CONTENTS

SYLLABUS

PAPER – III

MH–403 Sociology and Research Methods

UNIT-1

Scientific Inquiry: The scientific approach to human inquiry in comparison to the common approach; Meaning and purpose of research; Attitudes consonant to the scientific method; Social Work Research Meaning, scope and importance.

UNIT-II

Introduction of Research Strategies: Quantitative Methods; Qualitative Methods; Content, Scope, underlying assumption and strategy. Elements of sample; Rationale for sampling; Qualities of good sample/sampling strategy, Meaning and significance of a random and non-random sampling procedure; Simple and Stratified Random Sampling. Data collection and processing-Sources of data; Primary and secondary data; Interviewing and observation (Structured and Unstructured); simple instruments of data collection; data processing.

UNIT-III

Data Presentation: Graphical and tabular presentation of data-bar graphs, pie diagrams, histograms, polygons and line graphs; use, Unvaried and multivariate tables.

UNIT-IV

Statistical Tools: Percentage, ratios and proportions; Measures of Central Tendency (Mean, Mode and Median) Their computation, use, relative strengths and limitations.

UNIT-V

Writing Skills: Research Report Writing.

UNIT—I

SCIENTIFIC INQUIRY

NOTES

STRUCTURE

LEARNING OBJECTIVES

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

- understand the concept of socientific approach to human inquiry and common sense approach;
- state the meaning and purpose of research;
- explain the attitude consonant of scientific method;
- discuss the meaning, scope and importance of social work research.

1.1 INTRODUCTION

Scientific method refers to a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge. To be termed scientific, a method of inquiry must be based on gathering observable, empirical and measurable evidence subject to specific principles of reasoning. A scientific method consists of the collection of data through observation and experimentation, and the formulation and testing of hypotheses.

Although procedures vary from one field of inquiry to another, identifiable features distinguish scientific inquiry from other methodologies of knowledge. Scientific researchers propose hypotheses as explanations of phenomena, and

NOTES

design experimental studies to test these hypotheses. These steps must be repeatable in order to dependably predict any future results. Theories that encompass wider domains of inquiry may bind many independently derived hypotheses together in a coherent, supportive structure. This in turn may help form new hypotheses or place groups of hypotheses into context.

Among other facets shared by the various fields of inquiry is the conviction that the process must be objective to reduce biased interpretations of the results. Another basic expectation is to document, archive and share all data and methodology so they are available for careful scrutiny by other scientists, thereby allowing other researchers the opportunity to verify results by attempting to reproduce them. This practice, called full disclosure, also allows statistical measures of the reliability of these data to be established.

1.2 THE SCIENTIFIC APPROACH TO HUMAN INQUIRY IN COMPARISON TO THE NATIVE OR COMMON SENSE APPROACH

A scientific method or process is considered fundamental to the scientific investigation and acquisition of new knowledge based upon verifiable evidence. In addition to employing the scientific method in their research, sociologists explore the social world with several different purposes in mind. Like the physical sciences (i.e., chemistry, physics, etc.), sociologists can be and often are interested in predicting outcomes given knowledge of the variables and relationships involved. This approach to doing science is often termed positivism (though perhaps more accurately should be called empiricism). The positivist approach to social science seeks to explain and predict social phenomena, often employing a quantitative approach. But unlike the physical sciences, sociology (and other social sciences, like anthropology) also often seek simply to understand social phenomena.

Max Weber labeled this approach Verstehen, which is German for understanding. In this approach the goal is to understand a culture or phenomenon on its own terms rather than trying to develop a theory that allow for prediction. Both approaches employ a scientific method as they make observations and gather data, propose hypotheses, and test their hypotheses in the formulation of theories. These steps are outlined in more detail below.

Sociologists use observations, hypotheses and deductions to propose explanations for social phenomena in the form of theories. Predictions from these theories are tested. If a prediction turns out to be correct, the theory survives. If not, the theory is modified or discarded. The method is commonly taken as the underlying logic of scientific practice. Science is essentially an extremely cautious means of building a supportable, evidenced understanding of our natural world.

The essential elements of a scientific method are iterations and recursions of the following four steps:

1. Characterization (operationalization or quantification, observation and measurement).
2. Hypothesis (a theoretical, hypothetical explanation of the observations and measurements).
3. Prediction (logical deduction from the hypothesis).
4. Experiment (test of all of the above; in the social sciences, true experiments are often replaced with a different form of data analysis).

Characterization

A scientific method depends upon a careful characterization of the subject of the investigation. While seeking the pertinent properties of the subject, this careful thought may also entail some definitions and observations; the observation often demands careful measurement and/or counting.

The systematic, careful collection of measurements or counts of relevant quantities is often the critical difference between pseudo-sciences, such as alchemy, and a science, such as chemistry. Scientific measurements are usually tabulated, graphed, or mapped, and statistical manipulations, such as correlation and regression, performed on them. The measurements might be made in a controlled setting, such as a laboratory, or made on more or less inaccessible or unmanipulatable objects such as human populations. The measurements often require specialized scientific instruments such as thermometers, spectroscopes, or voltmeters, and the progress of a scientific field is usually intimately tied to their invention and development.

Measurements demand the use of operational definitions of relevant quantities (*a.k.a.* operationalization). That is, a scientific quantity is described or defined by how it is measured, as opposed to some more vague, inexact or idealized definition. The operational definition of a thing often relies on comparisons with standards: the operational definition of mass ultimately relies on the use of an artifact, such as a certain kilogram of platinum kept in a laboratory in France. In short, to operationalize a variable means creating an operational definition for a concept someone intends to measure. How this is done is very important as it should be done with enough precision that an independent researchers should be able to use your description of your measurement and repeat it.

The scientific definition of a term sometimes differs substantially from its natural language usage. For example, sex and gender are often used interchangeably in common discourse, but have distinct meanings in sociology. Scientific quantities are often characterized by their units of measure which can later be described in terms of conventional physical units when communicating the work.

Measurements in scientific work are also usually accompanied by estimates of their uncertainty. The uncertainty is often estimated by making repeated measurements of the desired quantity. Uncertainties may also be calculated by consideration of the uncertainties of the individual underlying quantities that are used. Counts of things, such as the number of people in a nation at a particular time, may also have an uncertainty due to limitations of the method used. Counts may only represent a sample of desired quantities, with an uncertainty that depends upon the sampling method used and the number of samples taken (see the central limit theorem).

Hypothesis Development

A hypothesis includes a suggested explanation of the subject. It will generally provide a causal explanation or propose some association between two variables. If the hypothesis is a causal explanation, it will involve at least one dependent variable and one independent variable.

Variables are measurable phenomena whose values can change (e.g., class status can range from lower- to upper-class). A dependent variable is a variable whose values are presumed to change as a result of the independent variable. In other words, the value of a dependent variable depends on the value of the independent variable. Of course, this assumes that there is an actual relationship between the two variables. If there is no relationship, then the value of the dependent variable does not depend on the value of the independent variable. An independent variable is a variable whose value is manipulated by the experimenter (or, in the case of non-experimental analysis, changes in the society and is measured). Perhaps an example will help clarify. In a study of the influence of gender on promotion, the independent variable would be gender/sex. Promotion would be the dependent variable. Change in promotion is hypothesized to be dependent on gender.

Scientists use whatever they can — their own creativity, ideas from other fields, induction, systematic guessing, etc. — to imagine possible explanations for a phenomenon under study. There are no definitive guidelines for the production of new hypotheses. The history of science is filled with stories of scientists claiming a flash of inspiration, or a hunch, which then motivated them to look for evidence to support or refute their idea.

Prediction

A useful hypothesis will enable predictions, by deductive reasoning, that can be experimentally assessed. If results contradict the predictions, then the hypothesis under examination is incorrect or incomplete and requires either revision or abandonment. If results confirm the predictions, then the hypothesis might be correct but is still subject to further testing. Predictions refer to experimental designs with a currently unknown outcome. A prediction (of an unknown) differs from a consequence (which can already be known).

Experiment

Once a prediction is made, an experiment is designed to test it. The experiment may seek either confirmation or falsification of the hypothesis.

Scientists assume an attitude of openness and accountability on the part of those conducting an experiment. Detailed record keeping is essential, to aid in recording and reporting on the experimental results, and providing evidence of the effectiveness and integrity of the procedure. They will also assist in reproducing the experimental results.

The experiment's integrity should be ascertained by the introduction of a control. Two virtually identical experiments are run, in only one of which the factor being tested is varied. This serves to further isolate any causal phenomena. For example in testing a drug it is important to carefully test that the supposed effect of the drug is produced only by the drug. Doctors may do this with a double-blind study: two virtually identical groups of patients are compared, one of which receives the drug and one of which receives a placebo. Neither the patients nor the doctor know who is getting the real drug, isolating its effects. This type of experiment is often referred to as a true experiment because of its design. It is contrasted with alternative forms below.

Once an experiment is complete, a researcher determines whether the results (or data) gathered are what was predicted. If the experimental conclusions fail to match the predictions/hypothesis, then one returns to the failed hypothesis and re-iterates the process - modifying one's theory or developing a new one. If the experiment appears successful - i.e. fits the hypothesis - the experimenter often will attempt to publish the results so that others (in theory) may reproduce the same experimental results, verifying the findings in the process.

An experiment is not an absolute requirement. In observation based fields of science actual experiments must be designed differently than for the classical laboratory based sciences. Due to ethical concerns and the sheer cost of manipulating large segments of society, sociologists often turn to other methods for testing hypotheses. In lieu of holding variables constant in laboratory settings, sociologists employ statistical techniques (e.g., regression) that allow them to control the variables in the analysis rather than in the data collection. For instance, in examining the effects of gender on promotions, sociologists may control for the effects of social class as this variable will likely influence the relationship.

NOTES

Unlike a true experiment where these variables are held constant in a laboratory setting, sociologists use statistical methods to hold constant social class (or, better stated, partial out the variance accounted for by social class) so they can see the relationship between gender and promotions without the interference of social class. Thus, while the true experiment is ideally suited for the performance of science, especially because it is the best method for deriving causal relationships, other methods of hypothesis testing are commonly employed in the social sciences.

Evaluation and Iteration

The scientific process is iterative. At any stage it is possible that some consideration will lead the scientist to repeat an earlier part of the process. For instance, failure of a hypothesis to produce interesting and testable predictions may lead to reconsideration of the hypothesis or of the definition of the subject.

It is also important to note that science is a social enterprise, and scientific work will become accepted by the community only if it can be verified. Crucially, experimental and theoretical results must be reproduced by others within the scientific community. All scientific knowledge is in a state of flux, for at any time new evidence could be presented that contradicts a long-held hypothesis. For this reason, scientific journals use a process of peer review, in which scientists' manuscripts are submitted by editors of scientific journals to (usually one to three) fellow (usually anonymous) scientists familiar with the field for evaluation. The referees may or may not recommend publication, publication with suggested modifications, or, sometimes, publication in another journal. This serves to keep the scientific literature free of unscientific work, helps to cut down on obvious errors, and generally otherwise improves the quality of the scientific literature. Work announced in the popular press before going through this process is generally frowned upon. Sometimes peer review inhibits the circulation of unorthodox work, and at other times may be too permissive. The peer review process is not always successful, but has been very widely adopted by the scientific community.

The reproducibility or replication of scientific observations, while usually described as being very important in a scientific method, is actually seldom reported, and is in reality often not done. Referees and editors often reject papers purporting only to reproduce some observations as being unoriginal and not containing anything new. Occasionally reports of a failure to reproduce results are published - mostly in cases where controversy exists or a suspicion of fraud develops. The threat of failure to replicate by others, however, serves as a very effective deterrent for most scientists, who will usually replicate their own data several times before attempting to publish.

Sometimes useful observations or phenomena themselves cannot be reproduced. They may be rare, or even unique events. Reproducibility of observations and replication of experiments is not a guarantee that they are correct or properly understood. Errors can all too often creep into more than one laboratory.

NOTES

Native or Common Sense Approach

Common sense approach, based on a strict construction of the term, consists of what people in common would agree on: that which they "sense" as their common natural understanding. Some people (such as the authors of Merriam-Webster Online) use the phrase to refer to beliefs or propositions that — in their opinion — most people would consider prudent and of sound judgment, without reliance on esoteric knowledge or study or research, but based upon what they see as knowledge held by people "in common".

Thus "common sense" (in this view) equates to the knowledge and experience which most people already have, or which the person using the term believes that they do or should have. However this is not the common dictionary definition. The most common meaning to the phrase is good sense and sound judgement in practical matters. It has nothing to do with what other people may think or feel.

Whatever definition one uses, identifying particular items of knowledge as "common sense" becomes difficult. Philosophers may choose to avoid using the phrase when using precise language. But common sense remains a perennial topic in epistemology and many philosophers make wide use of the concept or at least refer to it. Some related concepts include intuitions, pre-theoretic belief, ordinary language, the frame problem, foundational beliefs, good sense, endoxa, and axioms.

Common-sense ideas tend to relate to events within human experience (such as good will), and thus appear commensurate with human scale. Humans lack any commonsense intuition of, for example, the behavior of the universe at subatomic distances or of speeds approaching that of light. Often ideas that may be considered to be true by common sense are in fact false. It is seldom used as a professional Word.

Principles

In its strongest original formulation, it could be thought of as a set of five principles:

1. The unity of the scientific method – i.e., the logic of inquiry is the same across all sciences (social and natural).
2. The goal of inquiry is to explain and predict. Most positivists would also say that the ultimate goal is to develop the law of general understanding,

by discovering necessary and sufficient conditions for any phenomenon (creating a perfect model of it). If the law is known, we can manipulate the conditions to produce the predicted result.

3. Scientific knowledge is testable. Research can be proved only by empirical means, not argumentations. Research should be mostly deductive, i.e. deductive logic is used to develop statements that can be tested (theory leads to hypothesis which in turn leads to discovery and/or study of evidence). Research should be observable with the human senses (arguments are not enough, sheer belief is out of the question). Positivists should prove their research using the logic of confirmation.
4. Science does not equal common sense. Researchers must be careful not to let common sense bias their research.
5. The relation of theory to practice – science should be as value-free as possible, and the ultimate goal of science is to produce knowledge, regardless of any politics, morals, or values held by those involved in the research. Science should be judged by logic, and ideally produce universal conditionals:

- For all conditions of X, if X has property P and P=Q, then X has property Q.
- Statements must be true for all times and places.

There are now no fewer than twelve distinct epistemologies that are referred to as positivism. Many of these approaches dispose of a number of these claims. For example, most contemporary social researchers do not believe in the existence of general social laws.

1.3 MEANING AND PURPOSE OF RESEARCH

In the broadest sense of the word, the definition of research includes any gathering of data, information and facts for the advancement of knowledge. Reading a factual book of any sort is a kind of research. Surfing the internet or watching the news is also a type of research.

Research must be systematic and follow a series of steps and a rigid standard protocol. These rules are broadly similar but may vary slightly between the different fields of social science.

Research is defined as human activity based on intellectual application in the investigation of matter. The primary purpose for applied research is discovering, interpreting, and the development of methods and systems for the advancement of human knowledge on a wide variety of scientific matters of our world and the universe. Research can use the scientific method, but need not do so.

THE PURPOSE OF RESEARCH

NOTES

The purpose of the research process is to produce new knowledge, which takes three main forms :

- Exploratory research, which structures and identifies new problems.
- Constructive research, which develops solutions to a problem.
- Empirical research, which tests the feasibility of a solution using empirical evidence.

Research can also fall into two distinct types:—

- *Primary research*
- Secondary research

Research is often conducted using the hourglass model Structure of Research. The hourglass model starts with a broad spectrum for research, focusing in on the required information through the methodology of the project (like the neck of the hourglass), then expands the research in the form of discussion and results.

The ultimate aims of research are to generate measurable and testable data, gradually adding to the accumulation of human knowledge. Ancient philosophers believed that all answers could be achieved through deduction and reasoning, rather than measurement. Science now uses established research methods and standard protocols to test theories thoroughly.

It is important to remember that science and philosophy are intertwined and essential elements of human advancement, both contributing to the way we view the world. Scientific research, however, allows us to test hypotheses and lay solid foundations for future research and study.

No theory or hypothesis can ever be completely proved or disproved, but research enables us to make valid assumptions about the universe. This gradual accumulation of knowledge dictates the overall direction of science and philosophy.

Observation and Description

The first stage of any research is to observe the world around us and to ask questions about why things are happening. Every phenomenon in the universe has a reason behind it, and the aims of research are to understand and evaluate what is happening.

However simple the phenomenon or however easy it appears to be to generate logical and intuitive answers, scientific research demands rigorous testing for a truth to be accepted. Describing the overall behavior of the subject is the first stage of any research, whether it is a case study or a full-blown 'true experimental design'.

NOTES

Predict

This stage is where you must make a statement of intent and develop a strong hypothesis. This must be testable, with aims of research being to prove or disprove this statement. At this stage, you may express your personal opinion, favoring one side or the other. You must make a statement predicting what you expect the final answer to be.

You must, however, keep an open mind and understand that there is a chance that you may be wrong. Research is never about right or wrong, but about arriving at an answer, which improves our knowledge of natural processes.

Determination of the Causes

This is often the 'business end' for many areas of scientific research and is where one of the predictions is tested, usually by manipulating and controlling variables. The idea is to generate numerical data that can determine the cause with one of the many statistical tests.

For example, a small-scale global warming study might study Antarctic ice cores to determine the historical levels of carbon dioxide throughout history. In this experiment, time would be the manipulated variable, showing how levels of the greenhouse gas have changed over time.

Statistical procedures are then utilized to either prove or disprove the hypothesis and prediction. Of course, very little research gives such a black and white answer, but opens up new areas of potential study, focusing on a specific direction.

Explain

After determining the causes, the next layer of the research process is to try to find possible explanations of 'Why?' and 'How?' things are happening. For most areas, this stage involves sifting through and reviewing earlier studies about similar phenomena. Most research is built upon the work of previous researchers, so there should be a wealth of literature resources available.

If we look at a topical example, Global Warming is an area with which most of us are familiar, and has been the subject of thousands of studies. Intuitively, most of us would state that humanity pumping carbon dioxide into the atmosphere is responsible for a worldwide rise in temperatures.

The aims of research may be to establish 'What are the underlying causes and relationships between the different processes fueling this trend?' In most cases, it is necessary to review earlier research and try to separate the better quality sources from the inaccurate or poorly designed studies.

It is equally important to take into account any opposing points of view and accept that they may be equally valid. Explanation is about coming up with viable reasons, and you must try to be as objective and unbiased as possible.

For example, for global warming, there is an opposing view that temperature rises are natural, and that the effect of human society is making little difference. At this stage, personal opinion must be put aside and both sides of the debate must be given equal credence.

NOTES

New Directions

Whatever the final answer, it can be used to promote a healthy debate and discussion about the validity of the results. The aims of research can then be fine-tuned, or may serve to open up new areas of interest. Either way, the store of human knowledge has been enriched and increased.

1.4 ATTITUDE CONSONANT TO THE SCIENTIFIC METHOD

Along with this scientific method, a scientific attitude is also important. The significant aspect of attitude consonant of scientific method is as follows:

1. The researcher tries to be impartial.
2. Measurements are more reliable than subjective results.
3. Experiments are designed to test (disprove), not just confirm.
4. Data are recorded so they can be studied, and so the experiments can be replicated.
5. Certain common errors (like *contamination* or *experimenter* bias) are guarded against.
6. Statistical analysis is used to study the data.

The Scientific Method is not a checklist that researchers consciously follow. It is implied in the way they do science.

Sometimes it may seem that errors occur, as a false hypothesis survives for a long time. Sometimes there actually are errors, and even fraud. But mostly, the false paths are natural, and are corrected eventually, merely by further testing. Sometimes, this may involve looking at the situation from an entirely different perspective.

1.5 SOCIAL WORK RESEARCH : MEANING, SCOPE AND IMPORTANCE

Social work research refers to research conducted by social scientists. Social work research methods may be divided into two broad categories:

- *Quantitative designs* approach social phenomena through quantifiable evidence, and often rely on statistical analysis of many cases (or across intentionally designed treatments in an experiment) to create valid and reliable general claims.

- Qualitative designs emphasize understanding of social phenomena through direct observation, communication with participants, or analysis of texts, and may stress contextual and subjective accuracy over generality.

Social scientists employ a range of methods in order to analyse a vast breadth of social phenomena; from census survey data derived from millions of individuals, to the in-depth analysis of a single agents' social experiences; from monitoring what is happening on contemporary streets, to the investigation of ancient historical documents.

The methods rooted in classical sociology and statistics have formed the basis for research in other disciplines, such as political science, media studies, and market research.

SCOPE

Social work research is the scientific study of society. More specifically, social research examines a society's attitudes, assumptions, beliefs, trends, stratifications and rules. The scope of social work research can be small or large, ranging from the self or a single individual to spanning an entire race or country. Popular topics of social research include poverty, racism, class issues, sexuality, voting behavior, gender constructs, policing and criminal behavior.

Social work research determines the relationship between one or more variables. For example, sex and income level are variables. Social scientists will look for underlying concepts and cause-and-effect relationships of a social issue. Before even beginning research, scientists must formulate a research question. For example, a researcher might ask if there is a relationship between a person's sex and his or her income level. Do men have higher incomes than women? Are women most likely to be poor?

A third variable, race, can be added to the question. Then the social scientist can pose a research question: Does race and sex affect a person's income level? Social scientists will then collect data, organize and analyze information and create a report of their findings. People conducting social research must also consider ethics, biases and the reliability and validity of the research they're conducting. They must decide which form of sampling to use, how to measure information, how to analyze data and present their findings.

Research can be conducted using surveys, reports, observation, questionnaires, focus groups, historical accounts, personal diaries and census statistics. There are two types of research: qualitative research and quantitative research. Qualitative research is inductive, meaning the researcher creates hypotheses and abstractions from collected data. Most data is collected via words or pictures and mostly from people. Researchers are interested in how people make sense of their lives and in the research process itself.

