

RESEARCH METHODOLOGY

STUDY MATERIAL

THIRD SEMESTER

MCM3C13

For

**M.Com
(2019 ADMISSION ONWARDS)**



UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION

Calicut University P.O, Malappuram, Kerala, India 673 635

190613

UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION

STUDY MATERIAL

THIRD SEMESTER

M.Com

(2019 ADMISSION ONWARDS)

Research Methodology

MCM3C13

Prepared by:

*Dr. Yakoob.C
Reader and Research Guide,
S S College, Areekode.*

©Reserved

CONTENTS

Name of Unit

1. RESEARCH METHODOLOGY – BASIC CONCEPTS
2. SIGNIFICANCE OF RESEARCH
3. CLASSIFICATION OF RESEARCH
4. APPLICATION OF RESEARCH
5. METHODS OF RESEARCH
6. RESEARCH PROCESS
7. RESEARCH PROBLEM
8. LITERATURE REVIEW
9. DEVELOPMENT OF HYPOTHESIS
10. RESEARCH DESIGN
11. POPULATION CENSUS AND SAMPLE STUDY
12. SAMPLE DESIGN
13. DETERMINATION OF SAMPLE SIZE
14. DATA COLLECTION
15. MEASUREMENT AND DATA
16. SCALING
17. DATA PREPARATION AND PROCESSING
18. DATA ANALYSIS
19. INTERPRETATION OF DATA
20. RESEARCH REPORT WRITING

UNIT 1 RESEARCH METHODOLOGY : BASIC CONCEPTS

Introduction

Quite commonly, these days people talk of research – both in academic institutions and outside. Research in common parlance refers to search for further knowledge. More than a set of skills, research is a way of thinking. It is critically examining various aspects of your day to day work, or operations, formulating guiding principles and procedures, and developing and testing new theories, for common good.

Some people consider research as a progressive movement from the known to the unknown. It is actually a voyage towards discovery. We all possess the vital instinct of inquisitiveness. When the unknown confronts us, our inquisitiveness makes us probe and attain understanding of the unknown. This inquisitiveness is the mother of all knowledge and the method, which one employs for obtaining knowledge of whatever be the unknown, can be termed as research.

Basic concepts

Research involves both theoretical and practical aspects. A large number of concepts, principles, procedures and practices govern successful conduct and completion of research. Definitions, meaning, characteristic features, classification, objectives, approaches, significance, process, formulation of problem, literature review, hypothesis formulation, research designs etc will promote the understanding of research in its proper sense.

Definitions

In fact research is the art of scientific investigation. Dictionary definition of research is "careful investigation or inquiry, through searching for new facts in any branch of knowledge". This definition is exhaustive in the sense that it includes every study under research, but at the same time, it is not specific. Research is an academic activity, and the term should be more specific and to be used in technical sense.

According to Clifford Woody "research comprises defining and redefining problems, formulating hypotheses, collecting and evaluating data, making deductions reaching conclusions, testing such conclusions and thus formulating solution or theories."

D. Stessinger and M Stephenson defined research as "manipulation of concepts and symbols for the purpose of generalizing, verifying and extending knowledge, in order to aid in building theory or to promote practice of art."

According to Kerlinger, scientific research is "systematic, controlled, empirical and critical investigation, of propositions about presumed relationships between various phenomena".

These definitions are more inclusive in nature, and focuses on the functions of research, ie, defining problems, formulating hypothesis, collecting data, manipulation, generalizing, building theory, etc.

Research is, thus, original contribution to the existing stock of knowledge, making for its advancement. It is the pursuit of truth with the help of study, observation, comparison

and experiment. In short, the search for knowledge through objective and systematic method of finding solution to a problem, is research. Systematic approach concerning generalization and the formulation of theory is also research.

Meaning

Generally research refers to scientific and deep search for knowledge. Research can also be viewed as scientific and systematic search for pertinent information on a specific topic. Literal meaning research is a careful investigation or inquiry specially through search for new facts in any branch of knowledge.