NOTES

Quantitative research is the complete opposite and most often involves numbers and set data. Quantitative data is efficient but focuses only on the end result, not the process itself, as qualitative research does. Quantitative data is precise and is often the result of surveys or questionnaires.

Even though social research is most often conducted by social scientists or sociologists, it is an interdisciplinary study crossing into subjects like criminology, the study of crime; politics, the study of power; economics, the study of money and business; psychology, study of the mind; philosophy, study of beliefs and morals; and anthropology, the study of culture.

IMPORTANCE

Social science or soft science studies the human aspects of the world. It includes an in-depth study and evaluation of human behavior by using scientific methods in either quantitative or qualitative manner. The subject matter of the soft science can be subjective or objective. In contrast to hard science or natural science, social science reveals subjective, objective, inter-subjective and structural aspects of the society. In simple words, it measures the social developments of the society and finds the emerging drawbacks too. The work of social science is to watch where the society has been heading and what more can be done to benefit the entire race.

If we dig out the historical evidences, we will find that social science has been a constant part of the development of the human race. It is interesting to know that great philosophers like Plato and Aristotle were advocates of social science and their philosophies are based on the objectives of social sciences only. So, social science is very much present into our day-to-day life and has a pivotal role in the societal growth and development.

Social research refers to research conducted by social scientists (primarily within sociology, but also within other disciplines such as social policy, human geography, political science, social anthropology and education. Sociologists and other social scientists study diverse things: from census data on hundreds thousands of human beings, through the in-depth analysis of a life of a single important person to monitoring what is happening on a streets today - or what was happening few hundreds years ago.

Social scientists use many different methods in order to describe, explore and understand social life. Social methods can generally be subdivided into two broad categories. Quantitative methods are concerned with attempts to quantify social phenomena and collect and analyse numerical data, and focus on the links among a smaller number of attributes across many cases. Qualitative methods, on the other hand, emphasise personal experiences and interpretation over quantification, are more concerned with understanding the meaning of social

NOTES

phenomena and focus on links among a larger number of attributes across relatively few cases. While very different in many aspects, both qualitative and quantitative approaches involve a systematic interaction between theories and data.

Common tools of quantitative researchers include surveys, questionnaires, and secondary analysis of statistical data that has been gathered for other purposes (for example, censuses or the results of social attitudes surveys). Commonly used qualitative methods include focus groups, participant observation, and other techniques.

SUMMARY

- Scientific method refers to a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge.
- Common sense approach, based on a strict construction of the term, consists of what people in common would agree on: that which they "sense" as their common natural understanding.
- Research is defined as human activity based on intellectual application in the investigation of matter. In the broadest sense of the word, the definition of research includes any gathering of data, information and facts for the advancement of knowledge.
- Social work research refers to research conducted by social scientists. Social work research methods may be divided into two broad categories *i.e.,* Quantitative and Qualitative.

REVIEW QUESTIONS

1. What do you understand by scientific approach to human inquiry?
2. Give a berief description about common sense approach.
3. State the meaning of "Research".
4. What are the principal aims of research?
5. What are the major scopes of social work research?

FURTHER READINGS

- Kothari, C. R.; *Research Methodology: Methods and Techniques;* Publisher: New Age International.
- King, Gary, Keohane, Robert O., & Verba, Sidney (1994). *Designing Social Inquiry.* Princeton, NJ: Princeton University Press.
- Merton, Robert K. (1949:1968). *Social Theory and Social Structure.* New York: Free Press.

UNIT— II

INTRODUCTION OF RESEARCH STRATEGIES

STRUCTURE

LEARNING OBJECTIVES

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

- discuss the concept of quantitative and qualitative research methods;
- state the meaning, elements and classification of sampling;
- explain the method of data collection and data processing.

2.1 INTRODUCTION

Social scientists are divided into camps of support for particular research techniques. These disputes relate to the historical core of social theory (positivism and antipositivism; structure and agency). While very different in many aspects, both qualitative and quantitative approaches involve a systematic interaction between theory and data. The choice of method often depends largely on what the researcher intends to investigate. For example, a researcher concerned with drawing a statistical generalization across an entire population may administer a survey questionnaire to a representitive sample population. By contrast, a researcher who seeks full contextual understanding of an individuals' social actions may choose ethnographic participant observation or open-ended interviews. Studies will commonly combine, or 'triangulate', quantitative and qualitative methods as part of a 'multi-strategy' design. For instance, a quantitative study may be performed to gain statistical patterns or a target sample, and then combined with a qualitative interview to determine the play of agency.

2.2 QUANTITATIVE METHOD: CONTENT, SCOPE, AND STRATEGY

In the social sciences, quantitative research refers to the systematic empirical investigation of quantitative properties and phenomena and their relationships. The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena. The process of measurement is central to quantitative research because it provides the fundamental connection between empirical observation and mathematical expression of quantitative relationships.

Quantitative research is used widely in social sciences such as sociology, anthropology, and political science. Research in mathematical sciences such as physics is also 'quantitative' by definition, though this use of the term differs in context. In the social sciences, the term relates to empirical methods, originating in both philosophical positivism and the history of statistics, which contrast qualitative research methods.

Qualitative methods produce information only on the particular cases studied, and any more general conclusions are only hypotheses. Quantitative methods can be used to verify, which of such hypotheses are true.

Strategy and Scope

Quantitative methods are research techniques that are used to gather quantitative data - information dealing with numbers and anything that is measurable. Statistics, tables and graphs, are often used to present the results of these methods. They are therefore to be distinguished from qualitative methods.

In most physical and biological sciences, the use of either quantitative or qualitative methods is uncontroversial, and each is used when appropriate. In the social sciences, particularly in sociology, social anthropology and psychology, the use of one or other type of method has become a matter of controversy and even ideology, with particular schools of thought within each discipline favouring one type of method and pouring scorn on to the other. Advocates of quantitative methods argue that only by using such methods can the social sciences become truly scientific; advocates of qualitative methods argue that quantitative methods tend to obscure the reality of the social phenomena under study because they underestimate or neglect the non-measurable factors, which may be the most important. The modern tendency (and in reality the majority tendency throughout the history of social science) is to use eclectic approaches. Quantitative methods might be used with a global qualitative frame. Qualitative methods might be used to understand the meaning of the numbers produced by quantitative methods. Using quantitative methods, it is possible to give precise and testable expression to qualitative ideas. This combination of quantitative and qualitative data gathering is often referred to as mixed-methods research.

Examples of Quantitative Research

- Research that consists of the percentage amounts of all the elements that make up Earth's atmosphere.
- Survey that concludes that the average patient has to wait two hours in the waiting room of a certain doctor before being selected.
- An experiment in which group x was given two tablets of Aspirin a day and Group y was given two tablets of a placebo a day where each participant is randomly assigned to one or other of the groups.

The numerical factors such as two tablets, percent of elements and the time of waiting make the situations and results quantitative.

2.3 QUALITATIVE METHODS

Qualitative method is a method of inquiry appropriated in many different academic disciplines, traditionally in the social sciences, but also in market research and further contexts. Qualitative researchers aim to gather an in-depth understanding of human behavior and the reasons that govern such behavior. The qualitative method investigates the why and how of decision making, not just what, where, when. Hence, smaller but focused samples are more often needed, rather than large samples.

Qualitative methods produce information only on the particular cases studied, and any more general conclusions are only hypotheses (informative guesses). Quantitative methods can be used to verify which of such hypotheses are true.

Strategy and Scope

Qualitative research is often used for policy and program evaluation research since it can answer certain important questions more efficiently and effectively than quantitative approaches. This is particularly the case for understanding how and why certain outcomes were achieved (not just what was achieved) but also answering important questions about relevance, unintended effects and impact of programs such as: Were expectations reasonable? Did processes operate as expected? Were key players able to carry out their duties? Were there any unintended effects of the program? Qualitative approaches have the advantage of allowing for more diversity in responses as well as the capacity to adapt to new developments or issues during the research process itself. While qualitative research can be expensive and time-consuming to conduct, many fields of research employ qualitative techniques that have been specifically developed to provide more succinct, cost-efficient and timely results. Rapid Rural Appraisal is one formalised example of these adaptations but there are many others.

Data Collection

Qualitative researchers may use different approaches in collecting data, such as the grounded theory practice, narratology, storytelling, classical ethnography, or shadowing. Qualitative methods are also loosely present in other methodological approaches, such as action research or actor-network theory. Forms of the data collected can include interviews and group discussions, observation and reflection field notes, various texts, pictures, and other materials.

Qualitative research often categorizes data into patterns as the primary basis for organizing and reporting results. Qualitative researchers typically rely on the following methods for gathering information: Participant Observation, Non-participant Observation, Field Notes, Reflexive Journals, Structured Interview, Semi-structured Interview, Unstructured Interview, and Analysis of documents and materials.

The ways of participating and observing can vary widely from setting to setting. Participant observation is a strategy of reflexive learning, not a single method of observing. In participant observation researchers typically become members of a culture, group, or setting, and adopt roles to conform to that setting. In doing so, the aim is for the researcher to gain a closer insight into the culture's practices, motivations and emotions. It is argued that the researchers' ability to understand the experiences of the culture may be inhibited if they observe without participating.

Some distinctive qualitative methods are the use of focus groups and key informant interviews. The focus group technique involves a moderator facilitating a small group discussion between selected individuals on a particular topic. This is a particularly popular method in market research and testing new initiatives with users/workers.

NOTES

One traditional and specialized form of qualitative research is called cognitive testing or pilot testing which is used in the development of quantitative survey items. Survey items are piloted on study participants to test the reliability and validity of the items.

In the academic social sciences the most frequently used qualitative research approaches include the following:

1. Ethnographic Research, used for investigating cultures by collecting and describing data that is intended to help in the development of a theory. This method is also called "ethnomethodology" or "methodology of the people". An example of applied ethnographic research, is the study of a particular culture and their understanding of the role of a particular disease in their cultural framework.
2. Critical Social Research, used by a researcher to understand how people communicate and develop symbolic meanings.
3. Ethical Inquiry, an intellectual analysis of ethical problems. It includes the study of ethics as related to obligation, rights, duty, right and wrong, choice etc.
4. Foundational Research, examines the foundations for a science, analyses the beliefs and develops ways to specify how a knowledge base should change in light of new information.
5. Historical Research, allows one to discuss past and present events in the context of the present condition, and allows one to reflect and provide possible answers to current issues and problems. Historical research helps us in answering questions such as: Where have we come from, where are we, who are we now and where are we going?
6. Grounded Theory, is an inductive type of research, based or "grounded" in the observations or data from which it was developed; it uses a variety of data sources, including quantitative data, review of records, interviews, observation and surveys.
7. Phenomenology, describes the "subjective reality" of an event, as perceived by the study population; it is the study of a phenomenon.
8. Philosophical Research, is conducted by field experts within the boundaries of a specific field of study or profession, the best qualified individual in any field of study to use an intellectual analyses, in order to clarify definitions, identify ethics, or make a value judgment concerning an issue in their field of study.

Data Analysis

The most common analysis of qualitative data is observer impression. That is, expert or bystander observers examine the data, interpret it via forming an

NOTES

impression and report their impression in a structured and sometimes quantitative form.

Coding

Coding is an interpretive technique that both organizes the data and provides a means to introduce the interpretations of it into certain quantitative methods. Most coding requires the analyst to read the data and demarcate segments within it. Each segment is labeled with a "code" – usually a word or short phrase that suggests how the associated data segments inform the research objectives. When coding is complete, the analyst prepares reports via a mix of: summarizing the prevalence of codes, discussing similarities and differences in related codes across distinct original sources/contexts, or comparing the relationship between one or more codes.

Some qualitative data that is highly structured (e.g., open-end responses from surveys or tightly defined interview questions) is typically coded without additional segmenting of the content. In these cases, codes are often applied as a layer on top of the data. Quantitative analysis of these codes is typically the capstone analytical step for this type of qualitative data.

Contemporary qualitative data analyses are sometimes supported by computer programs. These programs do not supplant the interpretive nature of coding but rather are aimed at enhancing the analyst's efficiency at data storage/retrieval and at applying the codes to the data. Many programs offer efficiencies in editing and revising coding, which allow for work sharing, peer review, and recursive examination of data.

A frequent criticism of coding method is that it seeks to transform qualitative data into quantitative data, thereby draining the data of its variety, richness, and individual character. Analysts respond to this criticism by thoroughly expositing their definitions of codes and linking those codes soundly to the underlying data, therein bringing back some of the richness that might be absent from a mere list of codes.

Recursive Abstraction

Some qualitative datasets are analyzed without coding. A common method here is recursive abstraction, where datasets are summarized, those summaries are then further summarized, and so on. The end result is a more compact summary that would have been difficult to accurately discern without the preceding steps of distillation.

A frequent criticism of recursive abstraction is that the final conclusions are several times removed from the underlying data. While it is true that poor initial summaries will certainly yield an inaccurate final report, qualitative analysts can respond to this criticism. They do so, like those using coding method, by documenting the reasoning behind each summary step, citing examples from the data where statements were included and where statements were excluded from the intermediate summary.

Mechanical Techniques

Some techniques rely on leveraging computers to scan and sort large sets of qualitative data. At their most basic level, mechanical techniques rely on counting words, phrases, or coincidences of tokens within the data. Often referred to as content analysis, the output from these techniques is amenable to many advanced statistical analyses.

Mechanical techniques are particularly well-suited for a few scenarios. One such scenario is for datasets that are simply too large for a human to effectively analyze, or where analysis of them would be cost prohibitive relative to the value of information they contain. Another scenario is when the chief value of a dataset is the extent to which it contains "red flags" (e.g., searching for reports of certain adverse events within a lengthy journal dataset from patients in a clinical trial) or "green flags" (e.g., searching for mentions of your brand in positive reviews of marketplace products).

A frequent criticism of mechanical techniques is the absence of a human interpreter. And while masters of these methods are able to write sophisticated software to mimic some human decisions, the bulk of the "analysis" is nonhuman. Analysts respond by proving the value of their methods relative to either a) hiring and training a human team to analyze the data or b) letting the data go untouched, leaving any actionable nuggets undiscovered.

2.4 ASSUMPTIONS OF QUANTITATIVE AND QUALITATIVE METHODS

As we have defined quantitative research, let's compare it with qualitative research, against which it is usually contrasted. While quantitative research is based on numerical data analyzed statistically, qualitative research uses non-numerical data. Qualitative research is actually an umbrella term encompassing a wide range of methods, such as interviews, case studies, ethnographic research and discourse analysis, to name just a few. The difference between quantitative and qualitative research is often seen as quite fundamental, leading people to talk about 'paradigm wars' in which quantitative and qualitative research are seen as belligerent and incompatible factions. Many researchers define themselves as either quantitative or qualitative. Where does this idea come from?

This idea is linked to what is seen as the different underlying philosophies and worldviews of researchers in the two 'paradigms' (also called 'epistemologies' or sometimes 'assumptions'). According to this view, two fundamentally different worldviews underlie quantitative and qualitative research. The quantitative view

NOTES

is described as being 'realist' or sometimes 'positivist', while the worldview underlying qualitative research is viewed as being 'subjectivist'.

What does this mean? Realists take the view that what research does is to uncover an existing reality. 'The truth is out there' and it is the job of the researcher to use objective research methods to uncover that truth. This means that the researcher needs to be as detached from the research as possible, and use methods that maximize objectivity and minimize the involvement of the researcher in the research. This is best done using methods taken largely from the natural sciences and then transposed to social sciences. Positivism is the most extreme form of this worldview. According to positivism, the world works according to fixed laws of cause and effect.

However, the view that there is a true reality that we can measure completely objectively is problematic. We are all part of the world we are observing, and cannot completely detach ourselves from what we are researching. Historical research has shown that what is studied and what findings are produced are influenced by the beliefs of the people doing the research and the political/social climate at the time the research is done.

If one looks at research from a quantitative versus qualitative perspective,qualitative researchers are subjectivists. In contrast to the realist view that the truth is out there and can be objectively measured and found through research, subjectivists point to the role of human subjectivity in the process of research. Reality is not 'out there' to be objectively and dispassionately observed by us, but is at least in part constructed by us and by our observations. There is no pre-existing objective reality that can be observed. The process of our observing reality changes and transforms it, and; therefore, subjectivists are relativistic. All truth can only be relative and is never definitive as the positivists claim.

Besides, they are different in some other dimensions as shown in the Table below:

Table 2.1. Quantitative and Qualitative Paradigm Assumptions

Assumptions	Questions	Quantitative	Qualitative
Ontological Assumption	What is the nature of reality?	Reality is objective and singular apart from the researcher.	Reality is subjective and multiple as seen by participants in a study.
Epistemological Assumption	What is the relationship of the researcher to that researched?	Researcher is independent from that being researched.	Researcher interacts with that being researched.

Axiological Assumption	What is the role of values?	Value-free and unbiased	Value-laden and biased
Rhetorical Assumption	What is the language of research?	Formal, Based on set definitions, Impersonal voice, and Use of accepted quantitative words	Informal, Evolving decisions, Personal voice, and Accepted qualitative words
Methodological Assumption	What is the process of research?	Deductive process, Cause and effect, Static design - *categories isolated* before study, Generalizations leading to prediction, explanation, and understanding, and Accurate and reliable through validity and reliability	Inductive process, Mutual simultaneous *shaping of factors,* Emerging design - categories identified during research process, Context-bound, Patterns, theories developed for understanding, and Accurate and reliable through *verification.*

2.5 ELEMENTS OF SAMPLING

Sampling is that part of statistical practice concerned with the selection of individual observations intended to yield some knowledge about a population of concern, especially for the purposes of statistical inference.

Each observation measures one or more properties (weight, location, etc.) of an observable entity *enumerated to distinguish objects or individuals.* Survey weights often need to be applied to the data to adjust for the sample design. Results from probability theory and statistical theory are employed to guide practice.

The sampling comprises several stages:

- Defining the population of concern;
- Specifying a sampling frame, a set of items or events possible to measure;
- Specifying a sampling method for selecting items or events from the frame;
- Determining the sample size;
- Implementing the sampling plan;
- Sampling and data collecting;
- Reviewing the sampling process.

2.6 MEANING OF SAMPLING

Sampling is that part of statistical practice concerned with the selection of an unbiased or random subset of individual observations within a population of individuals intended to yield some knowledge about the population of concern, especially for the purposes of making predictions based on statistical inference. Sampling is an important aspect of data collection.

Researchers rarely survey the entire population for two reasons (Adèr, Mellenbergh, & Hand, 2008): the cost is too high, and the population is dynamic in that the individuals making up the population may change over time. The three main advantages of sampling are that the cost is lower, data collection is faster, and since the data set is smaller it is possible to ensure homogeneity and to improve the accuracy and quality of the data.

Each observation measures one or more properties (such as weight, location, color) of observable bodies distinguished as independent objects or individuals. In survey sampling, survey weights can be applied to the data to adjust for the sample design. Results from probability theory and statistical theory are employed to guide practice. In business and medical research, sampling is widely used for gathering information about a population.

2.7 QUALITIES OF GOOD SAMPLING

A good and purposeful sampling selects information rich cases for indepth study. Size and specific cases depend on the study purpose.

There are about 16 different types of purposeful sampling. They are briefly described below for you to be aware of them.

Extreme and deviant case sampling: This involves learning from highly unusual manifestations of the phenomenon of interest, suchas outstanding successes, notable failures, top of the class, dropouts, exotic events, crises.

Intensity sampling: This is information rich cases that manifest the phenomenon intensely, but not extremely, such as good students,poor students, above average/ below average.

Maximum variation sampling: This involves purposefully picking a wide range of variation on dimentions of interest. This documents unique or diverse variations that have emerged in adapting to different conditions. It also identifies important common patterns that cut across variations.

Homogenious sampling: This one reduces variation, simplifies analysis, facilitates group interviewing. Like instead of having the maximum number of nationalities as in the above case of maximum variation, it may focus on one nationality say Americans only.

NOTES

Typical case sampling: It involves taking a sample of what one would call typical, normal or average for a particular phenomenon,

Stratified purposeful sampling: This illustrates charecteristics of particular subgroups of interest and facilitates comparisons between the different groups.

Critical case sampling: This permits logical generalization and maximum application of information to other cases like "If it is true for this one case, it is likely to be true of all other cases. You must have heard statements like if it happenned to so and so then it can happen to anybody. Or if so and so passed that exam, then anybody can pass.

Snowball or chain sampling: This particular one identifies, cases of interest from people who know people who know what cases are information rich, that is good examples for study, good interview subjects. This is commonly used in studies that may be looking at issues like the homeless households. What you do is to get hold of one and he/she will tell you where the others are or can be found. When you find those others they will tell you where you can get more others and the chain continues.

Criterion sampling: Here, you set a criteria and pick all cases that meet that criteria for example, all ladies six feet tall, all white cars, all farmers that have planted onions. This method of sampling is very strong in quality assurance.

Theory based or operational construct sampling: It is about finding manifestations of a theoretical construct of interest so as to elaborate and examine the construct.

Confirming and disconfirming cases: Elaborating and deepening initial analysis like if you had already started some study, you are seeking further information or confirming some emerging issues which are not clear, seeking exceptions and testing variation.

Opportunistic Sampling: This involves following new leads during field work, taking advantage of the unexpected flexibility.

Random purposeful sampling: This adds credibility when the purposeful sample is larger than one can handle. Reduces judgement within a purposeful category. But it is not for generalizations or representativeness.

Sampling politically important cases: This type of sampling attracts or avoids attracting attention undesired attention by purposisefully eliminating from the sample political cases. These may be individuals, or localities.

Convenience sampling: It is useful in getting general ideas about the phenomenon of interest. For example you decide you will interview the first ten people you meet tomorrow morning. It saves time, money and effort. It is the poorest way of getting samples, has the lowest credibility and yields information-poor cases.

Combination or mixed purposeful sampling: This combines various sampling strategies to achieve the desired sample. This helps in triangulation, allows for flexibility, and meets multiple interests and needs. When selecting a sampling strategy it is necessary that it fits the purpose of the study, the resources available, the question being asked and the constraints being faced. This holds true for sampling strategy as well as sample size.

2.8 MEANING AND SIGNIFICANCE OF A RANDOM AND NON-RANDOM SAMPLING PROCEDURE

A sample is a subject chosen from a population for investigation. A random sample is one chosen by a method involving an unpredictable component. Random sampling can also refer to taking a number of independent observations from the same probability distribution, without involving any real population. A probability sample is one in which each item has a known probability of being in the sample.

The sample usually will not be completely representative of the population from which it was drawn— this random variation in the results is known as sampling error. In the case of random samples, mathematical theory is available to assess the sampling error. Thus, estimates obtained from random samples can be accompanied by measures of the uncertainty associated with the estimate. This can take the form of a standard error, or if the sample is large enough for the central limit theorem to take effect, confidence intervals may be calculated.

Types of Random Sample

- A simple random sample is selected so that all samples of the same size have an equal chance of being selected from the population.
- A self-weighting sample, also known as an EPSEM (Equal Probability of Selection Method) sample, is one in which every individual, or object, in the population of interest has an equal opportunity of being selected for the sample. Simple random samples are self-weighting.
- Stratified sampling involves selecting independent samples from a number of subpopulations, group or strata within the population. Great gains in efficiency are sometimes possible from judicious stratification.
- Cluster sampling involves selecting the sample units in groups. For example, a sample of telephone calls may be collected by first taking a collection of telephone lines and collecting all the calls on the sampled lines. The analysis of cluster samples must take into account the intra-cluster correlation which reflects the fact that units in the same cluster are likely to be more similar than two units picked at random.

NOTES

Importance

When conducting research like surveys, it's not always possible to interview or analyze every single person or object of interest. Researchers have to choose only some people or objects to include in the study. However, this selection must be conducted carefully to ensure that the research results based on this small group—a sample—are accurate when applied to all the people or objects that exist (called the population in statistical terminology).

It is very important to develop a sampling plan that permits legitimate generalization from the survey results to the population of interest (unless the study specifically calls for a non-probability sample design). The key is the use of statistically derived random sampling procedures. These ensure that survey results can be defended as statistically representative of the population. Surveys that do not follow these procedures can produce results that lead to misguided market research, strategic, or policy decisions. Any so-called "survey" in which no attempt is made to randomly select respondents, such as call-in readers' or viewers' "polls", is likely to produce results that in no way reflect overall public opinion—even if many thousands of individuals participate.

Non-Random Sampling

Non-random (or "nonprobability") samples are selected by any kind of procedure that does not give all cases in the population equal chances to fall into the sample. Sometimes the context of the study allows or facilitates using a certain method of sampling, sometimes the researcher has the possibility of selecting the method. Various such procedures will be discussed below.

Whatever the procedure, it is always possible that it will favor certain types of cases in the population more than the others, in other words the data from the sample will be biased.

In descriptive studies the presence of bias is usully a grave handicap, because it can prohibit generalizing the results into the population. This is a difficulty that you will meet later in your project, when Assessing Non-Random Sampling and when writing the final chapter of your report, so it can be prudent to think about it in advance, when selecting the sampling method.

When assessing a non-random sample you should ask yourself: Will the results from the sample be the same that you would get from the population? Is it certain that the criterion that you have used in selecting the sample (e.g. the willingness of people to participate) has no relationship with those variables that you want to record from the sample? If there is correlation, your sample will be biased and you should consider constructing a new sample with less correlation.

As a contrast, in research and development projects the risks in using non-random samples are smaller, because often a bias can be compensated later. For

NOTES

example, it is common to use convenience sampling when selecting potential customers to a think-tank in order to develop an early product concept. The selection of persons will probably be biased, as well as the proposals from the think-tank, but normally the proposals will be rectified at a later stage when they are evaluated anew by another, larger group of people.

Common types of non-random samples include:

1. Convenience sample. A coincidental group, e.g. people at a meeting, might be specified as a sample. More exactly, the sample contains those persons in the group who are willing to take part. Such a sample is often heavily biased, but this can be accepted if the data obtained from the sample are not really going to be used, such as is the case in demonstrations in survey method classes at universities. Likewise, this is a possible method when you need a few potential customers to assist in product development, on the condition that the obtained results shall later be tested with a better sample from the intended customers.
2. Sample of volunteers is created when all the members of a population have the opportunity to participate in the sample, and all volunteers are accepted. If you insert a survey form in an Internet page and ask people to give their opinions on a topic, you will get this type of sample from all the readers of this page. Similarly, the persons who spontaneously send customer feedback to a company are a sample of volunteers from all the customers.

   A sample of volunteers can be a practical alternative when there is no list of the members of the population from which a random sample could be drawn, or when it is difficult to contact the people in a sample because their addresses are not known. The disadvantage is that it is difficult to assess the presence of bias, i.e. whether the opinions or other interesting properties of the volunteers deviate from those of the population. When considering this question, there are two questions to ask:

   - What is the population that you are aiming at? Have all the members of the target population equal chances to be included in the sample?
   - Is there any reason why the volunteers should differ from the rest of the population? For example, have they, or at least some of them, a special reason for volunteering?

   If you, for example, want to get a sample of those people that have bought your latest product, you can include in the package of the product a postage-paid form where the people can give their names and addresses. What would happen if you additionally asked the respondents to give their opinions of the product? Quite probably you would get answers

NOTES

mostly from people that have a strong opinion of your product, either a positive or negative one. The people with no definite view of your product would probably not so often bother to answer. The sample would thus risk being biased, and you would have to consider whether such a bias could be acceptable for your purposes.

3. Snowball sample. When interviewing members of a population, you can ask the interviewed persons to nominate other individuals who could be asked to give information or opinion on the topic. You then interview these new individuals and continue in the same way until the material gets saturated, *i.e.*, you get no new viewpoints from the new persons.

   Snowball sampling is a good method for such populations that are not well delimited nor well enumerated, for example the homeless. The drawback is that you get no exact idea of the factual distribution of the opinions in the target population. Besides, people usually propose people that they know well and who share their own views, which means that small groups of interest often are passed by unnoticed. One method for compensating this could be asking people to nominate both such persons who share the same views and such persons who are of the opposite opinion. Another method is to start the snowball chain from not one but several different people, perhaps from different social groups.

4. Sample that consists of all the available cases. Sometimes the researcher is interested in a population of which only a few cases or specimens are available for study, and these then must serve as a sample of the population. Typical such samples are:

   - Surviving cases.
   - Permitted cases.

Surviving cases among historical or archaeological material, when a large part of once relevant material have disappeared before researchers get at it, can be regarded as a kind of convenience sample even when it is the historical reality and not convenience that selects the sample. Both samples involve a similar handicap for research: if the disappearance of material, during the period until the study, has not been random nor proportional but instead somehow partial or selective, the remaining material will be biased and the researcher should try to assess the likely bias. You should ask yourself whether any of the following factors have affected differently on the preservation of different sorts of material:

- Has the material sometimes been selected for any purpose, for example in order to be kept in archives, libraries or museums?
- Have some objects in the material sometimes been replaced with new ones?

NOTES

- What sorts of things have, in prevous times, been regarded as rubbish, or worthy and proper to be preserved?
- Are there physical factors which can have affected differently on the preservation of various groups of material?

Permitted cases. When studying private enterprises it often happens that the management will not allow recording information from certain units in the organization. The management's decision perhaps is motivated by their judgement about the objectives of the study, but from the researcher's scientific point of view such a sample will often seem seriously biased.

2.9 INAPPROPRIATE METHODS OF SAMPLING

Overstepping the limits of population. You must not include in your sample items that are not members of the defined population. For example, in snowball sampling it often happens that some interviewed people nominate candidates that do not belong to the same population. Of course, you have often the option of altering your original delimitations.

Sample of typical cases. Often the goal when studying a heterogenous group is to find what is common and typical of the majority of the cases in the group. To this effect, sampling has sometimes been used so that the most typical cases are selected into the sample and all the extraordinary cases are left out. In the figure on the right, typical cases are marked with dots, and exceptional cases with the symbols + and x.

The selection of "typical" cases is not quite commendable because the researcher's prejudices (which can be biased) influence too much the final results of the inquiry. The researcher can, without noticing it, select mostly such cases which corroborate his preconceptions or hypotheses. To sum up, if you want to point out the average or the most common cases in the population, a better method is to classify all items of the population or a random sample of it, and note the most frequent type. When necessary, you can then continue the study of this class, which hereafter becomes the new population of study.

Sample of specialists. It might look like a sensible idea to ask directly those, usually few, people that know a lot about the topic, instead of asking a large sample of randomly selected laymen whose knowledge can be sporadic and opinions may diverge. In this way, we might, for example:

- Investigate consumer preferences of household devices by interviewing salespersons.
- Study life styles of tenants through a questionnaire to house managers or landlords.
- Test a new family car model by asking celebrated racing drivers to try and evaluate it.

- Assess the working atmosphere in a company by interviewing the managers.

The advantage in interviewing specialists is that you need to interview just a few people and in the discussion you get quickly to the point. Nevertheless, you should not think that a sample of "specialists" could be taken as a sample of "non-specialists". These are two different populations. You should not generalize the results from "specialists" to any other population than just the population of "specialists" whoever they may be.

If you anyway choose to interview specialists, you can do it, of course. If you then additionally want to gather the opinions of the average consumers, you should define these as a second population and select a suitable sample of it, too. One possibility is to make these two surveys in succession. You could perhaps use the results from the specialists as new hypotheses to be tested with another sample of the consumer population. In other words, you would use the interview of the specialists as a preliminary study only. Or the other way round - you can first consult ordinary consumers and then the specialists.

Normative sampling. Normative aspect is acceptable in development projects which aim at improving similar objects in the future, but it is better to keep it out of sampling because it is not compatible with the principles of representativeness and generalization.

Studying normatively only a "sample of the best exemplars" is quite a tradition in art history: you only take into account the great works of art. The idea is that the best cases are closest to the ideals that artists had in their time and in this sense they represent the truest art of the era. They, too, had the greatest influence into later development. However, it is self-evident that the best works are not typical of the era and they do not represent average works of art. This does not suggest that you should not study them, but if you do it, do not call it a "sample" if you mean that the population of your study are the great masters. Cf. the discussion under Demarcating the Study.

Later in the project, when analyzing the data, you can easily uphold the normative aspect if it is needed, by using the methods of Normative Case Study , Normative Comparison , Normative Classification, and Normative Study of Development, so there is no need to mix up the sampling procedure with normative considerations.

2.10 RANDOM VERSUS NON-RANDOM SAMPLES

In statistics, a sample is a subset of a population. Usually, the population is very large, making a complete enumeration of all the values in the population impractical or impossible. The sample represents a subset of manageable size; the sample size is the number of units in the sample. Samples are collected and statistics are calculated from the samples so that one can make inferences or extrapolations from the sample to the population. This process of collecting information from a sample is referred to as sampling.

Samples are selected in such a way as to avoid presenting a biased view of the population. The sample will be unrepresentative of the population if certain members of the population are excluded from any possible sample. For example, if a researcher is interested in the drug-usage patterns among teenagers, but collects the sample from schools, the sample is biased because it excludes teenagers not in school for a variety of reasons, such as lack of funds to attend or schooled at home. Biases may also occur if some members of the population are more likely or less likely to be included in the sample than other members of the population for a reason other than the sample design. So the sample collected from schools is also biased because students who miss a lot of school days because of a chronic illness will be less likely to be selected than students who attend regularly.

The best way to avoid a biased or unrepresentative sample, and thus to obtain a representative sample of the population, is to select a random sample, also known as a probability sample. A random sample is defined as a sample in which every individual member of the population has a non-zero probability of being selected as part of the sample. In a simple random sample, every individual member of the population has the same probability of being selected as every other individual member. Other types of random samples fall under the category of complex sample design.

A sample that is not random is called a non-random sample or a non-probability sample. Some examples of non-random samples are convenience samples, judgment samples, purposive samples, quota samples, and snowball samples.

2.11 SIMPLE AND STRATIFIED RANDOM SAMPLING

A simple random sample gives each member of the population an equal chance of being chosen. It is not a haphazard sample as some people think! One way of achieving a simple random sample is to number each element in the sampling frame (e.g. give everyone on the Electoral register a number) and then use random numbers to select the required sample.

Random numbers can be obtained using your calculator, a spreadsheet, printed tables of random numbers, or by the more traditional methods of drawing slips of paper from a hat, tossing coins or rolling dice.

The optimum sample is the one which maximises precision per unit cost, and by this criterion simple random sampling can often be bettered by other methods.

In statistics, a simple random sample is a subset of individuals (a sample) chosen from a larger set (a population). Each individual is chosen randomly and entirely by chance, such that each individual has the same probability of being chosen at any stage during the sampling process, and each subset of k individuals has the same probability of being chosen for the sample as any other subset of k individuals. This process and technique is known as simple random sampling, and should not be confused with Random Sampling.

In small populations and often in large ones, such sampling is typically done "without replacement" ('SRSWOR'), i.e., one deliberately avoids choosing any member of the population more than once. Although simple random sampling can be conducted with replacement instead, this is less common and would normally be described more fully as simple random sampling with replacement ('SRSWR'). Sampling done without replacement is no longer independent, but still satisfies exchangeability, hence many results still hold. Further, for a small sample from a large population, sampling without replacement is approximately the same as sampling with replacement, since the odds of choosing the same sample twice is low.

An unbiased random selection of individuals is important so that in the long run, the sample represents the population. However, this does not guarantee that a particular sample is a perfect representation of the population. Simple random sampling merely allows one to draw externally valid conclusions about the entire population based on the sample.

Conceptually, simple random sampling is the simplest of the probability sampling techniques. It requires a complete sampling frame, which may not be available or feasible to construct for large populations. Even if a complete frame is available, more efficient approaches may be possible if other useful information is available about the units in the population.

Advantages are that it is free of classification error, and it requires minimum advance knowledge of the population other than the frame. Its simplicity also makes it relatively easy to interpret data collected via SRS. For these reasons, simple random sampling best suits situations where not much information is available about the population and data collection can be efficiently conducted on randomly distributed items, or where the cost of sampling is small enough to make efficiency less important than simplicity. If these conditions are not true, stratified sampling or cluster sampling may be a better choice.

Advantages

- ideal for statistical purposes.

Disadvantages

- hard to achieve in practice;

- requires an accurate list of the whole population;
- expensive to conduct as those sampled may be scattered over a wide area.

NOTES

STRATIFIED RANDOM SAMPLING

In statistics, stratified sampling is a method of sampling from a population.

When sub-populations vary considerably, it is advantageous to sample each subpopulation (stratum) independently. Stratification is the process of grouping members of the population into relatively homogeneous subgroups before sampling. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. Then random or systematic sampling is applied within each stratum. This often improves the representativeness of the sample by reducing sampling error. It can produce a weighted mean that has less variability than the arithmetic mean of a simple random sample of the population.

Stratified Sampling Strategies

1. Proportionate allocation uses a sampling fraction in each of the strata that is proportional to that of the total population. If the population consists of 60% in the male stratum and 40% in the female stratum, then the relative size of the two samples (three males, two females) should reflect this proportion.
2. Optimum allocation (or Disproportionate allocation) - Each stratum is proportionate to the standard deviation of the distribution of the variable. Larger samples are taken in the strata with the greatest variability to generate the least possible sampling variance.

A real-world example of using stratified sampling would be for a political survey. If the respondents needed to reflect the diversity of the population, the researcher would specifically seek to include participants of various minority groups such as race or religion, based on their proportionality to the total population as mentioned above. A stratified survey could thus claim to be more representative of the population than a survey of simple random sampling or systematic sampling.

Similarly, if population density varies greatly within a region, stratified sampling will ensure that estimates can be made with equal accuracy in different parts of the region, and that comparisons of sub-regions can be made with equal statistical power. For example, in Ontario a survey taken throughout the province might use a larger sampling fraction in the less populated north, since the disparity in population between north and south is so great that a sampling fraction based on the provincial sample as a whole might result in the collection of only a handful of data from the north.

Randomized stratification can also be used to improve population representativeness in a study.

Disadvantages

It is not useful when there are no similar subgroups. It cannot be used when amount of data in subgroups is not equal but total data in a subgroup are of equal importance as it gives more importance to subgroups with more data.

2.12 SOURCES OF DATA: PRIMARY AND SECONDARY DATA

Data refers to information or facts usually collected as the result of experience, observation or experiment or premises. Data may consist of numbers, words, or images, particularly as measurements or observations of a set of variables. Data are often viewed as a lowest level of abstraction from which information and knowledge are derived.

You might be reading a newspaper regularly. Almost every newspaper gives the minimum and the maximum temperatures recorded in the city on the previous day. It also indicates the rainfall recorded, and the time of sunrise and sunset. In your school, you regularly take attendance of children and record it in a register. For a patient, the doctor advises recording of the body temperature of the patient at regular intervals.

If you record the minimum and maximum temperature, or rainfall, or the time of sunrise and sunset, or attendance of children, or the body temperature of the patient, over a period of time, what you are recording is known as data. Here, you are recording the data of minimum and maximum temperature of the city, data of rainfall, data for the time of sunrise and sunset, and the data pertaining to the attendance of children.

As an example, the class-wise attendance of students, in a school, is as recorded in Table 2.2.

Table 2.2. Class-wise Attendance of Students

Class	No. of Students Present
VI	42
VII	40
VIII	41
IX	35
X	36
XI	32
XI1	30
Total	256

NOTES

Table 2.2 gives the data for class-wise attendance of students. Here the data comprise 7 observations in all. These observations are, attendance for class VI, VII, and so on. So, data refers to the set of observations, values, elements or objects hder consideration.

The complete set of all possible elements or objects is called a population. Each of the elements is called a piece of data. Data also refers to the known facts or things used as basis for inference or reckoning facts, information, material to be processed or stored.

Nature of Data

For understanding the nature of data, it becomes necessary to study about the various forms of data, as shown below :

- Qualitative and Quantitative Data
- Continuous and Discrete Data
- Primary and Secondary Data

Qualitative and Quantitative Data

Let us consider a set of data given in Table 2.3.

Table 2.3. Management-wise Number of Schools

Management	No. of Schools
Government	4
Local Body	8
Private Aided	10
Private Unaided	2
Total	24

In Table 2.3, number of schools have been shown according to the management of schools. So the schools have been classified into 4 categories, namely, Government Schools, Local Body Schools, Private Aided Schools and Private Unaided Schools. A given school belongs to any one of the four categories. Such data is shown as Categorical or Qualitative Data. Here the category or the quality referred to is management. Thus categorical or qualitative data result from information which has been classified into categories. Such categories are listed alphabetically or in order of decreasing frequencies or in some other conventional way. Each piece of data clearly belongs to one classification or category.

We frequently come across categorical or qualitative data in the form of schools categorised according to Boys, Girls and Co-educational; Students' Enrolment categorised according to SC, ST, OBC and 'Others'; number of persons employed in various categories of occupations, and so on.

Let us consider another set of data given in Table 2.4.

NOTES

Table 2.4. Number of Schools according to Enrolment

Enrolment	No. of Schools
Upto 50	6
51 - 100	15
101 - 200	12
201 - 300	8
Above 300	4
Total	45

In Table 2.4, number of schools have been shown according to the enrolment of students in the school. Schools with enrolment varying in a specified range are grouped together, e.g. there are 15 schools where the students enrolled are any number between 51 and 100. As the grouping is based on numbers, such data are called Numerical or Quantitative Data. Thus, numerical or quantitative data *result from counting or measuring.* We frequently come across numerical data in newspapers, advertisements etc. related to the temperature of the cities, cricket averages, incomes, expenditures and so on.

Continuous and Discrete Data

Numerical or quantitative data may be continuous or discrete depending on the nature of the elements or objects being observed.

Let us consider the Table 2.5 depicting the heights of students of a class.

Table 2.5. Heights of Students of a Class

Height	No. of Students
4'8"- 4' 10"	2
4'10" - 5'0"	2
5'0" - 5'2"	5
5'2" - 5'4"	8
5'4" - 5'6"	12
5'6" - 5'8"	10
5'8"- 5'10"	2
Total	41

Table 2.5 gives the data pertaining to the heights of students of a class. Here the element under observation is the height of the students. The height varies from 4' 8" to 5' 10". The height of an individual may be anywhere from 4' 8" to 5'10". Two students may vary by almost zero inch height. Even if we take two adjacent points, say 4' 8.00" and 4' 8.01" there may be several values between the two points. Such data are called Continuous Data, as the height is continuous. Continuous Data arise from the measurement of continuous attributes or vdriables, in which individual may differ by amounts just approaching zero.

Weights and heights of children; temperature of a body; intelligence and achievement level of students, etc. are the examples of continuous data.

Let us consider Table 2.3 showing the number of students enrolled and the number of schools according to enrolment. Let us,consider the enrolment of 2 schools as 60 and 61. Now in between 60 and 61, there cannot be any number, as the enrolment will always be in whole numbers. Thus there is a gap of one unit from 60 to 61. Such data, where the elements being observed have gaps are called Discrete Data.

Discbete Data are characterised by ,gaps in the scale, for which no real values may ever be found. Such data are usually expressed in whole numbers. The size of a family, enrolment of children, number of books etc. are the examples of discrete data. Generally data arising from meagurement are continuous, while data arising from counting or arbitrary classification are discrete.

The achievement scores of students, though presented in discrete form may be considered to constitute continuous data, since a score of 24 represents any point between 23.5 and 24.5. Actually achievement is a continuous attribute or variable.

All peasurements of continuous attributes are approximate in character and as such do not prodide a basis for distinguishing between continuous and discrete data. The distinction is made on the basis of variable being measured. 'Height' is a continuous variable but number of children would give discrete data.