Some people consider research as a movement from the known to unknown. It is a long voyage in quest of discovery. It is quenching our never ending quest for knowledge it is the solution for inquisitiveness and understanding unknown.

Research is an academic activity and as such the term may be used in a technical sense. Research comprises identifying and defining problem or issue, formulating hypothesis or suggested solutions, collecting, organizing and evaluating relevant data, making deduction, reaching conclusions, testing conclusions, verifying hypotheses making generalizations and building theory. Thus research refers to systematic method consisting of enunciating problem, designing the plan of study, formulating possible propositions, collecting and analyzing data, reaching valid conclusions regarding solution to problem or certain generalization of some theoretical formulation.

Characteristics of research

From these definitions, it is clear that research is a process for collecting, analyzing and interpreting information n to answer questions. So, to qualify as research the process must have certain characteristics.

- **Systematic**

Different activities of research must be performed systematically. It must be undertaken and finished in a step by step approach, starting from identification of a problem and ending on finding a solution or suggestions for improvement.

- **Problem based**

- Researches should be problem or issue based. It must address an issue or solve practical problem, and thus add to existing stock of knowledge,

- **Generalizing**

- Research may lead to generalizing concepts or theories through deduction or induction methods. Research must establish facts or principles applicable in similar general cases.

- **Control**

- In most researches, causes and effects are linked. Studying certain cause, isolating others is called control. Research will give true results only if the variables and their causes are properly controlled.

- Validity
 - Conclusions and findings revealed by research must be valid in similar situations and could be verified by anybody. It should give consistent results.
- Empirical
 - Conclusions drawn are to be based upon solid evidence, should be gathered from information collected from real life experiences and observations.
- Critical
 - Research could be subjected to critical scrutiny of procedures used and method employed. The process of investigation must be fool proof, and free from any drawbacks. The procedures adopted must be able to withstand critical testing and reasoning.

Objectives

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out truth which is hidden and which has not been discovered as yet. Though each research study has its own specific purpose, we mention some general objectives of research below.

- To gain familiarity with a phenomenon to achieve new insights into it. Studies with this objective in view are termed as exploratory or formularize research studies.
- To portray accurately the characteristics of a particular individual situation or a group. These kinds of studies are known as descriptive research studies.
- To determine the frequency with which some event occurs or it is associated with something else. Studies with this objective in view are known as diagnostic research studies
- To test a hypothesis of a causal relationship between variables. Such studies are known as hypothesis testing research studies.

Approaches to research

Descriptions and meanings of research bring to light the fact that there are two basic approaches to research – quantitative approach and qualitative approach.

Quantitative approach

This approach involves generation of data in quantitative form which can be subjected to rigorous quantitative analysis in a formal and rigid fashion. This approach can be further sub classified into inferential, experimental and simulation approaches to research.

The purpose of inferential approach is to form a data base to infer characteristics or relationships of population. This usually means survey research where a sample of population is studied, questioned or observed, to determine the characteristics, and it is then inferred that the population has the same characteristics or not.

Experimental approach is characterized by much greater control over the research environment and, in this case some variables are manipulated to observe their effect on other variables.

Simulation approach involves the construction of an artificial environment within which the relevant information and data can be generated. This permits an observation of the dynamic behavior of a system or its sub systems under controlled conditions.

The term simulation in the context of business and social sciences applications refers to operation of a numerical model that represents the structure of dynamic process. Given the values of initial conditions, parameters and exogenous variables, simulation is a process to represent behaviors of the process over time. Simulation approach can also be useful in building models for understanding future conditions.

Qualitative approach

Qualitative approach to research is concerned with subjective assessment of attitudes, opinions and behavior. Research in such a situation is a function of researcher's insights and impressions. Such an approach to research generates results whether in non quantitative form or in the form which is not subjected to rigorous quantitative analysis. Generally, the techniques of focus group interview, projective techniques and depth interviews are used.

Scientific method in research

Research is a scientific, in-depth study of an issue or topic. Accordingly, for a clear perception of the term research, one should know the meaning of scientific method. The two terms – research and scientific method, are closely related.