Primary and Secondary Data

The data collected by or on behalf of the person or people who are going to make use of the data refers to primary data. For example, the attendance of children, the result of examinations conducted by you are primary data. If you contact the parents of the children and ask about their educational qualifications to relate them to the performance of the children, this also gives primary data. Actually, when an individual personally collects data or information pertaining to an event, a definite plan or design, it refers to primary data.

Sometimes an investigator may use the data already collected by you, such as the school attendance of children, or performance of students in various subjects. etc, for his/her study, then the data are secondary data. The data used by a person or people other than the people by whom or for whom the data were collected refers to secondary data. For many reasons we may have to use secondary data, which should be used carefully, since the data could have beeb collected with a purpose different from that of the investigator and may lose some detail or may not be fully relevant. For using secondary data, it is always useful to know :

(a) how the data have been collected and processed;

(*b*) the accuracy of data;

(*c*) how far the data have been summarised;

(*d*) how comparable the data are with other tabulations; and

(*e*) how to interpret the data, especially when figures collected for one purpose are used for another purpose.

Secondary Data

In research, Secondary data is collecting and possibly processing data by people other than the researcher in question. Common sources of secondary data for social science include censuses, large surveys, and organizational records. In sociology primary data is data you have collected yourself and secondary data is data you have gathered from primary sources to create new research. In terms of historical research, these two terms have different meanings. A primary source is a book or set of archival records. A secondary source is a summary of a book or set of records.

Advantages to the secondary data collection method are - (1) it saves time that would otherwise be spent collecting data, (2) provides a larger database (usually) than what would be possible to collect on ones own However there are disadvantages to the fact that the researcher cannot personally check the data so it's reliability may be questioned.

Secondary Data Analysis

There are two different types of sources that need to be established in order to conduct a good analysis. The first type is a primary source which is the initial material that is collected during the research process. Primary data is the data that the researcher is collecting themselves using methods such as surveys,direct observations, interviews, as well as logs(objective data sources). Primary data is a reliable way to collect data because the researcher will know where it came from and how it was collected and analyzed since they did it themselves. Secondary sources on the other hand are sources that are based upon the data that was collected from the primary source. Secondary sources take the role of analyzing, explaining, and combining the information from the primary source with additional information.

Secondary data analysis is commonly known as second-hand analysis. It is simply the analysis of preexisting data in a different way or to answer a different question than originally intended. Secondary data analysis utilizes the data that was collected by someone else in order to further a study that you are interested in completing.

In contrast to secondary data, primary data comes from observations made by the researchers themselves. This often creates credibility issues that do not arise with secondary data.

NOTES

Combining Data with Secondary Data

Where It's Used

For what different purposes can data from archives be used? The first and simplest case would be for descriptive purposes, such as a phone book. A particular contribution of the data archives can be made to comparative research, both, across nations and over time. In the early years of data archives, when secondary analysis was not yet a popular research strategy, the idea of comparative research based on archival data was promoted in conferences already some 40 years ago. In the first case this would allow for comparative analysis over time, in the second for comparative analysis across societies or nations. Therefore, the design of comparative surveys is crucial for making empirical knowledge cumulative over space and time.

Combing Data From a Different Source with Different Time Periods

Equally important are longitudinal studies which can be compiled over time. For example, in a research project on "Attitudes Towards Technology" it is of crucial importance to include data collected in the fifties and sixties in order to answer the research question whether potential threats from new technologies have decreased the level of technology acceptance or whether tendencies to reject new developments concentrate on particular technologies only, and if so, under what circumstances.

Combining Existing Secondary Data Sources with New Primary Data Sources

Imagine that we could get hold of a good collection of surveys taken in earlier years, such as detailed studies about changes going on in this phase and hopefully additional studies in the years to come. Analyzing this data base over time could give us a good picture of what changes actually have taken place in the orientation of the population and of the extent to which new technical concepts did have an impact on subgroups of the population. Furthermore, data archives can help to prepare studies on change over time by monitoring what questions have been asked in earlier years and alerting principal investigators to important questions which should be repeated in planned research projects. Actually, data archives should consider including funds in their budgets which allow them to collect data for relevant questions in order to avoid interruptions in important time series.

Technical Challenges in Combining Data Sets

A number of methodological and technical requirements have to be observed and should be implemented rigorously. Just to mention the most important: Some methodologists require that the questions should be functionally equivalent, whereas others claim that the question texts must be phrased identically. Frequently, it is not the linguistic identity which matters. Sometimes it is much more important, whether the questions are understood by the respondent in the same way. Thus, a thermometer or scale used as a representation for intensity of attitudes in the more developed societies may be replaced by a ladder in less developed societies. Both, thermometer and ladder, would still measure the same dimension in the conceptual world of the respective respondents. A second requirement would be comparability of samples, thus, a cross-national representative random sample would be hard to compare with the local quota sample in one community in a different nation. Several other factors have to be controlled as well, in particular contextual influences at the time of field work or political or environmental events, which are related to the topic of the research.

Collecting, Reviewing, and Analyzing Secondary Data

The Design and Purpose of Research

Secondary data analysis consists of collecting data that was compiled through research by another person and using that data to get a better understanding of a concept. A good way to begin your research using secondary data that you are collecting to further support your concept is to clearly define the goals of your research and the design that you anticipate using. An important thing to remember when defining your plan is to ensure that you have established what kind of data you plan on using for your research and the exact goal. Establishing what type of research design is an important component. In terms of using secondary data for research it helps to create an outline of what the final product will look like consisting of all the types of data to be used along with a list of sources that were used to compile the research. In order to use secondary data three steps must be completed:

1. locate the data
2. evaluate the data
3. verify the data

Locating the data can be easily done with the advancements of searching sources online. However, people need to be aware of the details when searching online since pages can be out of date or poorly put together. Therefore, use caution and pay attention to whether it is a reliable data source online and check when the last update was. To evaluate the data a researcher must carefully examine the secondary data they are considering to ensure that it meets their needs and purpose of study. The person must look at the population and what the sample strategy and type were. It is also important to look at when the data was collected, how it was collected, how it was coded and edited, along with the operational definitions of measures that were used. Finally, the data must be verified to ensure good quality material to be used in new research.

NOTES

Determining the Types of Data and Information Needed to Conduct Analysis

Data and information collection for secondary data analysis will depend entirely upon the subject that is central to the focal point of the study. The purpose of conducting secondary data analysis is to further develop an improved understanding of the subject matter at hand. Some important types of data and information that should be collected and summarized include demographic information, information gathered by government agencies (i.e. the Census), and social science surveys. There is also the possibility of reanalyzing data that was collected in experimental studies or data collected with qualitative measures that can be applied in secondary data analysis. The most important component is to ensure that the information and data being collected needs to relate to the subject of study.

Determine the Quality of Sources of Data

In secondary data analysis, most individuals who do not have much experience in research training or technical expertise can be trained accordingly. However, this advantage is not without difficulty as the individual must be able to judge the quality of the data or information that has been gathered. These key tips will assist you in assessing the quality of the data: Determine the original purpose of the data collection, attempt to discover the credentials of the source(s) or author(s) of the information, consider if the document is a primary or secondary source, verify that the source well-referenced, and finally find out the: date of the publication; the intended audience, and coverage of the report or document.

Challenges of Secondary Data Analysis

Advantages

Using secondary data can allow for the analyses of social processes in what would otherwise be inaccessible settings. It also saves time and money since the work has already been done to collect the data. That lets the researcher avoid problems with the data collection process. Using someone else's data can also facilitate a comparison with other data samples and allow multiple sets of data to be combined. There is also the chance that other variables could be included, resulting in a more diverse sample than would have been feasible before.

Disadvantages

There are several things to take into consideration when using preexisting data. Secondary data does not permit the progression from formulating a research question to designing methods to answer that question. It is also not feasible for a secondary data analyst to engage in the habitual process of making observations and developing concepts. These limitations hinder the ability of the researcher to focus on the original research question. Data quality is always a concern because it's source may not be trusted. Even data from official records may be bad because the data is only as good as the records themselves. There are six questions that a secondary analyst should be able to answer about the data they wish to analyze.

1. What were the agency's or researcher's goals when collecting the data?
2. What data was collected and what is it supposed to measure?
3. When was the data collected?
4. What methods were used? Who was responsible and are they available for questions?
5. How is the data organized?
6. What information is known about the success of that data collection? How consistent is the data with data from other sources?

2.13 DATA COLLECTION

Data collection is a term used to describe a process of preparing and collecting data - for example as part of a process improvement or similar project. The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, to pass information on to others. Primarily, data is collected to provide information regarding a specific topic.

Data collection usually takes place early on in an improvement project, and is often formalised through a data collection plan which often contains the following activity.

- Pre collection activity – Agree goals, target data, definitions, methods
- Collection – data collection

Present Findings – usually involves some form of sorting analysis and/or presentation.

Prior to any data collection, pre-collection activity is one of the most crucial steps in the process. It is often discovered too late that the value of their interview information is discounted as a consequence of poor sampling of both questions and informants and poor elicitation techniques. After pre-collection activity is fully completed, data collection in the field, whether by interviewing or other methods, can be carried out in a structured, systematic and scientific way.

A formal data collection process is necessary as it ensures that data gathered is both defined and accurate and that subsequent decisions based on arguments embodied in the findings are valid. The process provides both a baseline from which to measure from and in certain cases a target on what to improve.

Types of Data Collection

NOTES

- By mail questionnaires;
- By personal interview.

Other main types of collection include census, sample survey, and administrative by-product and each with their respective advantages and disadvantages. A census refers to data collection about everyone or everything in a group or population and has advantages, such as accuracy and detail and disadvantages, such as cost and time. A sample survey is a data collection method that includes only part of the total population and has advantages, such as cost and time and disadvantages, such as accuracy and detail. Administrative by-product data is collected as a byproduct of an organization's day-to-day operations and has advantages, such as accuracy, time simplicity and disadvantages, such as no flexibility and lack of control.

Primary Data Collection Methods

In primary data collection, you collect the data yourself using methods such as interviews and questionnaires. The key point here is that the data you collect is unique to you and your research and, until you publish, no one else has access to it.

There are many methods of collecting primary data and the main methods include:

- questionnaires
- interviews
- focus group interviews
- observation
- case-studies
- diaries
- critical incidents
- portfolios.

Questionnaires

Questionnaires are a popular means of collecting data, but are difficult to design and often require many rewrites before an acceptable questionnaire is produced.

Advantages:—

- Can be used as a method in its own right or as a basis for interviewing or a telephone survey.
- Can be posted, e-mailed or faxed.
- Can cover a large number of people or organisations.

NOTES

- Wide geographic coverage.
- Relatively cheap.
- No prior arrangements are needed.
- Avoids embarrassment on the part of the respondent.
- Respondent can consider responses.
- Possible anonymity of respondent.
- No interviewer bias.

Disadvantages:—

- Design problems.
- Questions have to be relatively simple.
- Historically low response rate (although inducements may help).
- Time delay whilst waiting for responses to be returned.
- Require a return deadline.
- Several reminders may be required.
- Assumes no literacy problems.
- No control over who completes it.
- Not possible to give assistance if required.
- Problems with incomplete questionnaires.
- Replies not spontaneous and independent of each other.
- Respondent can read all questions beforehand and then decide whether to complete or not. For example, perhaps because it is too long, too complex, uninteresting, or too personal.

The Response Process

While the process is simple and straight-forward, there are many opportunities for error.

- The question must be read.
- The question must be understood.
- The respondent must create a response.
- This response must be translated into the categories or values present for the question.

Fundamental Concerns

We are asking about a person's knowledge, attitudes, beliefs, feelings, motivations, anticipations, future plans or past behavior. What can we do to motivate people to respond and to respond truthfully? For example, one older study found that about 20 percent of those without library cards claimed to have one when asked. There is a strong tendency to give answers that are socially desirable, make the respondent look good, or that will please the researcher.

While we begin with an assumption of truthfulness, it may be useful to use variants of the same question to capture more of reality.

Another concern is whether or not the respondent knows enough to provide a meaningful answer. Some one who has never used the teen collection may find it difficult to indicate if it is valued.

Validity is the degree to which we are measuring what we need to measure. The questionnaire should gather valid responses. Reliability is the degree to which we receive the same measurement over time. Would we receive the same response if the respondent had answered the questionnaire earlier or later?

Cost

At the beginning and throughout the process, you need to consider these questions:

1. How much money will be needed to collect this data?
2. How much time will be needed to collect this data?
3. Is cost-sharing possible and will that be helped by sponsorship or endorsement?
4. How many completed questionnaires will be needed and what response rate does that require?

If possible, begin with another's questionnaire, especially if you are doing a use and user survey [do receive permission first]. This is less expensive, but there are other advantages. Others should have validated their instrument. You will be able to compare your findings with theirs and built upon previous generalizations. Knowledge can cumulate. The ERIC data base includes a variety of questionnaires used in previous use and user surveys. Library Literature is also quite useful via the --survey tag. Ideally, you would keep changes to a minimum to facilitate comparison, but you may build upon an existing instrument by adding additional questions or making essential changes.

The Process

Ordinarily, there are six steps:

1. Identifying what information is needed
2. Deciding what sort of questionnaire to use
3. Creating the first draft
4. Editing and revising
5. Pre testing and revising
6. Specifying procedures for its use.

A few deep thoughts are needed at the beginning:

- Are you certain that this study is worth doing?
- Are your research questions and key variables clearly identified?

- What answers do you need to have?
- What sort of questions are likely to gather those answers?
- What problems are likely to be encountered in getting a good response?

The number of questions must be limited to insure a good response. Response rate declines rapidly as the number of questions, especially those that require time and thought are added. Questions can usually be divided into two categories: (1) absolutely necessary and (2) interesting. Be certain that the information to be gathered is not available elsewhere, i.e. in census data or another report. Respondents are much less likely to respond to a question if they feel the answer is readily available.

Questions will need to be placed in a logical sequence.

Type of Information Sought

Typically, information sought falls into four categories:

1. Attitudes or what people see/understand about certain things
2. Beliefs or what people think is true [more strongly felt than attitudes]
3. Behavior or what people do
4. Attributes or what people are.

Attitude questions ask people to indicate if they favor or oppose, if they prefer or not, should or should not, right versus wrong, desirable versus undesirable. These questions require sensitive, thoughtful wording.

Belief questions ask people if something is true or false, correct or incorrect, accurate or inaccurate.

Behavior questions ask people what they have done, what they do, or what they plan to do.

Attribute or demographic questions ask about age, income, education, and the like.

Question Type

Open-ended Questions

Open-ended questions provide no answer choices. They are easy to ask and allow for a wide variety of responses, including the creative and unusual. Open-ended questions are especially useful when you don't know the likely values or can't anticipate how the respondent will respond. The information gathered by open-ended questions could then be used to develop appropriate close-ended questions for another questionnaire.

These questions force the respondent to think and allow the respondent to clarify and explain a response. If the respondent takes needed time and makes the effort, responses can be illuminating and yield much useful information.

The response rate will be lower because the blank space is demanding and intimidating, especially for those who don't like to write. Illegible handwriting may be a problem.

Since responses are not really ordered, analysis requires considerable time and effort. It may be difficult to measure and classify responses. Responses may be off base because there is inadequate guidance from the instrument itself.

Close-ended Questions

These questions provide specific answer choices although there may be an "other" value with brief space for adding an additional value. With close-ended questions, there is always the possibility that the right question will not be asked and valuable information will not be gathered. GIGO certainly applies here.

Ordered

Ordered close-ended questions require respondents to select a particular response. The responses are easily selected. They take little time, at least by most respondents.

These questions require well defined variables and values. They work best when there are a small number of reasonable answer possibilities.

Unordered

Unordered questions ask respondents to rank values and are useful for identifying priorities. Requires well defined variables and values. Unordered close-ended questions are not reliable if there are more than five values. Most respondents find ranking after one or two values to be difficult.

Criteria for ranking must be clearly identified and the order of the values must make sense to the respondent.

Partially Close-Ended

These questions may be ordered or not, but they do include the "other" option which adds some flexibility and provides the opportunity to add information not otherwise captured by the instrument.

More flexibility may mean better and more valid responses. New values may be selected from the leading "other" values, but this will make analysis more challenging.

Number of Values

Close-ended questions may be categorized by the number of values. Two value questions are dichotomous. These questions are easy to ask and are quickly answered. Analysis is straight-forward and quick.

However, two alternatives are usually not enough. Some times, respondents will select the first value so the ordering of the values has some impact. Each value must be exclusive. The researcher must know the notable alternatives.

Multiple choice questions [more than two values]are also easy to analyze, but do require more time and effort. These questions require more time and effort by the respondent. Typically, respondents tend to choose from the top or middle of the list.

2.14 INTERVIEWING AND OBSERVATION

Interviewing is a technique that is primarily used to gain an understanding of the underlying reasons and motivations for people's attitudes, preferences or behaviour. Interviews can be undertaken on a personal one-to-one basis or in a group. They can be conducted at work, at home, in the street or in a shopping centre, or some other agreed location.

Personal Interview

Advantages:

- Serious approach by respondent resulting in accurate information.
- Good response rate.
- Completed and immediate.
- Possible in-depth questions.
- Interviewer in control and can give help if there is a problem.
- Can investigate motives and feelings.
- Can use recording equipment.
- Characteristics of respondent assessed – tone of voice, facial expression, hesitation, etc.
- Can use props.
- If one interviewer used, uniformity of approach.
- Used to pilot other methods.

Disadvantages:

- Need to set up interviews.
- Time consuming.
- Geographic limitations.
- Can be expensive.
- Normally need a set of questions.
- Respondent bias – tendency to please or impress, create false personal image, or end interview quickly.
- Embarrassment possible if personal questions.
- Transcription and analysis can present problems – subjectivity.
- If many interviewers, training required.

NOTES

TYPES OF INTERVIEW

Structured

- Based on a carefully worded interview schedule.
- Frequently require short answers with the answers being ticked off.
- Useful when there are a lot of questions which are not particularly contentious or thought provoking.
- Respondent may become irritated by having to give over-simplified answers.

Semi-structured

The interview is focused by asking certain questions but with scope for the respondent to express him or herself at length.

Unstructured

This also called an in-depth interview. The interviewer begins by asking a general question. The interviewer then encourages the respondent to talk freely. The interviewer uses an unstructured format, the subsequent direction of the interview being determined by the respondent's initial reply. The interviewer then probes for elaboration – 'Why do you say that?' or, 'That's interesting, tell me more' or, 'Would you like to add anything else?' being typical probes.

The following section is a step-by-step guide to conducting an interview. You should remember that all situations are different and therefore you may need refinements to the approach.

Planning an Interview:

- List the areas in which you require information.
- Decide on type of interview.
- Transform areas into actual questions.
- Try them out on a friend or relative.
- Make an appointment with respondent(s) – discussing details of why and how long.
- Try and fix a venue and time when you will not be disturbed.

Conducting an Interview:

• Personally	–	arrive on time be smart smile employ good manners find a balance between friendliness and objectivity.
• At the start	–	introduce yourself re-confirm the purpose assure confidentiality – if relevant specify what will happen to the data.

• The questions	–	speak slowly in a soft, yet audible tone of voice control your body language know the questions and topic ask all the questions.
• Responses	–	recorded as you go on questionnaire written verbatim, but slow and time-consuming summarised by you taped – agree beforehand – have alternative method if not acceptable consider effect on respondent's answers proper equipment in good working order sufficient tapes and batteries minimum of background noise.
• At the end	–	ask if the respondent would like to give further details about anything or any questions about the research thank them.

Telephone Interview

This is an alternative form of interview to the personal, face-to-face interview.

Advantages:

- Relatively cheap.
- Quick.
- Can cover reasonably large numbers of people or organisations.
- Wide geographic coverage.
- High response rate – keep going till the required number.
- No waiting.
- Spontaneous response.
- Help can be given to the respondent.
- Can tape answers.

Disadvantages:

- Often connected with selling.
- Questionnaire required.
- Not everyone has a telephone.
- Repeat calls are inevitable – average 2.5 calls to get someone.
- Time is wasted.
- Straightforward questions are required.
- Respondent has little time to think.
- Cannot use visual aids.

- Can cause irritation.
- Good telephone manner is required.
- Question of authority.

Getting Started

Locate the respondent:

- Repeat calls may be necessary especially if you are trying to contact people in organisations where you may have to go through secretaries.
- You may not know an individual's name or title – so there is the possibility of interviewing the wrong person.
- You can send an advance letter informing the respondent that you will be telephoning. This can explain the purpose of the research.

Getting them to agree to take part:

- You need to state concisely the purpose of the call – scripted and similar to the introductory letter of a postal questionnaire.
- Respondents will normally listen to this introduction before they decide to co-operate or refuse.
- When contact is made respondents may have questions or raise objections about why they could not participate. You should be prepared for these.

Ensuring Quality

- Quality of questionnaire – follows the principles of questionnaire design. However, it must be easy to move through as you cannot have long silences on the telephone.
- Ability of interviewer – follows the principles of face-to-face interviewing.

Smooth Implementation

- Interview schedule – each interview schedule should have a cover page with number, name and address. The cover sheet should make provision to record which call it is, the date and time, the interviewer, the outcome of the call and space to note down specific times at which a call-back has been arranged. Space should be provided to record the final outcome of the call – was an interview refused, contact never made, number disconnected, etc.
- Procedure for call-backs – a system for call-backs needs to be implemented. Interview schedules should be sorted according to their status: weekday call-back, evening call-back, weekend call-back, specific time call-back.

Comparison of Postal, Telephone and Personal Interview Surveys

The table below compares the three common methods of postal, telephone and interview surveys – it might help you to decide which one to use.

NOTES

Table 2.6. Comparison of the three common methods of surveys

	Postal survey	Telephone survey	Personal interview
Cost (assuming a good response rate)	Often lowest	Usually in-between	Usually highest
Ability to probe	No personal contact or observation	Some chance for gathering additional data through elaboration on questions, but no personal observation	Greatest opportunity for observation, building rapport, and additional probing
Respondent ability to complete at own convenience	Yes	Perhaps, but usually no	Perhaps, if interview time is prearranged with respondent
Interview bias	No chance	Some, perhaps due to voice inflection	Greatest chance
Ability to decide who actually responds to the questions	Least	Some	Greatest
Impersonality	Greatest	Some due to lack of face-to-face contact	Least
Complex questions	Least suitable	Somewhat suitable	More suitable
Visual aids	Little opportunity	No opportunity	Greatest opportunity
Potential negative respondent reaction	'Junk mail'	'Junk calls'	Invasion of privacy
Interviewer control over interview environment	Least	Some in selection of time to call	Greatest
Time lag between soliciting and receiving response	Greatest	Least	May be considerable if a large area involved
Suitable types of questions	Simple, mostly dichotomous	Some opportunity for open-ended	Greatest opportunity

	(yes/no) and multiple choice	questions especially if interview is recorded	for open-ended questions
Requirement for technical skills in conducting interview	Least	Medium	Greatest
Response rate	Low	Usually high	High

Focus Group Interviews

A focus group is an interview conducted by a trained moderator in a non-structured and natural manner with a small group of respondents. The moderator leads the discussion. The main purpose of focus groups is to gain insights by listening to a group of people from the appropriate target market talk about specific issues of interest.

Observation

Observation involves recording the behavioural patterns of people, objects and events in a systematic manner. Observational methods may be:

- structured or unstructured
- disguised or undisguised
- natural or contrived
- personal
- mechanical
- non-participant
- participant, with the participant taking a number of different roles.

Structured or Unstructured

In structured observation, the researcher specifies in detail what is to be observed and how the measurements are to be recorded. It is appropriate when the problem is clearly defined and the information needed is specified.

In unstructured observation, the researcher monitors all aspects of the phenomenon that seem relevant. It is appropriate when the problem has yet to be formulated precisely and flexibility is needed in observation to identify key components of the problem and to develop hypotheses. The potential for bias is high. Observation findings should be treated as hypotheses to be tested rather than as conclusive findings.

Disguised or Undisguised

In disguised observation, respondents are unaware they are being observed and thus behave naturally. Disguise is achieved, for example, by hiding, or using hidden equipment or people disguised as shoppers.

In undisguised observation, respondents are aware they are being observed. There is a danger of the Hawthorne effect – people behave differently when being observed.

Natural or Contrived

Natural observation involves observing behaviour as it takes place in the environment, for example, eating hamburgers in a fast food outlet.

In contrived observation, the respondents' behaviour is observed in an artificial environment, for example, a food tasting session.

Personal

In personal observation, a researcher observes actual behaviour as it occurs. The observer may or may not normally attempt to control or manipulate the phenomenon being observed. The observer merely records what takes place.

Mechanical

Mechanical devices (video, closed circuit television) record what is being observed. These devices may or may not require the respondent's direct participation. They are used for continuously recording on-going behaviour.

Non-participant

The observer does not normally question or communicate with the people being observed. He or she does not participate.

Participant

In participant observation, the researcher becomes, or is, part of the group that is being investigated. Participant observation has its roots in ethnographic studies (study of man and races) where researchers would live in tribal villages, attempting to understand the customs and practices of that culture. It has a very extensive literature, particularly in sociology (development, nature and laws of human society) and anthropology (physiological and psychological study of man). Organisations can be viewed as 'tribes' with their own customs and practices.

The role of the participant observer is not simple. There are different ways of classifying the role:

- Researcher as employee.
- Researcher as an explicit role.
- Interrupted involvement.
- Observation alone.

Researcher as Employee

The researcher works within the organisation alongside other employees, effectively as one of them. The role of the researcher may or may not be explicit and this will have implications for the extent to which he or she will be able to move around and gather information and perspectives from other sources. This role is appropriate when the researcher needs to become totally immersed and experience the work or situation at first hand.

There are a number of dilemmas. Do you tell management and the unions? Friendships may compromise the research. What are the ethics of the process? Can anonymity be maintained? Skill and competence to undertake the work may be required. The research may be over a long period of time.

Researcher as an Explicit Role

The researcher is present every day over a period of time, but entry is negotiated in advance with management and preferably with employees as well. The individual is quite clearly in the role of a researcher who can move around, observe, interview and participate in the work as appropriate. This type of role is the most favoured, as it provides many of the insights that the complete observer would gain, whilst offering much greater flexibility without the ethical problems that deception entails.

Interrupted Involvement

The researcher is present sporadically over a period of time, for example, moving in and out of the organisation to deal with other work or to conduct interviews with, or observations of, different people across a number of different organisations. It rarely involves much participation in the work.

Observation Alone

The observer role is often disliked by employees since it appears to be 'eavesdropping'. The inevitable detachment prevents the degree of trust and friendship forming between the researcher and respondent, which is an important component in other methods.

Choice of Roles

The role adopted depends on the following:

- Purpose of the research: Does the research require continued longitudinal involvement (long period of time), or will in-depth interviews, for example, conducted over time give the type of insights required?
- Cost of the research: To what extent can the researcher afford to be committed for extended periods of time? Are there additional costs such as training?
- The extent to which access can be gained: Gaining access where the role of the researcher is either explicit or covert can be difficult, and may take time.
- The extent to which the researcher would be comfortable in the role: If the researcher intends to keep his identity concealed, will he or she also

feel able to develop the type of trusting relationships that are important? What are the ethical issues?

- The amount of time the researcher has at his disposal: Some methods involve a considerable amount of time. If time is a problem alternate approaches will have to be sought.

2.15 DATA PROCESSING

Data processing is the act of handling or manipulating data in some fashion. Regardless of the activities involved in it, processing tries to assign meaning to data. Thus, the ultimate goal of processing is to transform data into information. Data processing is the process through which facts and figures are collected, assigned meaning, communicated to others and retained for future use. Hence we can define data processing as a series of actions or operations that converts data into useful information. We use the term 'data processing system' to include the resources that are used to accomplish the processing of data.

Data Processing Activities

As discussed above, data processing consists of those activities which are necessary to transform data into information. Man has in course of time devised certain tools to help him in processing data. These include manual tools such as pencil and paper, mechanical tools such as filing cabinets, electromechanical tools such as adding machines and typewriters, and electronic tools such as calculators and computers.

Many people immediately associate data processing with computers. As stated above, a computer is not the only tool used for data processing, it can be done without computers also. However, computers have outperformed people for certain tasks. There are some other tasks for which computer is a poor substitute for human skill and intelligence.

Summary

- In the social sciences, quantitative research refers to the systematic empirical investigation of quantitative properties and phenomena and their relationships. The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena.
- Qualitative method is a method of inquiry appropriated in many different academic disciplines, traditionally in the social sciences, but also in market research and further contexts.
- Sampling is that part of statistical practice concerned with the selection of an unbiased or random subset of individual observations within a

population of individuals intended to yield some knowledge about the population of concern, especially for the purposes of making predictions based on statistical inference. Sampling is an important aspect of data collection.

- A sample is a subject chosen from a population for investigation. A random sample is one chosen by a method involving an unpredictable component.
- Non-random (or "nonprobability") samples are selected by any kind of procedure that does not give all cases in the population equal chances to fall into the sample.
- A simple random sample gives each member of the population an equal chance of being chosen.
- Data refers to information or facts usually collected as the result of experience, observation or experiment or premises. Data may consist of numbers, words, or images, particularly as measurements or observations of a set of variables.
- Data collection is a term used to describe a process of preparing and collecting data - for example as part of a process improvement or similar project.

Review Questions

1. Differentiate between quantitative and qualitative method.
2. What is sampling? State the elements of sampling.
3. Distinguish between simple and stratified random sampling.
4. What are the main sources of data?
5. Why is appropriate method of data collection essential? Explain.
6. State the importance of interviewing and observation.

Further Readings

- Ahuja Ram, *Research Methods*, Publisher: Rawat Publication, Jaipur.
- Bhanwar Lal Garg, Renu Kavdia, Sulochana Agrawal and Umesh Kumar Agrawal, *An Introduction to Research Methodology*, RBSA Pub, 2002, New Delhi.
- C M Chaudhary, *Research Methodology*.
- Kothari, C. R.; *Research Methodology: Methods and Techniques*; Publisher: New Age International.

UNIT – III

DATA PRESENTATION

NOTES

STRUCTURE

LEARNING OBJECTIVES

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

- understand the method of graphical and tabular presentation of data;
- state the concept of various typres of graphs such as pie-graph, histograms, polygons and line graphs;
- explain the concept of unvaried and multivariate tables.

3.1 INTRODUCTION

In general, most evaluations conducted by local programs would lend themselves to descriptive analysis of data. Descriptive analysis is a way of summarizing and aggregating results from groups. If an evaluation has been conducted which employs a control group, or measures changes in program participants over time, then it might be appropriate to employ inferential analysis in which a decision is made about whether the particular results of the study are "real". More emphasis will be placed on descriptive analysis in this fact sheet.

Many reports rely on narrative information to present most, if not all, of the necessary information. Narrative information may be presented in three ways: standard writing style; tables; and/or, figures, diagrams, maps, and charts.

Standard writing style, that is, the use of sentences and paragraphs, is often the best way to present information, especially to audiences that are not accustomed to working with charts, graphs, tables, numbers, etc. It is the only way to present information such as examples and explanations. If standard writing style is used to summarize the results of open ended questions ("What do you like most about the program?"), it is often useful to give some indication of how often a particular response was given.

Tables represent narrative or numerical information in tabular fashion. A table arranges information in rows or columns, so that data elements may be referred to easily. They provide a clear and succinct way to present data, and are often more simple and understandable than standard writing style. They also facilitate the interpretation of data.

Figures, diagrams, maps and charts present verbal information visually. They often describe information more clearly than several paragraphs of description. Common forms of figures are: flow charts; organization charts; GANT charts; and/or maps.

- Flow charts are particularly useful for presenting relationships and/or describing the sequence of events and the location and result of decisions.
- Organization charts are useful for presenting the chain of responsibility in a program.
- GANT charts list a set of tasks. They indicate the time each task is to be performed and by whom.
- Maps visually describe certain geographical areas. They are useful in describing different conditions for individual geographical areas.

Data refers to information or facts usually collected as the result of experience, observation or experiment or premises. Data may consist of numbers, words, or images, particularly as measurements or observations of a set of variables. Data are often viewed as a lowest level of abstraction from which information and knowledge are derived.

You might be reading a newspaper regularly. Almost every newspaper gives the minimum and the maximum temperatures recorded in the city on the previous day. It also indicates the rainfall recorded, and the time of sunrise and sunset. In your school, you regularly take attendance of children and record it in a register. For a patient, the doctor advises recording of the body temperature of the patient at regular intervals.

If you record the minimum and maximum temperature, or rainfall, or the time of sunrise and sunset, or attendance of children, or the body temperature of the patient, over a period of time, what you are recording is known as data. Here, you are recording the data of minimum and maximum temperature of the city, data of rainfall, data for the time of sunrise and sunset, and the data pertaining to

NOTES

the attendance of children. As an example, the class-wise attendance of students, in a school, is as recorded in Table 3.1.

Table 3.1. Class-wise Attendance of Students

Class	No. of Students Present
VI	42
VII	40
VIII	41
IX	35
X	36
XI	32
XI1	30
Total	256

Table 3.1 gives the data for class-wise attendance of students. Here the data comprise 7 observations in all. These observations are, attendance for class VI, VII, and so on. So, data refers to the set of observations, values, elements or objects hder consideration.

The complete set of all possible elements or objects is called a population. Each of the elements is called a piece of data. Data also refers to the known facts or things used as basis for inference or reckoning facts, information, material to be processed or stored.

Nature of Data

For understanding the nature of data, it becomes necessary to study about the various forms of data, as shown below :

- Qualitative and Quantitative Data
- Continuous and Discrete Data
- Primary and Secondary Data

Qualitative and Quantitative Data

Let us consider a set of data given in Table 3.2.

Table 3.2. Management-wise Number of Schools

Management	No. of Schools
Government	4
Local Body	8
Private Aided	10
Private Unaided	2
Total	24

NOTES

In Table 3.2, number of schools have been shown according to the management of schools. So the schools have been classified into 4 categories, namely, Government Schools, Local Body Schools, Private Aided Schools and Private Unaided Schools. A given school belongs to any one of the four categories. Such data is shown as Categorical or Qualitative Data. Here the category or the quality referred to is management. Thus categorical or qualitative data result from information which has been classified into categories. Such categories are listed alphabetically or in order of decreasing frequencies or in some other conventional way. Each piece of data clearly belongs to one classification or category.

We frequently come across categorical or qualitative data in the form of schools categorised according to Boys, Girls and Co-educational; Students' Enrolment categorised according to SC, ST, OBC and 'Others'; number of persons employed in various categories of occupations, and so on.

Let us consider another set of data given in Table 3.3.

Table 3.3. Number of Schools according to Enrolment

Enrolment	No. of Schools
Upto 50	6
51 - 100	15
101 - 200	12
201 - 300	8
Above 300	4
Total	45

In Table 3.3, number of schools have been shown according to the enrolment of students in the school. Schools with enrolment varying in a specified range are grouped together, e.g. there are 15 schools where the students enrolled are any number between 51 and 100. As the grouping is based on numbers, such data are called Numerical or Quantitative Data. Thus, numerical or quantitative data result from counting or measuring. We frequently come across numerical data in newspapers, advertisements etc. related to the temperature of the cities, cricket averages, incomes, expenditures and so on.

CONTINUOUS AND DISCRETE DATA

Numerical or quantitative data may be continuous or discrete depending on the nature of the elements or objects being observed.

Let us consider the Table 3.4 depicting the heights of students of a class.

Table 3.4. Heights of Students of a Class

Height	No. of Students
4'8"- 4' 10"	2

NOTES

4'10" - 5'0"	2
5'0" - 5'2"	5
5'2" - 5'4"	8
5'4" - 5'6"	12
5'6" - 5'8"	10
5'8"- 5'10"	2
Total	41

Table 3.4 gives the data pertaining to the heights of students of a class. Here the element under observation is the height of the students. The height varies from 4' 8" to 5' 10". The height of an individual may be anywhere from 4' 8" to 5'10". Two students may vary by almost zero inch height. Even if we take two adjacent points, say 4' 8.00" and 4' 8.01" there may be several values between the two points. Such data are called Continuous Data, as the height is continuous. Continuous Data arise from the measurement of continuous attributes or vdriables, in which individual may differ by amounts just approaching zero. Weights and heights of children; temperature of a body; intelligence and achievement level of students, etc. are the examples of continuous data.

Let us consider Table 3.3 showing the number of students enrolled and the number of schools according to enrolment. Let us,consider the enrolment of 2 schools as 60 and 61. Now in between 60 and 61, there cannot be any number, as the enrolment will always be in whole numbers. Thus there is a gap of one unit from 60 to 61. Such data, where the elements being observed have gaps are called Discrete Data.

Discbete Data are characterised by gaps in the scale, for which no real values may ever be found. Such data are usually expressed in whole numbers. The size of a family, enrolment of children, number of books etc. are the examples of discrete data. Generally data arising from meagurement are continuous, while data arising from counting or arbitrary classification are discrete.

The achievement scores of students, though presented in discrete form may be considered to constitute continuous data, since a score of 24 represents any point between 23.5 and 24.5. Actually achievement is a continuous attribute or variable.

All peasurements of continuous attributes are approximate in character and as such do not prodide a basis for distinguishing between continuous and discrete data. The distinction is made on the basis of variable being measured. 'Height' is a continuous variable but number of children would give discrete data.

Primary and Secondary Data

The data collected by or on behalf of the person or people who are going to make use of the data refers to primary data. For example, the attendance of

NOTES

children, the result of examinations conducted by you are primary data. If you contact the parents of the children and ask about their educational qualifications to relate them to the performance of the children, this also gives primary data. Actually, when an individual personally collects data or information pertaining to an event, a definite plan or design, it refers to primary data.

Sometimes an investigator may use the data already collected by you, such as the school attendance of children, or performance of students in various subjects. etc, for his/her study, then the data are secondary data. The data used by a person or people other than the people by whom or for whom the data were collected refers to secondary data. For many reasons we may have to use secondary data, which should be used carefully, since the data could have beeb collected with a purpose different from that of the investigator and may lose some detail or may not be fully relevant. For using secondary data, it is always useful to know :

(a) how the data have been collected and processed;

(b) the accuracy of data;

(c) how far the data have been summarised;

(d) how comparable the data are with other tabulations; and

(e) how to interpret the data, especially when figures collected for one purpose are used for another purpose.

Secondary Data

In research, Secondary data is collecting and possibly processing data by people other than the researcher in question. Common sources of secondary data for social science include censuses, large surveys, and organizational records. In sociology primary data is data you have collected yourself and secondary data is data you have gathered from primary sources to create new research. In terms of historical research, these two terms have different meanings. A primary source is a book or set of archival records. A secondary source is a summary of a book or set of records.

Advantages to the secondary data collection method are - 1) it saves time that would otherwise be spent collecting data, 2) provides a larger database (usually) than what would be possible to collect on ones own However there are disadvantages to the fact that the researcher cannot personally check the data so it's reliability may be questioned.

Secondary Data Analysis

There are two different types of sources that need to be established in order to conduct a good analysis. The first type is a primary source which is the initial material that is collected during the research process. Primary data is the data that the researcher is collecting themselves using methods such as surveys,direct

NOTES

observations, interviews, as well as logs(objective data sources). Primary data is a reliable way to collect data because the researcher will know where it came from and how it was collected and analyzed since they did it themselves. Secondary sources on the other hand are sources that are based upon the data that was collected from the primary source. Secondary sources take the role of analyzing, explaining, and combining the information from the primary source with additional information.

Secondary data analysis is commonly known as second-hand analysis. It is simply the analysis of preexisting data in a different way or to answer a different question than originally intended. Secondary data analysis utilizes the data that was collected by someone else in order to further a study that you are interested in completing.

In contrast to secondary data, primary data comes from observations made by the researchers themselves. This often creates credibility issues that do not arise with secondary data.

3.2 TABULAR PRESENTATION OF DATA

It is cumbersome to study or interpret large data without grouping it, even if it is arranged sequentially. For this, the data are usually organised into groups called classes and presented in a table which gives the frequency in each group. Such a frequency table gives a better overall view of the distribution of data and enables a person to rapidly comprehend important characteristics of the data.

For example, a test of 50 marks is administered on a class of 40 students and the marks obtained by these students are as listed below in Table 3.5.

Table 3.5

35, 40, 22, 32, 41, 18, 20, 40, 36, 29, 24, 28, 28, 31, 39, 37, 27, 29, 40, 35, 38, 30,
45, 26, 20, 25, 32, 31, 42, 28, 33, 32, 29, 26, 48, 32, 16, 46, 18, 44.

By going through the marks of 40 students listed in Table 3.5, you may be able to see that the marks vary from 16 to 48, but if you try to comprehend the overall performance it is a difficult proposition.

Now conbider the same set of marks, arranged in a tabular form, as shown in Table 3.6.

Table 3.6

Marks	No. of Students
45 - 49	3
40 - 44	6
35 - 39	6

30 - 34	8
25 - 29	10
20 - 24	4
15 - 19	3
Total	40

From Table 3.6 one can easily comprehend the distribution of marks *e.g.*, 10 students have scores from 25 to 29, while only 7 students have a score lower than 50% etc. Various terms related to the tabulation of data are being discussed below:

Table 3.6 shows the marks arranged in descending order of magnitude and their corresponding frequencies. Such a table is known as frequency distribution. A grouped frequency distribution has a minimum of two columns - the first has the classes arranged in some meaningful order, and a second has the corresponding frequencies. The classes are also referred to as class intervals. The range of scores or values in each class interval is the same. In the given example the first class interval is from 45 to 49 having a range of 5 marks *i.e.*, 45, 46, 47, 48 and 49. Here 45 is the lower class limit and 49 is the upper class limit. As discussed earlier the score of 45 may be anywhere from 44.5 to 45.5, so the exact lower class limit is 44.5 instead of 45. Similarly, the exact upper class limit is 49.5 instead of 49. The range of the class interval is 49.5 - 44.5 = 5 *i.e.*, the difference between the upper limit of class interval and the lower limit of class interval.

For the presentation of data in the form of a frequency distribution for grouped data, a number of steps are required. These steps are :

1. Selection of non-overlapping classes.
2. Enumeration of data values that fall in each class.
3. Construction of the table.

Let us consider the score of 120 students of class X of a school in Mathematics, shown in Table 3.7.

Table 3.7. Mathematics score of 120 class X Students

71 85 41 88 98 45 75 66 81 38 52 67 92 62 83 49 64 52 90 61 58 63 91 5748
75 89 73 64 80 67 76 65 76 65 61 68 84 72 57 77 63 52 56 41 60 55 75 53 45
37 91 57 40 73-66 76 52 88 62 78 68 55 67 39 65 44 47 58 68 42 90 89 39 69
48 82 91 39 85 44 71 68 56 48 90 44 62 47 83 80 96 69 88 24 44 38 74 93 39
72 56 46 71 80 46 54 77 58 81 70 58 51 78 64 84 50 95 87 59

First we have to decide about the number of classes. We usually have 6 to 20 classes of equal length. If the number of scores/events is quite large, we usually have 10 to 20 classes. The number of classes when less than 10 is considered only when the number of scores values is not too large. For deciding the exact number of classes to be taken, we have to find out the range of scores. In Table 3.7 scores vary from 37 to 98 so the range of the score is 62 (98.5 - 36.5 = 62).

NOTES

The length of class interval preferred is 2, 3, 5, 10 and 20. Here if we take class length of 10 then the number of class intervals will be 62/10 = 6.2 or 7 which is less than the desired rumber of classes. If we take class length of 5 then the number of class intervals will be (62/5 = 12.4 or 13 which is desirable.