In research, we study the nature, reasons and the consequences of set of circumstances which are controlled experimentally or observed as they appear. Researcher is usually interested in particular results, the reactions of those results, and generalizations. On other hand, the philosophy common to all research methods and techniques, although they may vary considerably, from one science to another, is usually given the name of scientific method. In this context, Karl Pearson wrote "the scientific method is one and same in various branches of science and that method is the method of all logically trained minds."

Scientific method is the pursuit of truth and determined by logical considerations. The ideal of science is to achieve a systematic interrelation of facts. Scientific method attempts to achieve this ideal by experimentation, observation and logical arguments, from accepted postulates, and a combination of these three processes, in varying proportions.

In scientific method, logic aids to formulating propositions explicitly and accurately so that their possible alternatives become clear. Further, logic develops the consequences of such alternatives, and then these are compared with observations or phenomenon, it becomes possible for the researcher or the scientist to state which alternative is the most in harmony, with observed facts. All this is done through experimentation and survey investigations, which constitute integral parts of scientific method.

Experimentation is done to test hypothesis and to discover new relationships, if any, among variables. However, sometimes the conclusions drawn on the basis of experimental data may be misleading for faulty assumptions, poorly designed experimentation, badly executed experiments or faulty interpretation. As such the researcher must pay all possible attention while developing the experimental design and drawing inferences. The purpose of survey investigations may also be to provide scientifically gathered information to work as a basis for the researchers for their conclusions.

Principles of scientific method

The scientific method is based on certain basic postulates or principles, which can be stated as under.

- It relies on empirical evidence
- It utilizes relevant concepts
- It is committed to only objective considerations.
- It aims making only adequate and correct statements about population
- It results into probabilistic predictions
- Its methodology is made known to all concerned for critical scrutiny.
- It aims at formulating most general axioms or theory building.

Thus scientific method encourages rigorous, method oriented investigation, wherein the researcher is guided by rules of logical reasoning, based on experimentation and observation. It is a method wherein the investigation proceeds in an orderly manner and a method that implies internal consistency.

Criteria of good Research

Certain conditions must be satisfied to make a research successful and fruitful. Following are those conditions or criteria.

- Certainty – the purpose of the research should be clearly defined and common concepts to be used. Research should remove ambiguity.
- Continuity – the research procedure used should be described in sufficient detail to permit another researcher to repeat the research for further advancement, keeping the continuity of what has already been attained.
- Objectivity – the procedural design of the research should be carefully planned to yield results that are as objective as possible.
- Transparency – the researcher should report with complete frankness, flaws in procedural design and estimate their effects upon the findings.
- Validity – analysis of data should be sufficiently adequate to reveal its significance and the methods of analysis used should be appropriate. The validity and reliability of data should be checked carefully.
- Data based – conclusions should be confined to those justified by the data of the research and limited to those for which the data provide an adequate basis.
- Confidential – greater confidence in research is warranted if the researcher is experienced, has a good reputation in research and is a person of integrity.

Qualities of good research

Research will be critically evaluated, examined and valued. Some researchers are good while certain others are bad and required to be resubmitted after correction and rectification. Following are qualities of good research.

- Systematic – it means that research is structured with specified steps to be taken in a specified sequence in accordance with well defined set of rules. Systematic characteristic of the research does not rule out creative thinking but it certainly does reject the use of guessing and intuition in arriving at conclusions.
- Logical – this implies that research is guided by the rules of logical reasoning and logical process of induction and deduction are of great value in carrying out research. Induction is the process of reasoning from a part to the whole where as deduction is the process of reasoning from some premise to a conclusion which follows from that very premise. In fact, logical reasoning makes research more meaningful in the context of decision making.
- Empirical – by empirical it is implied that research is related basically to one or more aspects of a real situation and deals with practical, concrete data that provides a basis for external validity of research results.
- Replicable – this quality allows research results to be verified by replicating the study and thereby building a sound basis for decisions.