Now, where to start the first class interval ? The highest score of 98 is included in each of the three class intervals of length 5 *i.e.*, 94 - 98, 95 - 99 and 96 - 100. We choose the interval 95- 99 as the score 95 is multiple of 5. So the 13 classes will be 95 - 99, 90 - 94, 85 - 89, 80 -84, ... , 35 - 39. Here, we have two advantages. One, the mid points of the classes are whole numbers, which sometimes you will have to use. Second, when we start with the multiple of the length of class interval, it is easier to mark tallies. When the size of class interval is 5, we start with 0, 5, 10, 15, 20 etc.

To know about these advantages, you may try the other combinations also *e.g.*, 94 -98, 89 -93, 84 - 88, 79 -83 etc. You will observe that marking tallies in such classes is a bit more difficult. You may also take the size of the class interval as 4. There you will observe that the mid points are not whole numbers. So, while selecting the size of the class interval and the limits of the classes, one has to be careful.

After writing the 13 class intervals in descending order and putting tallies against the concerned class interval for each of the scores, we present the frequency distribution as shown in Table 3.8.

Table 3.8. Frequency Distribution of Mathematics Scores of 120 Class X Students

Scores	Tally	No. of Students
95 -99	III	3
90 - 94	IIII III	8
85 - 89	IIII III	8
80 - 84	IIII IIII	10
75 - 79	IIII IIII	10
70 - 74	IIII IIII	10
65 - 69	IIII IIII IIII	14
60 - 64	IIII IIII I	11
55 - 59	IIII IIII III	13
50 - 54	IIII III	8
45 - 49	IIII IIII	10
40 - 44	IIII III	8
35 - 39	IIII II	7
	Total	120

NOTES

Procedure for Writing the Class Intervals

At the top we write the first class interval which is 95 - 99. Then we find the second class interval by substracting 5 points from the corresponding figures *i.e.*, 90 - 94, and write it under 95- 99. On substracting 5 from 90- 94, the third class interval will be 85 - 89. The procedure is to be followed till we reach the class interval having the lowest score.

Procedure for Marking the Tallies

Let us take the first score in the first row *i.e.*, 71. The score of 71 is in the class interval 70 -74 (70, 71, 72, 73, 74) so a tally (/) is marked against 70 - 74. The second score in the first row is 85, which lies in the class interval 85 - 89 (85, 86, 87, 88, 89), so a tally (/) is marked against 86 - 89. Similarly, by taking, all the 120 scores, tallies are put one by one. While marking the tallies, put your finger on the scores, as a mistake can reduce the whole process to naughlt. The total tallies should be 120 *i.e.*, total number of scores. When against a particular class interval there are four tallies (////) and you have to mark the fifth tally, cross the four tallies (////) to make it 5. So while marking the tallies we make the cluster of 5 tallies. By counting the number of tallies, the frequencies are recorded against each of the class intervals. It completes the construction of table.

In Table 3.8, the exact limits of class interval 95 - 99 are 94.5 and 99.5, as the score of 95 ranges from 94.5 to 99.5 and the score of 99 ranges from 98.5 to 99.5, making the exact range from (94.5 to 99.5. As discussed earlier the data are continuous based on the nature of the variable. The class interval, though customarily arranged in descending order, can also be arranged in ascending order.

3.3 GRAPHICAL PRESENTATION OF DATA

Most people show lack of interest or have no time to go through facts and figures given in a daily newspaper or a magazine. But if these figures are graphically presented, they become easier to grasp and catch the eye and have a more lasting effect on the reader's mind.

The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that it lacks details and is less accurate. In our study, we have the following graphs:

1. Bar Graphs
2. Pie Charts
3. Frequency Polygon
4. Histogram

NOTES

Bar Graphs

This is the simplest type of graphical presentation of data. The following types of bar graphs are possible: *(a)* Simple bar graph *(b)* Double bar graph *(c)* Divided bar graph.

Pie Charts

Sometimes a circle is used to represent a given data. The various parts of it are proportionally represented by sectors of the circle. Then the graph is called a Pie Graph or Pie Chart.

Frequency Polygon

In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value. These points are joined by straight lines. These straight lines may be extended in both directions to meet the X - axis to form a polygon.

Histogram

A two dimensional frequency density diagram is called a histogram. A histogram is a diagram which represents the class interval and frequency in the form of a rectangle.

3.4 TYPES OF GRAPHICAL PRESENTATION OF DATA

Here only a few of the standard graphic forms of representing the data are being discussed as listed below (Beriefly we have discussed above):

- Histogram
- Bar Diagram or Bar Graph
- Frequency Polygon
- Cumulative Frequency Curve or Ogive

Histogram

The most common form of graphical presentation of data is histogram. For plotting a histogram, one has to take a graph paper. The values of the variable are taken on the horizontal axis/scale known as X-axis and the frequencies are taken on the vertical axis/scale known as Y-axis. For each class interval a rectangle is drawn with the base equal to the length of the class interval and height according to the frequency of the C.I. When C.I. are of equal length, which would generally be the case in the type of data you are likely to handle in school situations, the heights of rectangles must be proportional to the frequencies of the Class Intervals.

When the C.I. are not of equal length, the areas of rectangles must be proportional to the frequencies indicated (most likely you will not face this type of situation). As the C.I.s for any variable are in continuity, the base of the rectangles also extends from one boundary to the other in continuity. These boundaries of the C.I.s are indicated on the horizontal scale.

The frequencies for determining the heights of the rectangles are indicated on the vertical scale of the graph.

Let us prepare a histogram for the frequency distribution of mathematics score of 120 Class X students (Table 3.8).

For this, on the horizontal axis of the graph one has to mark the boundaries of the class intervals, starting from the lowest, which is 34.5 to 39.5. So the points on X-axis will be 34.5, 39.5, 44.5, 49.5, ... 99.5. Now on the vertical axis of the graph, the frequencies from 1 to 14 are to be marked. The height of the graphical presentation is usually taken as 60 to 75% of the width. Here, we take 1 cm on X-axis representing 5 scores and 1 cm on Y-axis representing a frequency of 2. For plotting the first rectangle, the base to be taken is 34.5 -39.5 and the height is 7, for the second the base is 39.5 - 44.5 and the height is 8, and so on.

The histogram will be as shown in Figure 3.1.

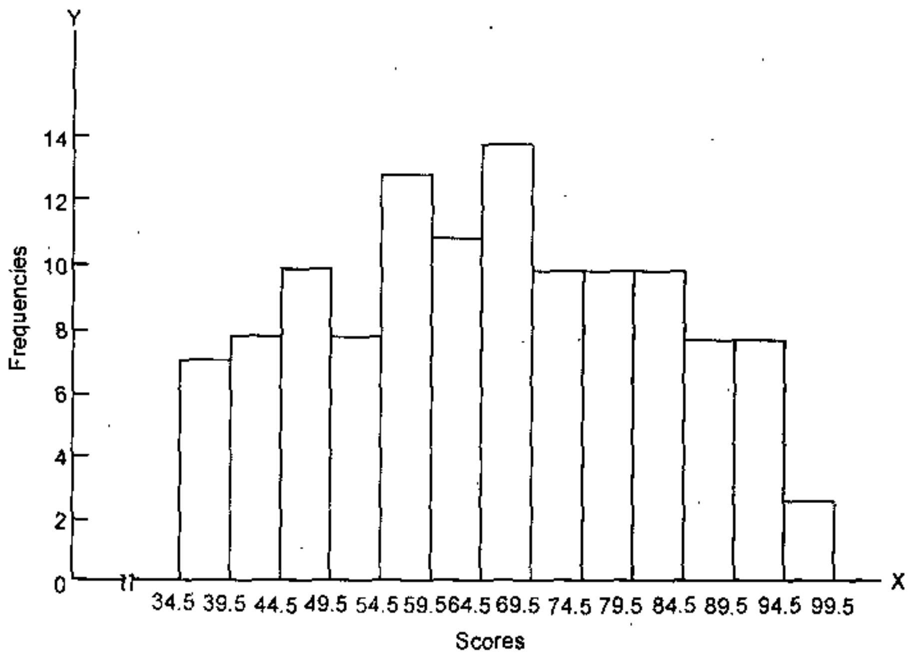

Fig. 3.1. Distribution of Mathematics Scores

Let us re-group the data of Table 3.8 by having the length of class intervals as 10, as shown in Table 3.9.

Table 3.9. Frequency Distribution of Mathematics Scores

Scores	Frequency
90 - 99	11
80 - 89	18

NOTES

70 - 79	20
60 - 69	25
50 - 59	21
48 - 49	18
30 - 39	7
Total	120

To plot the histogram, we mark the boundaries of the class intervals on X-axis. Here the points will be 29.5, 39.5, 49.5,. . .-. . . . , 99.5. On the y-axis, the frequencies to be marked are from 1 to 25. On X-axis, a distance of 1 cm represents a scoare of 10, while on Y-axis, 1 cm represents a frequency of 5. The histogram will be as shown in Figure 3.2.

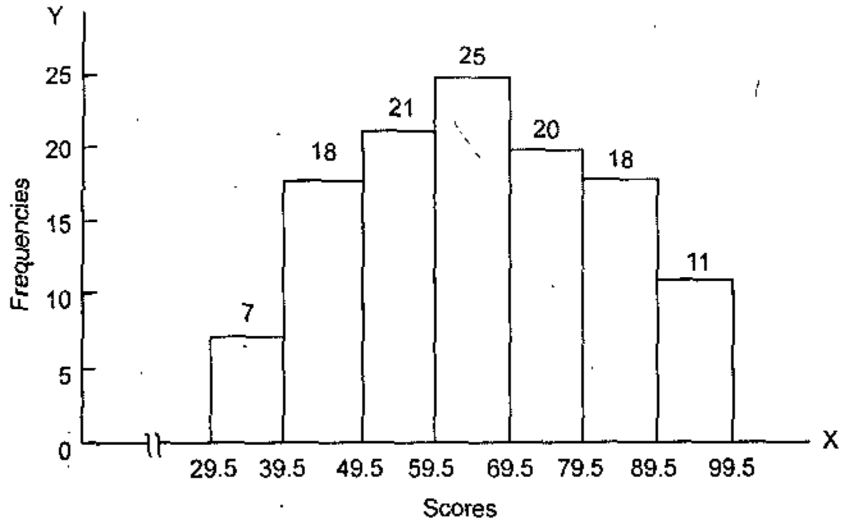

FIG. 3.2. DISTRIBUTION OF MATHEMATICS SCORES

If we observe Figures 3.1 and 3.2, we find that figure 3.2 is simpler than Figure 3.1. Figure 3.1 is complex because the number of class intervals is more. If we further increase the number of class intervals, the figure obtained will be still more complex. So for plotting the histogram for a given data, usually we prefer to have less number of class intervals.

BAR DIAGRAM OR BAR GRAPH

If the variable is discrete, then a histogram cannot be constructed as the classes are not comparable in terms of magnitude. However, a simple graphical presentation, quite similar to histogram, known as bar graph, may be constructed. In a particular town, total number of schools is 24 and the management-wise distribution of schools is as shown earlier in Table 3.2.

Management	No. of Schools
Govenment	4

NOTES

Local Body	8
Private Aided	10
Private Unaided	2
Total	24

The bar graph will be as shown below in Figure 3.3.

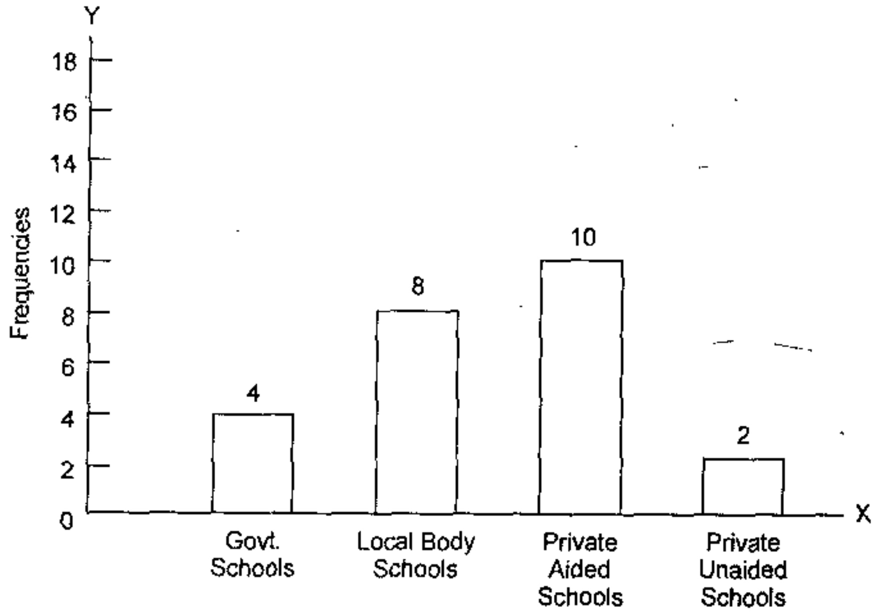

FIG. 3.3. MANAGEMENT-WISE DISTRIBUTION OF SCHOOLS IN A TOWN

For a discrete variable the unit of measure on the horizontal axis is not important. Neither are the classes related to each other. So the bars are equally spaced and are of equal width on the horizontal axis. However, the height of the bars are proportionate to the respective frequencies. Bar graphs are frequently used for pictorial presentation of discrete data. If two variables are used simultaneously, even then bar graphs may be quite effective. For example, if alongwith the total number of schools (management-wise) the number of boys' schools, girls' schools and co-education schools are also to be indicated then this can be done on the same graph paper by using different colours, each indicating the sex-wise category. For each management there will be 4 bars having different colours indicating different categories.

Frequency Polygon

For plotting a frequency polygon, as in case of histogram, the values of the variable are taken on the horizontal axis of the graph and the frequencies are taken on the vertical axis of the graph. In the case of a frequency polygon, one has to indicate the mid points of the C.I. on the horizontal axis, instead of indicating the boundaries of the interval, Here the mid point of the intervals just before the lowest interval and just after the highest interval are also to be indicated.

NOTES

Now by taking the mid points one by one, the points above them are to be plotted corresponding to the frequencies of the intervals. In case of the two additional mid points, the frequency being zero, the points to be plotted are on the X-axis itself. The adjoining points so plotted are to be jained by straight line segments.

Let us again consider the frequency distribution of mathematics scores shown in Table 3.9 and prepare the frequency polygon for the same. The mid points of the C.I.s are respectively 34.5, 44.5, 54.5, ... 94.5. Two additional mid points required are 24.5 and 104.5. Now on the horizontal axis of the graph locate the points 24.5, 34.5, 44.5, ...94.5, 104.5 as shown in Figure 3.4.

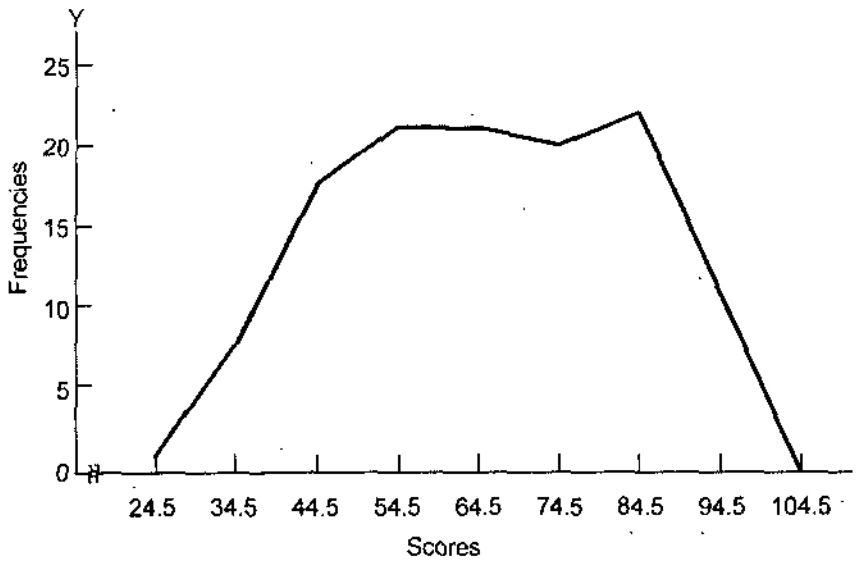

Fig. 3.4. Frequency Polygon of Mathematics Scores

Take the points above the plotted points by taking the heights as 0,7, 18, 21, 25, 20, 18, 11 and 0 respectively. Join these points in a sequence. The frequency polygon obtained will be as shown in Figure 3.4.

Compare the Figure 3.2 and 3.4. You will find that if in Fig. 3.2 you join the mid points of the tops of the rectangle and extend them to one interval on either end of the figure with zero frequency, the figure so obtained will be the frequency polygon shown in Fig. 3.4.

The primary purpose of frequency polygon is to show the shape of the distribution. When two or more frequency distributions are to be compared, the relative frequency polygons are constructed against the same set of axes. Any difference in the shape of these distributions becomes visible. Frequency polygon has an advantage over the histogram.

NOTES

CUMULATIVE FREQUENCY CURVE OR OGIVE

For plotting a cumulative frequency curve or Ogive, first of all cumulative frequencies against each of the intervals are to be written. If we take the frequency distribution of Table 3.9, it will be as shown in Table 3.10.

Table 3.10. Cumulative Frequency Distribution of Scores

Scores	Frequency	Cumulative Frequency
30 - 39	7	7
40 - 49	18	25
50 - 59	21	46
60 - 69	25	71
70 - 79	20	91
80 - 89	18	109
90 - 99	11	120

For getting the cumulative frequencies of a C.I. we take the cumulative frequencies upto the previous interval and add the frequency of that interval into it. Here C.F. indicates that upto 39.5 there are 7 cases, upto 49.5 there are 25 cases, upto 59.5 there are 46 cases, and so on. The difference between the construction of the frequency polygon and ogive is that for frequency polygon, one takes the mid points of the C.I. on horizontal axis, while for ogive one takes the upper boundary of the C.I. on horizontal axis. Again on the vertical axis, in case of Ogive one takes cumulative frequency/cumulative percentage instead of frequency only. The cumulative frequency curve or Ogive for the given data in Table 3.10, will be as shown in Fig. 3.5.

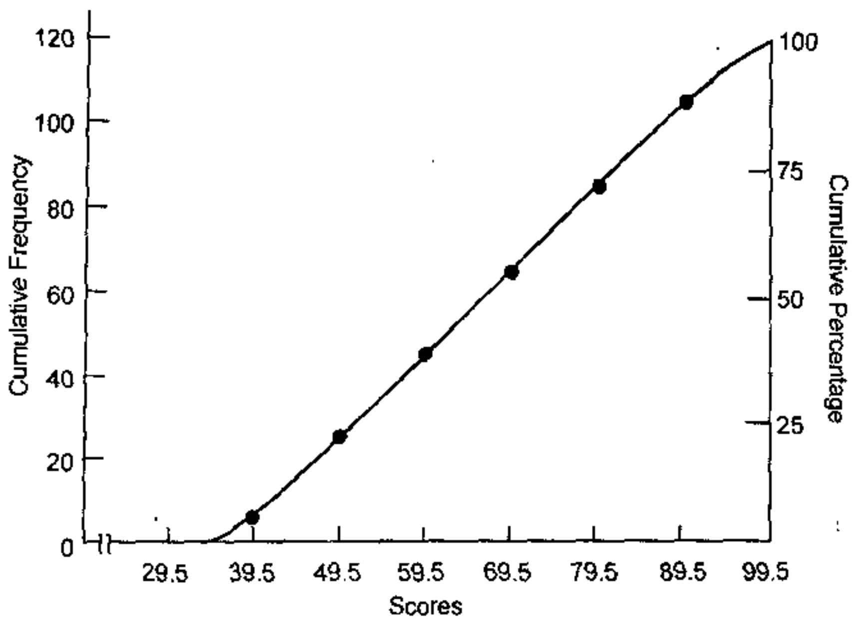

FIG. 3.5. CUMULATIVE FREQUENCY CURVE OR OGIVE

NOTES

In Fig. 3.5, the curve starts from 29.5 (0 Cumulative Frequency) and moves upto 99.5 (120 C.F.). In this case the points have been joined in a sequence with a smoothened cuve, instead of straight line segments. From ogive we can easily find out a point on horizontal axis upto which the specified number of cases or the specified percentage of cases will be available. The only difference between the cumulative frequency curve and ogive is that for cumulative frequency curve, on vertical axis, we take cumulative frequencies, while in case of ogive we also have to take cumulative percentages.

3.5 UNVARIED AND MULTIVARIATE TABLES

The main use of Tables is to obtain univariate or bivariate frequency tables with optional row, column and corner percentages and optional univariate and bivariate statistics. Tables of mean values of a variable can also be obtained.

Both univariate/bivariate tables and bivariate statistics can be output to a file so that can be used with a report generating program, or can be input to Graph ID or other packages such as Excel for graphical display.

Univariate Tables: Both univariate frequencies and cumulative univariate frequencies may be generated for any number of input variables and may also be expressed as percentages of the weighted or unweighted total frequency. In addition, the mean of a cell variable can be obtained.

Bivariate Tables: Any number of bivariate tables may be generated. In addition to the weighted and/or unweighted frequencies, a table may contain frequencies expressed as percentages based on the row marginals, column marginals or table total, and the mean of a cell variable. These various items may be placed in a single table with a possible six items per cell, or each may be obtained as a distinct table.

Multivariate

Multivariate Data Analysis refers to any statistical technique used to analyze data that arises from more than one variable. This essentially models reality where each situation, product, or decision involves more than a single variable. The information age has resulted in masses of data in every field. Despite the quantum of data available, the ability to obtain a clear picture of what is going on and make intelligent decisions is a challenge. When available information is stored in database tables containing rows and columns, Multivariate Analysis can be used to process the information in a meaningful fashion.

Multivariate analysis methods typically used for:

- Consumer and market research.

NOTES

- Quality control and quality assurance across a range of industries such as food and beverage, paint, pharmaceuticals, chemicals, energy, telecommunications, etc.
- Process optimization and process control.
- Research and development.

With Multivariate Analysis you can:

- Obtain a summary or an overview of a table. This analysis is often called Principal Components Analysis or Factor Analysis. In the overview, it is possible to identify the dominant patterns in the data, such as groups, outliers, trends, and so on. The patterns are displayed as two plots
- Analyze groups in the table, how these groups differ, and to which group individual table rows belong. This type of analysis is called Classification and Discriminant Analysis
- Find relationships between columns in data tables, for instance relationships between process operation conditions and product quality. The objective is to use one set of variables (columns) to predict another, for the purpose of optimization, and to find out which columns are important in the relationship. The corresponding analysis is called Multiple Regression Analysis or Partial Least Squares (PLS), depending on the size of the data table

Tools for Multivariate Analysis

Among the various, multivariate tools available, The Unscrambler stands out as an all-in-one multivariate data analysis software product. This product and related ones from CAMO are proven tools that have enabled different organizations solve their Multivariate Analysis requirements.

The main use of TABLES is to obtain univariate or bivariate frequency tables with optional row, column and corner percentages and optional univariate and bivariate statistics. Tables of mean values of a variable can also be obtained.

Summary

- Tables represent narrative or numerical information in tabular fashion. A table arranges information in rows or columns, so that data elements may be referred to easily.
- The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that it lacks details and is less accurate.
- This is the simplest type of graphical presentation of data. The following types of bar graphs are possible: *(a)* Simple bar graph *(b)* Double bar graph *(c)* Divided bar graph.

NOTES

- In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value.
- Multivariate Data Analysis refers to any statistical technique used to analyze data that arises from more than one variable. This essentially models reality where each situation, product, or decision involves more than a single variable.

REVIEW QUESTIONS

1. How is data presentated using tabular method?
2. Discuss various graphs used in graphical presentation of data.
3. What is unvariate and bi-variate table?
4. What do you understand by multivariate data analysis?
5. How is frequency polygon plotted? Discuss with an appropriate example.

FURTHER READINGS

- Garrett, H.E. (1956), *Elementary Statistics, Longmans*, Green & Co., New York.
- Guilford, J.P. (1965), *Fundamental Statistics in Psychology and Education*, Mc Graw Hill Book Company, New York.
- Hannagan, T.J. (1982), *Mastering Statistics*, The Macmillan Press Ltd., Surrey.
- Edward Tufte, *The Visual Display of Quantitative Information* (Cheshire, Conn.: Graphics Press, 1993).
- Howard Wainer, *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot* (Mahwah, N.J.: Lawrence Erlbaum, 1997).

UNIT – IV

STATISTICAL TOOLS

STRUCTURE

LEARNING OBJECTIVES

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

- know the concept of precentage, ratio and proportion as statistical tools;
- explain the concept of central tendency;
- understand the use and limitations of mean, mode and median.

4.1 INTRODUCTION

Statistical methods can be used to summarize or describe a collection of data; this is called descriptive statistics. In addition, patterns in the data may be modeled in a way that accounts for randomness and uncertainty in the observations, and are then used to draw inferences about the process or population being studied; this is called inferential statistics. Descriptive, and inferential statistics (predictive statistics) comprise applied statistics.

There is also a discipline called mathematical statistics, which is concerned with the theoretical basis of the subject. Moreover, there is a branch of statistics called exact statistics that is based on exact probability statements.

NOTES

In applying statistics to a scientific, industrial, or societal problem, it is necessary to begin with a process or population to be studied. This might be a population of people in a country, of crystal grains in a rock, or of goods manufactured by a particular factory during a given period. It may instead be a process observed at various times; data collected about this kind of "population" constitute what is called a time series.

For practical reasons, rather than compiling data about an entire population, a chosen subset of the population, called a sample, is studied. Data are collected about the sample in an observational or experimental setting. The data are then subjected to statistical analysis, which serves two related purposes: description and inference.

- Descriptive statistics can be used to summarize the data, either numerically or graphically, to describe the sample. Examples of numerical descriptors include the mean and standard deviation for continuous data, such as height, and frequency and percentage for categorical data, such as race.
- Inferential statistics is used to model patterns in the data, accounting for randomness and drawing inferences about the larger population. These inferences may take the form of answers to yes/no questions (hypothesis testing), estimates of numerical characteristics (estimation), descriptions of association (correlation), or modeling of relationships (regression). Other modeling techniques include ANOVA, time series, and data mining.

Some well known statistical tests and procedures are:

- Analysis of variance (ANOVA)
- Chi-square test
- Correlation
- Factor Analysis
- Mann-Whitney U
- Mean Square Weighted Deviation MSWD
- Pearson product-moment correlation coefficient
- Regression analysis
- Spearman's rank correlation coefficient
- Student's t-test
- Time Series Analysis

4.2 PERCENTAGE, RATIOS AND PROPORTION

Data are not only described in narrative, they are often described numerically. Three of the most basic types of summarization are:

- frequency distribution;
- percent; and
- average.

Each of these types of summarization may be presented as part of the text or arranged in tables or figures (graphs). Inclusion as part of text ("The average age for children served was 18 months") is an obvious way to report data.

Frequency distribution determines the number of units (e.g., people) which fall into each of a series of specified categories. In order to do a frequency distribution one must have categories. Reporting on age, for example, requires that you group the data first before constructing a frequency distribution (*e.g.*, "birth to 2 years," or "3 to 5 years"). The evaluation might look to see how many parents were members of particular racial or ethnic categories, how many were known to protective services, or how many were referred from a range of referral sources.

Percent is another useful way of describing data. A frequency count can be converted to percent by dividing the number of units for a particular category by the total number of units and multiplying by 100. Percents are often more easily understood than the corresponding frequency counts. Percents can be represented in the same manner as frequency counts. In addition, a pie chart is useful in breaking the total group of people into the percentage of the total represented by each category.

An average is a way of summarizing all of the information into one number. It can be used with data which is non-categorical numerical data. You cannot have a numerical average for gender or race, for example. Using a numerical average is very powerful, but it can also be misleading. A few data points which are very different from the others could substantially change the numerical average. For example, if the ages of children you serve are generally between 1 and 3 years, but you get one child who is 18, the average may be thrown off. Averages can be represented in tables or graphs.

Calculation

Ratio

A ratio is simply a comparison between a pair of numbers.

For example, if there are 3 boys and 4 girls in a room, the ratio of boys to girls is

3 Boys to 4 Girls OR 3:4 OR 3/4

Each of the above represents a method of displaying ratios.

Proportion

A proportion is two ratios written with an equal sign between them.

For example, the numbers listed below represent proportions:

4/3 = 8/6   4:5 = 80:100   2/3 = 200/300   8:10 = 16:20

The examples below further illustrate the concept of proportion:

If the ratio of girls to boys in a high school is 7:9 and there are 2135 girls, how many boys are there?

To solve, we must use the given information to write a proportion:

$$7 : 9 = 2135 : X$$

-Written in another way, we have 7/9 = 2135/X

-Now we must solve for X by cross-multiplication: *(a/b = c/d or a × d = b × c)*

For further information about cross-multiplication, please refer to the bottom of this page.

$$7/9 = 2135/X$$

$$7 \times X = 2135 \times 9$$

$$X = (2135 \times 9) / 7$$

$$X = 2745$$

Therefore, from the value attained for X, there are 2745 boys in the high school.

Percentage

The word percent means "per one hundred". For example, 50 percent means 50 per one hundred. Percents are represented by the symbol %. Therefore, 50 percent is written mathematically as 50 %. It is equivalent to 50/100 or 0.50. There are several ways in which percentages can be used in calculations. Several examples are illustrated below:

A. What is 45% of 110?

"45 % of" something means that you take 45 percent and multiply it by "that something".

In this example,

$$45\% \text{ of } 110 = (45/100) \times 110 = 49.5$$

B. What percent of 110 is 49.5?

In this case we are looking for the percentage. Therefore, we work backwards according to the same equation as in A:

$$X\% \text{ of } 110 = 49.5$$

Which can be written as:

$$X \times 110 = 49.5$$

Cross multiplying gives us:

$$X = 49.5/110$$

$$X = 45\%$$

C. What percent of 675 is 450?

$$X\% \text{ of } 675 = 450$$

$$X \times 675 = 450$$

So we have to solve for X:

$$X = 450/675$$

$$X = .667 \text{ or } 2/3$$

D. A blouse sits on a 20% off table. It has a sale price of Rs. 10. What was its original price?

In this example, our regular price is "b", and our discount is 20 percent less than "b". We were given the sale price to be Rs. 10. Therefore, our equation is written as follows:

$$b - (.20)b = \text{Rs. } 10$$

$$(0.8)b = \text{Rs. } 10$$

$$b = \text{Rs. } 10/0.8$$

$$b = \text{Rs. } 12.50$$

Therefore, its original price was Rs.12.50.

4.3 CENTRAL TENDENCY AND VARIABILITY

Central tendency is defined as the central point around which data revolve. There are three measures of central tendency and each one plays a different role in determining where the center of the distribution or the average score lies. First, the mean is often referred to as the statistical average. To determine the mean of a distribution, all of the scores are added together and the sum is then divided by the number of scores. The mean is the preferred measure of central tendency because it is used more frequently in advanced statistical procedures, however, it is also the most susceptible to extreme scores. For example, if the scores '8' '9' and '10' were added together and divided by '3', the mean would equal '9'. If the 10 was changed to 100, making it an extreme score, the mean would change drastically. The new mean of '8' '9' and '100' would be '39.'

The **median** is another method for determining central tendency and is the preferred method for highly skewed distributions. The media is simply the middle most occurring score. For an even number of scores there will be two middle numbers and these are simply added together and divided by two in order to determine the median. Using the same distribution as above, the scores '8' '9' and '10' would have a median of 9. By changing the '10' to a score of '100' you'll notice that the median of this new positively skewed distribution does not change. The median remains equal to '9.'

NOTES

Median

- Median (Md) always represents the exact center of a distribution of scores.
- It is the score of the case that is in the exact middle of a distribution: half the cases have scores higher and half the cases have scores lower than the case with the median score.
- How to find the median— first, the cases must be placed in order from the highest to the lowest score. Once this is done, find the central or middle case.
- When the number of cases (N) is odd, the value of the median is unambiguous because there will always be a middle cases; and, in this situation, the median is defined as the score exactly halfway between the scores of the two middle cases.
- If the number is even, there will be two middle scores. The median will be the average of the scores of the two middle cases.
- Since the median requires that scores be ranked from high to low, it cannot be calculated for variables measured at the nominal level.
- The score of nominal-level variables cannot be ordered: the scores are different from each other but do not form a mathematical scale of any sort.
- Therefore, the median can only be found either ordinal or interval-ratio data but is generally more appropriate for the former (the ordinal).

Mean

- The arithmetic average.
- It is the most commonly used measure of central tendency.
- It reports the average score of a distribution, and its calculation is straightforward.
- To compute the mean, add the scores and then divide by the number of scores.

Finally, the **mode** is the least used measure of central tendency. The mode is simply the most frequently occurring score. For distributions that have several peaks, the mode may be the preferred measure. There is no limit to the number of modes in a distribution. If two scores tie as the most frequently occurring score, the distribution would be considered bimodal. Three would be trimodal, and all distributions with two or more modes would be considered multimodal distributions.

Interestingly, in a perfectly normal distribution, the mean, median, and mode are exactly the same. As the skew of the distribution increases, the mean and median begin to get pulled toward the extreme scores. The mean gets pulled the most which is why it becomes less valid the more skewed the distribution. The median gets pulled a little and the mode typically remains the same. You can often tell how skewed a distribution is by the distance between these three measures of central tendency.

Measures of Variability

Variability refers to how spread apart the scores of the distribution are or how much the scores vary from each other. There are four major measures of variability, including the range, interquartile range, variance, and standard deviation. The range represents the difference between the highest and lowest score in a distribution. It is rarely used because it considers only the two extreme scores. The interquartile range, on the other hand, measures the difference between the outermost scores in only the middle fifty percent of the scores. In other words, to determine the interquartile range, the score at the 25^{th} percentile is subtracted from the score at the 75^{th} percentile, representing the range of the middle 50 percent of scores.

The variance is the average of the squared differences of each score from the mean. To calculate the variance, the difference between each score and the mean is squared and then added together. This sum is then divided by the number of scores minus one. When the square root is taken of the variance we call this new statistic the standard deviation. Since the variance represents the squared differences, the standard deviation represents the true differences and is therefore easier to interpret and much more commonly used. Since the standard deviation relies on the mean of the distribution, however, it is also affected by extreme scores in a skewed distribution.

4.4 INDIVIDUAL AND GROUP MEASUREMENTS

Measurement is the means of providing quantitative description where numbers are assigned to objects or events according to some rules. These objects or events may sometimes relate to the individuals or groups of individuals depending upon the purpose of measuring an attribute. The rules adopted become the criteria and help in providing systematic and objective mode for deriving accurate judgements about individuals or groups of individuals leading to solutions to various educational problems. The nature of rules may vary from simple to complex and we come across the lowest to highest forms of measurements. From lower to higher levels we have nominal, ordinal, interval and ratio measurement scales. The higher level scales require more restrictive rules and the obtained measurement may require one to undertake more operations.

You as a teacher might be coming across a variety of data pertaining to students' achievement or other characteristics, both of individuals or groups of individuals. We may often be interested in having a concise description of the performance of the group as a whole. In case there are more than one group one may like to compare the groups in terms of their typical performance. Such descriptions of group performances are known as measures of central tendency. Let us assume that we have got the scores of students of three sections of class IX with 40 students each in these sections. We may compute an index of the sets of scores of 40 students in each section which would represent the average performance of the three sections in a given subject. Such an index would be a measure of central tendency. It can very well be used to understand the nature of scores in each section and for making inter-group comparisons. The most commonly used measures of central tendency are Mode, Median and Mean.

NOTES

Scales of Measurement and Measures of Central Tendency

If we have data on some characteristics of students and we wish to describe the group we would be trying to look for an appropriate measure of central tendency. It would be pertinent to note that selection of appropriate measure of central tendency would largely depend upon the nature of data, more specifically the scale of measurement used for representing it. Several kinds of averages may be worked out. Here we will discuss the mode, median and mean for data collected on nominal, ordinal and interval scales respectively.

4.5 DATA ON NOMINAL SCALE AND THE MEASURE OF CENTRAL TENDENCY– THE MODE

Data obtained on the nominal scale is of classificatory type and mostly qualitative. We can count the numher of cases in each category and obtain the frequencies. We may then be interested in noting down the class which is most populous or popular. We frequently deal with 'scores' in measurement in-education. The score obtained by the largest number of individuals is the mode of that group of scores.

For example, if in a section of 40 students of class IX the number of students obtaining the score of 55 is the highest, 55 would be called the mode of the scores for that section. Generally such values are seen to be centrally located, with other values in either direction having relatively lower frequencies. Thus the mode presents a rough estimate of the most typical or the average score in a group of values. It is not essential to have precise scores of all the individuals of the group for finding out mode.

For continuous variables mode provides a quick measure which is less precise and less dependable as compared to other measures of central tendency. If you

NOTES

draw a frequency polygon or a histogram, you will notice the maximum height of this point or the bar.

Sometimes the scores of a group tend to concentrate on two distinctly separate places on the scale. In such a situation the distribution is said to be bimodal and the value or score with highest frequency cannot be said to be the mode. You may examine the following histogram and frequency polygon.

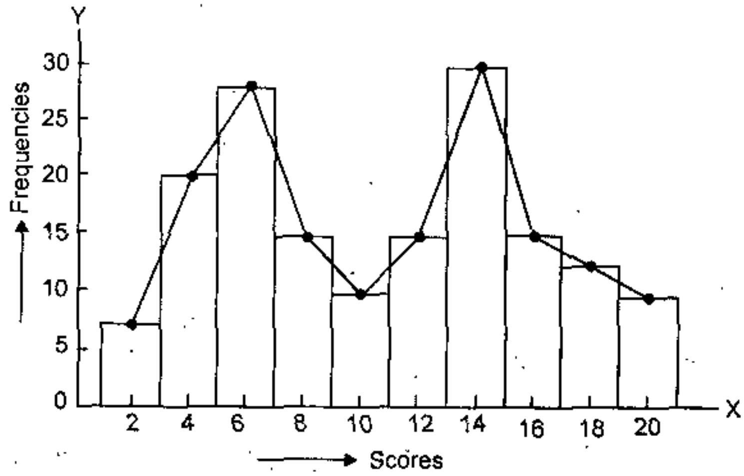

FIG. 4.1

In the above histogram and frequency polygon you may notice that the distribution has two peaks, one at score 6 and the other at score 14. Obviously 14 cannot be the only mode here. Hence it represents a bimodal distribution having two modes at 6 and 14. Some distributions 24 can even be multimodal *i.e.*, having more than two modes. We may define Mode as the point on the scale of measurement with largest frequency in relation to other frequency values in the neighbourhood.

Mode in Ungrouped Data

In a simple ungrouped set of measures, the mode is the single measure or score which occurs most frequently. For example, if the scores of ten students are 13, 12, 14, 15, 12, 14, 18, 12, 14, 14, the most frequent score is 14 as it has been obtained by 4 students. It is, thus, the mode for the given ungrouped data.

Calculating the Mode for the Grouped Data

When the data have been gouped in terms of class intervals and frequencies, the point of greatest concentration of frequencies or the peak in the frequency distribution happens to be the mode. In such a situation the made can be identified by inspection alone. The Mode is the mid point of the class interval having the greatest frequency. Because of this estimation, it is sometime referred to as Crude Mode.

Example 1: Find the mode for the frequency distribution given below :

Class Interval	Frequency
100 - 104	3
95 - 99	4
90 - 94	8
85 - 89	5
80 - 84	2

In the given distribution there are maximum (8) frequencies in the class interval 90 - 94. So the mid point, *i.e.*, 92, is the mode.

You may come across some distributions where more than one mode could be observable as in the following example :

Example 2 :

Class Interval	Frequency
90 - 99	4
80 - 89	12
70 - 79	6
60 - 69	7
50 - 59	9
40 - 49	15
30 - 39	7

The above distribution is a bimodal one which has the highest and near highest frequencies at two distinct points *i.e.*, in class interval 80 - 89 (f = 12) and 40 - 49 (f = 15). So it will have two modes at the mid points of the respective class intervals which are 84.5 and 44.5. Mode is usually interpreted as a simple, inspectional 'avarage' to roughly indicate the centre of concentration in the distribution.

Use of Mode

Mode may be used in the following type of educational situations :

- When the most typical value is wanted as a measure of central tendency. For instance, the most liked boy in the class, the most popular belief of students about vocational courses etc., etc.
- When a quick and approximate measure of central tendency is required.
- When data is incomplete or the distribution is skewed, where most of the values are towards the extremes.

Limitations of Mode

Mode has the limitations associated with the scale of measurement for which it stands. Mode car/ obviously not be subjected to further statistical analysis. It

NOTES

remains as only a rough estimate. Sometimes we may come across bimodal distributions (having two modes) and we do not easily find one composite measure. Yau may examine the following two situations and appreciate the limitations of mode :

Situation I : The scores of students in History for Class VII A are as follows :
22, 37, 45, 66, 32, 64, 65, 67, 66, 67, 65, 67, 38, 66, 66, 65, 32, 66, 67, 65, 64, 64, 67, 52, 47, 67 68, 67, 70

Situation II : The scores of students in Maths for Class IX A are as follows :
18, 20, 23, 24, 24, 25, 24, 24, 24, 30, 35, 40, 46, 48, 50, 56, 62, 62, 62, 62, 60, 47, 38, 62, 62, 24, 28, 62, 80

An inspection of situation I gives the mode of 67 while the adjacent scorer of 64, 65 and 66 seem to be equally potent to become mode. In situation II you notice a bimodal distribution having two modes at 24 and 62 as both seem to be equally frequent in their own places. We may thus conclude that mode is only a crude measure which can be of value when a quick and rough estimate of central tendency is required.

4.6 DATA ON ORDINAL SCALE AND THE MEASURE OF CENTRAL TENDENCY - THE MEDIAN

When data have been arranged in rank order the measure of central tendency may be found by locating a point that divides the whole distribution into two equal halves. Thus median may be defined as the point on the scale of measurement below and above which lie exactly 50 percent of the cases.

Median can therefore be found for truncated (incomplete) data provided we know the total number of cases and their possible placements on the scale. It may be noted that median is defined as a point and not as a score or any particular measurement.

Median in Ungrouped Data

For finding out median in ungrouped data let us study the following examples :

Example 3 : Find the Median for the scores :
2, 5, 9, 8, 17, 12, 14
Here; we have seven scores. On arranging them in ascending (or descending) order we may have the sequence of scores as under :

We find that there are 3 cases above and below 9 and 9 itself is the mid-point of unit interval 8.5 to 9.5. Thus 9 divides the whole distribution into two equal halves. Therefore 9 would be the Median in this case.

Example 4 : Find the Median for the set of scores given below :
12, 17, 18, 15, 20, 19

NOTES

The first step towards finding the Median would be to arrange the given *scores serially (ascending or descending order). You will get 12, 15, 17; 18, 19, 20.* Of the six scores here two scores are below 17 and two above 18. The scores 17 and 18 fall in the middle of the distribution. Here we are not in a position to define median unless we divide this interval covering both the middle scores (17 and 18) into two equal halves. The mid point of the interval 16.5 (lower limit of 17) to 18.5 (upper limit of 18) would be 17.5. Hence 17.5 is the point which divides the distribution into two equal halves. Since there lie exactly 50 percent cases below and above the point, 17.5 it is the Median of the given distribution.