Problems in research

Researchers in India, particularly those engaged in empirical research, are facing several problems, some of which are as follows;

- Insufficient training in methodology – lack of scientific training in the methodology of research is a great impediment for researchers in our country. There is a paucity of competent researcher. Many researchers take a leap in the dark without knowing research methods. Most of the work which goes in the name of research is not methodologically sound. Research to many researchers and even to their guides, is mostly a scissor and paste job without any insight shed on the collected materials. The consequence is obvious, vis, the research results, quite often, do not reflect the reality or realities. Thus a systematic study of research methodology is an urgent necessity. Before undertaking research project, researchers should be well equipped with all the methodological aspects. As such, efforts should be made to provide short duration intensive courses for meeting this requirement.
- Lack of interaction – there is insufficient interaction between the university research departments on one side and business establishments, government departments and research institutions on the other side. A great deal of primary data of non confidential nature remains untouched or untreated by the researchers for want of proper contacts. Efforts should be made to develop satisfactory liaison among all concerned for better and realistic researches. There is need for developing some mechanisms of university-industry interaction programmes so that academics can get ideas from practitioners on what needs to be researched and practitioners can apply the research done by the academics.

- Business secrecy – most of the business units in our country do not have the confidence that the material supplied by them to researchers will not be misused and as such they are often reluctant to supplying the needed information to researchers. The concept of secrecy seems to be sacrosanct to business organizations in the country so much that it proves an impermeable barrier to researchers. Thus there is the need for generating the confidence that information or data obtained from a business unit will not be misused.
- Overlapping studies – research studies overlapping one another are undertaken quite often for want of adequate information. This results in duplication and fritters away resources. This problem can be solved by proper compilation and revision, at regular intervals, of a list of subjects on which and the places where the research is going on. Due attention should be given towards identification of research problems in various disciplines of applied science which are of immediate concern to industries.
- Lack of code of conduct – there does not exist a proper code of conduct for researchers and inter university and inter departmental rivalries are also quite common. Hence, there is need for developing a code of conduct for researchers which, if adhered sincerely, can win over this problem.
- Inadequate secretarial assistance – many researchers in our country also face the difficulty of finding adequate and timely clerical and technical assistance. This causes unnecessary delays in the completion of research studies. Library management and functioning is not satisfactory at many places and much of the time and energy of researchers is spent in tracing out books, journals, reports etc, rather than in tracing out relevant material from them.
- Maintenance of old copies – there is also the problem that many of our libraries are not able to get copies of old and new Acts or Rules, reports and other government publications in time. This problem is felt more in libraries which are away in places from national or state capitals.
- Published data – there is also the difficulty of timely availability of published data from various government and other agencies doing this job in our country. Researcher also faces the problem on account of the fact that the published data vary quite significantly because of differences in coverage by the concerning agencies.

Review Questions

1. What do you mean by research
2. Define research and state its nature
3. Explain significance of research in modern times
4. State characteristics of research
5. What are the objectives of research
6. Explain approaches to research
7. What is qualitative approach to research
8. Describe scientific method in research
9. What principles govern scientific method
10. What are the criteria for good research

11. Explain qualities of good research
12. Elucidate problems in research

UNIT 2 SIGNIFICANCE OF RESEARCH

Anxiety and inquisitiveness is the basis of all human progress. Doubt is often better than over confidence, for it leads to inquiry and inquiry leads to invention. This reflects in the significance of research. Increased amount of research make progress possible. Following points reiterate the importance of research.

1. Scientific thought

Research inculcates scientific and inductive thinking and it promotes the development of logical habits of thinking and organization.

2. Applications

The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times. The increasingly complex nature of business and governance has focused attention on the use of research

- In solving operational problems. Research , as an aid to economic policy, has gained added importance, both for governance and business.

3. Policy making

Research provides the basis for nearly all government policies in our economic system. Government's budgets rest in part on an analysis of the needs and desires of people and on the availability of revenues to meet those needs. The cost of needs has to be equated to probable revenues and this is a field where research is mot needed.

4. Alternatives

Through research we can devise alternative policies and can as well examine the consequences of each of these alternatives. Government has to chalk out programmes for dealing with all facets of the country's various operations and most of these are related directly or indirectly to economic conditions.