Conclusion : *From the above two examples it may be concluded that the median* can be obtained by finding out the $\{(n + 1)/2\}^{th}$ term when scores have been arranged in ascending or descending order. If you examine example 1, there were 7 cases (odd number). The $\{(n + 1)/2\}^{th}$ term will be $(7+1)/2 = 4^{th}$ term, which happens to be 9. In example 2, there were 6 cases (even number) where $(n+1)^{th}$ term comes to be $(6+1)/2 = 3.5^{th}$ term. Thus Median is found by averaging the third and fourth term *i.e.,* $(17+18)/2 = 17.5$

After calculating the median, you may like to go back and check that half *the values do fall below and above the value you have identified as median.* This will help you to avoid making errors.

Example 5 : Find the median for the given set of scores :

16, 28, 32, 45, 75, 28, 26, 34, 37, 52, 18

We first arrange the scores in ascending order (we could arrange these in descending orders as 16, 18, 26, 28, 28, 32, 34, 37, 45, 52, 75

There are eleven scores in all. The $\{(n+1)/2\}^{th}$ term *i.e.,* $(11+1)/2 = 6^{th}$ term will be the Median.

Median = 32 (Ans.)

Example 6 : Find the Median for the scores given below:

29, 18, 34, 36, 49, 28, 53, 19, 24, 32

We arrange the scores in ascending order :

18, 19, 24, 28, 29, 32, 34, 36, 49, 53

There are ten scores in all. The $\{(n+1)/2\}^{th}$ term *i.e.,* $(10+1)/2 = 5.5^{th}$ term will be the Median

5^{th} term = 29

3^{th} term = 32

Median (5.5^{th} term) = $(29+32)/2 = 61/2 = 30.5$ (Ans.)

Calculation of Median in Grouped Data

As stated earlier, Median is a point on the scale of measurement below which lie exactly fifty percent cases. Obviously fifty percent cases will be above

NOTES

it. For calculating the median, the assumption made in case of grouped data is that frequencies are evenly distributed within the class interval.

Let us take an example to illustrate the point.

Example 7 : Find the Median for the distribution given below:

Class Interval	Frequency	
45 - 49	3	13 = number of cases above the interval containing Median
40 - 44	4	
35 - 39	6	8 = cases in the Median class (fm)
30 - 34	8 (fm)	19 = number of cases (fb) below the interval containing Median
25 - 29	7	
20 - 24	4	
15 - 19	5	
10 - 14	3	
	(N=40)	

In the above example there are a total of 40 cases. We have to find a point below and above which lie 20 cases. There are 13 cases in top 3 class intervals and 19 cases in the bottom four class intervals. The point segregating the values into two halves may be found in class interval 30 - 34 which has 8 cases in it. It is thus called the Median class. Assuming that these 8 frequencies are evenly distributed within the class interval 30 - 34 (exact limits 29.5 to 34.3), we may find the median point which has to be 1 case above 29.5 (or 7 cases below 34.5).

There are 8 cases covering a space of 5 units so one case would take 5/8 spaces. Hence the Median would be 29.5 + (1× 5)/8 = 29.5 + 0.625 = 30.13 (taking approximation upto two decimal points)

This type of calculation gives rise to the formula :

Median

$$= L + \frac{\frac{N}{2} - fb}{\text{fm}}$$

Where L = Lower limit of Median class

N = Total number of cases

fb = Cumulative Frequency below the Median class

fm = Frequency in Median class

i = Size of the class interval

Using this formula for the previous example you can see that

Median = 29.5 + {(20 - 19)/8} × 5

Now we will calculate Median using this formula : = 29.5 + 5/8 = 29.5 + 0.625

= 30.13

Interpretation of Median

The median represents the central point of data here frequencies are halved. In the above example there are fifty percent of cases namely 17, having scorer above 15, *i.e.*, the median.

NOTES

Use of Median

Median is used in the following situations :

- when incomplete distribution is given.
- when the point dividing the distribution into two equal parts is needed.
- when a distribution is markedly skewed. That is, one or more very extreme cases are there to one side of the distribution. Say, in a group of 20 students 18 of them are scoring very low marks say 15 to 40 out of 100 and two students score 95 and 100. Such distributions are known as skewed.
- when interest is limited to finding the placement of cases in the lower half or upper half of the distribution and not in finding how far they are from the central point.

Limitations of Median

Median is not dependent on all the observations and ignores their numerical values. It cannot be used as the centre of gravity of the distribution. Also, it cannot be used for inferential statistical analyses.

4.7 DATA ON EQUAL INTERVAL SCALE AND THE MEASURE OF CENTRAL TENDENCY - THE MEAN

Mean is calculated when the data are complete and presented on equal interval scale. It is most popularly known as the 'Arithmetic Mean'. Mean provides an accurate description of the sample and indirectly, that of the population. It is the sum of measurements* divided by their number.

$$\text{Mean} = \frac{\Sigma X}{N}$$

Where N = Number of cases

*Mean of a distribution of scores may be defined as the point on the scale of measurement obtained by dividing the sum of all the scores by the number of scores.

Calculating Mean for Ungrouped Data

When raw data are given the Mean is computed by adding all these values and dividing by the total number.

Example 8 : Compute Mean for the scores given below

25, 36, 18, 29, 30, 41, 49, 26, 16, 27

NOTES

Mean = (25 + 36 + 18 + 29 + 30 + 41 + 49 + 26 + 16 + 27)/10 = 297/10 = 29.7 (Answer)

Calculating Mean for Grouped Data

There cam be two situations of grouped data :

(i) when scores and frequencies are given; and

(ii) when data have been grouped *i.e.*, frequency is given for each class interval. In the second case we may compute either by long method or by the short method, using the Assumed Mean.

Interpretation of Mean

Mean reprresents the centre of gravity of the scores. Measurements in any sample are perfectly balanced about the mean. In other words the sum of deviations of scores from mean is always zero. In computation of mean all observations are taken into consideration. This is preferred because of its high reliability and its applicability to inferential statistics. The Mean indicates the average performance of the members of the group. It is used to give an idea about how varied are the scores from the central value.

Use of Mean

Mean is used when :

(i) Scores are nearly symmetrically distributed around a central point *i.e.*, distributions are not markedly skewed.

(ii) We wish to know the centre of gravity of a sample.

(iii) Central tendency with greatest stability is wanted.

(iv) When other statistics (standard deviation, coefficient of correlation etc.) for inferential purposes are to be calculated.

(v) Group performances are to be compared with accuracy and precision.

Limitations of Mean

Sometimes Mean of a distribution is highly misleading especially when some of the observations are too large or too small as compared to the others. If you want to study the average class size and there are 5 classes with 100 - 150 students, 10 classes having 50 to 100 students and 35 classes having 30 to 50 students each.

Then the Mean of 55.5 would not represent the typical case. Even within a class if 5 students' scores are 12, 15, 20, 25 and 100, the Mean of 34.4 can be misleading. There are situations where mean may not provide meaningful information.

4.8 RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

NOTES

In your dealings with a variety of data you may come across situations where these three measures of central tendency are very close to each other or at divergence. This largely depends upon the nature of the distribution. In a perfectly symmetrical unimodal distribution these three measures come very close to each other or even become identical. As the symmetry of the distribution varies the three measures (Mean, Median and Mode) register divergence. A very crude sort of relationship among them is shown by the equation :

$$Mode = 3\ Median - 2Mean$$

SUMMARY

- Statistical methods can be used to summarize or describe a collection of data; this is called descriptive statistics.
- Central tendency is defined as the central point around which data revolve. There are three measures of central tendency and each one plays a different role in determining where the center of the distribution or the average score lies. First, the mean is often referred to as the statistical average.
- Measurement is the means of providing quantitative description where numbers are assigned to objects or events according to some rules. These objects or events may sometimes relate to the individuals or groups of individuals depending upon the purpose of measuring an attribute.
- Data obtained on the nominal scale is of classificatory type and mostly qualitative. We can count the numher of cases in each category and obtain the frequencies. We may then be interested in noting down the class which is most populous or popular.
- When data have been arranged in rank order the measure of central tendency may be found by locating a point that divides the whole distribution into two equal halves. Thus median may be defined as the point on the scale of measurement below and above which lie exactly 50 percent of the cases.
- Mean is calculated when the data are complete and presented on equal *interval scale. It is most popularly known as the* 'Arithmetic Mean'.

REVIEW QUESTIONS

1. What do you understand by central tendency?
2. How is mode calculated? discuss with the help of an example.
3. What are the uses and limitations of the median?
4. How is mean calculated? state its uses also.
5. What is the relationship between mean, mode and median?

Reports are a common form of writing because of the inclusion of recommendations which are helpful in implementing the decision.

5.3 MAIN COMPONENTS OF A RESEARCH REPORT

The research report should contain the following components:

- Title and Cover Page,
- Summary of study design, findings and recommendations,
- Acknowledgements,
- Table of contents
  - — List of tables, figures (optional)
  - — List of abbreviations (optional)
  - — INTRODUCTION (statement of the problem in its local context, including relevant literature)
  - — OBJECTIVES
  - — METHODOLOGY
  - — RESEARCH FINDINGS
  - — DISCUSSION
  - — CONCLUSIONS AND RECOMMENDATIONS
  - — REFERENCES
  - — ANNEXES (data collection tools; tables)

5.4 REPORT STRUCTURE

Cover page

The cover page should contain the title, the names of the authors with their titles and positions, the institution that is publishing the report, and the month and year of publication. The title could consist of a challenging statement or question, followed by an informative subtitle covering the content of the study and indicating the area where the study was implemented.

Summary

The summary should be written only after the first or even the second draft of the report has been completed. It should contain:

- a very brief description of the problem (WHY this study was needed) — the main objectives (WHAT has been studied);
- the place of study (WHERE);
- the type of study and methods used (HOW);
- major findings and conclusions, followed by
- the major (or all) recommendations.

NOTES

The summary will be the first (and for busy health decision makers most likely the only) part of your study that will be read. Therefore, its writing demands thorough reflection and is time consuming. Several drafts may have to be made, each discussed by the research team as a whole.

As you will have collaborated with various groups during the drafting and implementation of your research proposal, you may consider writing different summaries for each of these groups.

Acknowledgements

It is good practice to thank those who supported you technically or financially in the design and implementation of your study. Also your employer who has allowed you to invest time in the study and the respondents may be acknowledged. Acknowledgements are usually placed right after the title page or at the end of the report, before the references.

Table of Contents

A table of contents is essential. It provides the reader a quick overview of the major sections of your report, with page references, so that (s)he can go through the report in a different order or skip certain sections.

List of Tables, Figures

If you have many tables or figures it is helpful to list these also, in a 'table of contents' type of format with page numbers.

List of Abbreviations (Optional)

If abbreviations or acronyms are used in the report, these should be stated in full in the text the first time they are mentioned. If there are many, they should be listed in alphabetical order as well. The list can be placed before the first chapter of the report.

The table of contents and lists of tables, figures, abbreviations should be prepared last, as only then can you include the page numbers of all chapters and sub-sections in the table of contents. Then you can also finalise the numbering of figures and tables and include all abbreviations.

Chapter 1: Introduction

The introduction is a relatively easy part of the report that can best be written after a first draft of the findings has been made. It should certainly contain some relevant (environmental/ administrative/ economic/ social) background data about the country, the health status of the population, and health service data which are related to the problem that has been studied. You may slightly comprise or make additions to the corresponding section in your research proposal, including additional literature, and use it for your report.

Then the statement of the problem should follow, again revised from your research proposal with additional comments and relevant literature collected during the implementation of the study. It should contain a paragraph on what you hope(d) to achieve with the results of the study.

Global literature can be reviewed in the introduction to the statement of the problem if you have selected a problem of global interest. Otherwise, relevant literature from individual countries may follow as a separate literature review after the statement of the problem. You can also introduce theoretical concepts or models that you have used in the analysis of your data in a separate section after the statement of the problem.

Chapter 2: Objectives

The general and specific objectives should be included as stated in the proposal. If necessary, you can adjust them slightly for style and sequence. However, you should not change their basic nature. If you have not been able to meet some of the objectives this should be stated in the methodology section and in the discussion of the findings. The objectives form the HEART of your study. They determined the methodology you chose and will determine how you structure the reporting of your findings.

Chapter 3: Methodology

The methodology you followed for the collection of your data should be described in detail. The methodology section should include a description of:

- the study type;
- major study themes or variables (a more detailed list of variables on which data was collected may be annexed);
- the study population(s), sampling method(s) and the size of the sample(s);
- data-collection techniques used for the different study populations;
- how the data was collected and by whom;
- procedures used for data analysis, including statistical tests (if applicable).

If you have deviated from the original study design presented in your research proposal, you should explain to what extent you did so and why. The consequences of this deviation for meeting certain objectives of your study should be indicated. If the quality of some of the data is weak, resulting in possible biases, this should be described as well under the heading 'limitations of the study'.

NOTES

Chapter 4: Research Findings

The systematic presentation of your findings in relation to the research objectives is the crucial part of your report.

The description of findings should offer a good combination or triangulation of data from qualitative and quantitative components of the study. There are two different ways in which you can present your findings:

Chapter 5: Discussion

The findings can now be discussed by objective or by cluster of related variables or themes, which should lead to conclusions and possible recommendations. The discussion may include findings from other related studies that support or contradict your own.

Chapter 6: Conclusions and Recommendations

The conclusions and recommendations should follow logically from the discussion of the findings. Conclusions can be short, as they have already been elaborately discussed in chapter 5. As the discussion will follow the sequence in which the findings have been presented (which in turn depends on your objectives) the conclusions should logically follow the same order.

It makes easy reading for an outsider if the recommendations are again placed in roughly the same sequence as the conclusions. However, the recommendations may at the same time be summarised according to the groups towards which they are directed.

References

The references in your text can be numbered in the sequence in which they appear in the report and then listed in this order in the list of references (Vancouver system). Another possibility is the Harvard system of listing in brackets the author's name(s) in the text followed by the date of the publication and page number, for example: (Shan 2000: 84). In the list of references, the publications are then arranged in alphabetical order by the principal author's last name.

You can choose either system as long as you use it consistently throughout the report.

Annexes or Appendices

The annexes should contain any additional information needed to enable professionals to follow your research procedures and data analysis.

Information that would be useful to special categories of readers but is not of interest to the average reader can be included in annexes as well.

Examples of information that can be presented in annexes are:

- tables referred to in the text but not included in order to keep the report short;
- lists of hospitals, districts, villages etc. that participated in the study;
- questionnaires or check lists used for data collection.

Note:

Never start writing without an outline. Make sure that all sections carry the headings and numbers consistent with the outline before they are word-processed. Have the outline visible on the wall so everyone will be aware immediately of any additions or changes, and of progress made.

Prepare the first draft of your report double-spaced with large margins so that you can easily make comments and corrections in the text.

Have several copies made of the first draft, so you will have one or more copies to work on and one copy on which to insert the final changes for revision.

5.5 STYLE OF WRITING

Remember that your reader:

- Is short of time.
- Has many other urgent matters demanding his or her interest and attention.
- Is probably not knowledgeable concerning 'research jargon'.

Therefore the rules are:

- Simplify. Keep to the essentials.
- Justify. Make no statement that is not based on facts and data.
- Quantify when you have the data to do so. Avoid 'large', 'small'; instead, say '50%', 'one in three'.
- Be precise and specific in your phrasing of findings.
- Inform, not impress. Avoid exaggeration.
- Use short sentences.
- Use adverbs and adjectives sparingly.
- Be consistent in the use of tenses (past or present tense). Avoid the passive voice, if possible, as it creates vagueness (e.g., 'patients were interviewed' leaves uncertainty as to who interviewed them) and repeated use makes dull reading.
- Aim to be logical and systematic in your presentation.

Layout of the Report

A good physical layout is important, as it will help your report:

- make a good initial impression,

- encourage the readers, and
- give them an idea of how the material has been organised so the reader can make a quick determination of what he will read first.

NOTES

Particular attention should be paid to make sure there is:

- An attractive layout for the title page and a clear table of contents.
- Consistency in margins and spacing.
- Consistency in headings and subheadings, *e.g.*, font size 16 or 18 bold, for headings of chapters; size 14 bold for headings of major sections; size 12 bold, for headings of sub-sections, etc.
- Good quality printing and photocopying. Correct drafts carefully with spell check as well as critical reading for clarity by other team-members, your facilitator and, if possible, outsiders.
- Numbering of figures and tables, provision of clear titles for tables, and clear headings for columns and rows, etc.
- Accuracy and consistency in quotations and references.

Common Weaknesses in Writing

Writing is always a challenging job, which requires courage. Starting is usually most difficult. Don't be afraid to make mistakes, otherwise you will never begin! However, it is good to be aware of common pitfalls, which you might try to avoid.

An almost universal weakness of beginning report writers is omitting the obvious. Hardly ever does the description of the country or area contain sufficient data to permit outsiders to follow the presentation of findings and discussion without problems. On the other hand, some data (*e.g.*, exact geographical location on the globe) could be left out which are usually in.

Endless description without interpretation is another pitfall. Tables need conclusions, not detailed presentation of all numbers or percentages in the cells which readers can see for themselves. The chapter discussion, in particular, needs comparison of data, highlighting of unexpected results, your own or others' opinions on problems discovered, weighing of pro's and con's of possible solutions. Yet, too often the discussion is merely a dry summary of findings.

Neglect of qualitative data is also quite common. Still, quotes of informants as illustration of your findings and conclusions make your report lively. They also have scientific value in allowing the reader to draw his/her own conclusions from the data you present. (Assuming you are not biased in your presentation!)

Sometimes qualitative data (*e.g.*, open opinion questions) are just coded and counted like quantitative data, without interpretation, whereas they may be providing interesting illustrations of reasons for the behaviour of informants or of their attitudes. This is serious maltreatment of data that needs correction.

NOTES

5.6 REVISING AND FINALISING THE TEXT

When a first draft of the findings, discussion and conclusions has been completed, all working group members and facilitators should read it critically and make comments.

The following questions should be kept in mind when reading the draft:

- Have all important findings been included?
- Do the conclusions follow logically from the findings? If some of the findings contradict each other, has this been discussed and explained, if possible? Have weaknesses in the methodology, if any, been revealed?
- Are there any overlaps in the draft that have to be removed?
- Is it possible to condense the content? In general a text gains by shortening. Some parts less relevant for action may be included in annexes. Check if descriptive paragraphs may be shortened and introduced or finished by a concluding sentence.
- Do data in the text agree with data in the tables? Are all tables consistent (with the same number of informants per variable), are they numbered in sequence, and do they have clear titles and headings?
- Is the sequence of paragraphs and subsections logical and coherent? Is there a smooth connection between successive paragraphs and sections? Is the phrasing of findings and conclusions precise and clear?

The original authors of each section may prepare a second draft, taking into consideration all comments that have been made. However, you might consider the appointment of two editors amongst yourselves, to draft the complete version.

In the meantime, other group members may (re)write the introductory sections (INTRODUCTION, OBJECTIVES and METHODOLOGY, adjusted from your original proposal).

Now a first draft of the SUMMARY can be written.

Finalising the Research Report

It is advisable to have one of the other groups and facilitators read the second draft and judge it on the points mentioned in the previous section. Then a final version of the report should be prepared. This time you should give extra care to the presentation and layout: structure, style and consistency of spelling (use spell check!).

Use verb tenses consistently. Descriptions of the field situation may be stated in the past tense (*e.g.*, 'Five households owned less than one acre of land.') Conclusions drawn from the data are usually in the present tense (*e.g.*, 'Food taboos hardly have any impact on the nutritional status of young children.)

Note: *– For a final check on readability you might skim through the pages and* read the first sentences of each paragraph. If this gives you a clear impression of the organisation and results of your study, you may conclude that you did the best you could.

Group Work

1. Make an outline for your report on a flipchart, after reviewing your objectives, your sources of information and the outcomes of your data analysis. Number proposed sections and subsections. Stick the outline to the wall in a visible place. Leave sufficient space between the lines for additions (more subsections, for example) and for changes.
2. Start writing, beginning with the chapter on findings. Decide with your facilitator whether you will interpret the data presenting it by variable, by objective or by study population. If you are unsure in the beginning which method of organising the presentation will work best, record your findings and interpretations by study population. In the second draft you can decide how to reorganise and shorten the presentation. Divide writing tasks among sub-groups of one or two persons.
3. Discuss your findings in relation to each other, to the objectives and to other literature, and write the chapter Discussion. Then list the major conclusions in relation to possible recommendations.
4. Develop at the same time the introductory chapters (background and statement of the problem, including new literature, objectives and methodology), adapting what you prepared for the proposal.
5. Finally, develop the summary following the outline given earlier in this module. Take at least half a day for this, working systematically.
6. Keep track of progress in writing and typing, making notes on the flipchart that has the outline of your report.
7. Go over the first draft with the group as a whole checking it for gaps, overlaps, etc. before the second draft is prepared. Have a facilitator from another group read the whole draft report before it is finalised.

SUMMARY

- A report is a very formal document that is written for a variety of purposes in the sciences, social sciences, engineering and business disciplines.

- Research report presents the tangible proof of the conducted research is the major intention of the academic assignment.
- It is advisable to have one of the other groups and facilitators read the second draft and judge it on the points mentioned in the previous section. Then a final version of the report should be prepared.

REVIEW QUESTIONS

1. What is report? Discuss the major types of report writings.
2. Discuss the appropriate structure of research report writing.
3. What are the major errors occur while writing a research report?
4. Explain the method of finalisation of report.

FURTHER READINGS

- Mildred G. Taylor, *How to Write a Research Paper*, Published at Pacific Book (June 1974).
- Fred Pyrczak and Randall R. Bruce, *Writing Empirical Research Reports: A Basic Guide for Students of the Social and Behavioral Sciences.*
- James D. Lester, *Principles of Writing Research Papers* (3rd Edition) (Penguin–Academics), Feb 18, 2010.