5. Decision making

Decision making may not be a part of research, but research certainly facilitates the decisions of the policy maker. Research is considered necessary with regard to the decision making process relating with allocation nation's scarce resources. The plight of cultivators, the problems of big and small business and industry, working conditions, trade union activities, problems of distribution, even the size and nature of defense services are matters requiring research.

6. Allocation of resources

Research is considered necessary with regard to the allocation of nation's scare resources. Resources have a wide variety of uses, and they may be allocated to the most viable activities which is possible through research and analysis.

7. Information

Another area in government, where research is necessary, is collecting information on the economic and social structure of the nation. Such information indicates what is happening in the economy and what changes are taking place. Collecting such statistical information is by no means an easy and routine task, but it involves a variety of research problem. These days nearly all governments maintain large staff of research technicians or experts to carry on this work.

8. Planning

Research has its special significance in solving various operations and planning problems of business and industry. Operations research and market research, along with motivational research are considered crucial and their results assist, in more than one way, in taking business decisions.

9. Societal problems

Research is equally important for social scientists in studying societal relationships and in seeming answers to various societal problems. It provides the intellectual satisfaction of knowing a few things just for the sake of knowledge and also has practical utility for the social scientist to know for the sake of being able to do something better or in a more efficient manners.

10. Academics

To the student community, particularly to those students who are to write a master's or Ph. D. thesis, research may mean careerism or a way to attain a high position in the social structure.

Thus research is the fountain of knowledge for the sake of knowledge and an important source of providing guidelines for solving different business, governmental and social problems. It is a sort of formal training which enables one to understand new developments in one's field in a better way.

Need for research

As the human society is expanding in terms of its material wealth, problems associated with such expansions also develop in various fields. To study such problems of varied nature and to find solutions to such problems, research serves as an essential aid.

It is the only way of analyzing a problem. Besides, research is essential for the researcher to acquire and familiarize himself with various tools of research. As evaluation is one of the important aspects of public or private sector projects, research will have immense value to evaluate the failure or success of those projects. Based on such evaluation, the projects can be aligned in a right direction.

Research has occupied a rightful place in all government departments, industrial concerns and business organizations. Research encourages interdisciplinary approaches to find solutions to problems and to make new discoveries. Research is a basic ingredient for development and therefore serves as a means for rapid economic development. Need for research can be summed up as follows:

1. Innovation research leads to discovery and innovation. Research helps us to discover the unknown be it physical world or the social world. New knowledge is

synonymous to a new born baby. It holds a promise of growth. It gives us the pleasure of watching it mature to newer opportunities. Thus research opens doors for more and more opportunities.

2. Decision making. Research inculcates the habit of logical and inductive thinking, thereby improving our decision making facilities. All business organizations and governments make their decisions after a careful analysis of the prevailing circumstances. Research tools are actively used by them for decisions regarding their policies and day to day work.

3. Trends. The strength to face the future is based on our ability to make certain predictions about the future. These predictions are not shots in the dark, but based on an analysis of the past trends. For example, the growth predictions made by the finance minister are based on a careful trend analysis of select economic indicators. The business cycles are predicted using techniques like barometric forecasting, index numbers etc.

4. Societal impact. Social scientists use research to understand the dynamics of social relationships. It helps us to understand the working social groups and social structures. As a result, several misconceptions are removed and new insights are gained which makes social progress possible. For example, the Hawthorne experiment made a path breaking revelation on the impact of work environment on the efficiency of the employees.

5. Government policy. Almost all the policies and budgets laid down by the government find their basis in research. The annual budget of the government stating the expected revenues and expenditures is drawn after a careful analysis of the government activities. Similarly the monetary and economic policies are framed after a research into the economic trends within the country as well as the trends in major trading partners of the country for the period under study. Government is assisted by various bodies like the National Council for Applied and Economic Research, FICCI, Farmers Union, Trade Unions etc, which provide the data based on the field specific research. Thus government decisions and policies rely on research conducted by the government departments as well as their interested bodies.

Research in Business

Research is widely used in business organizations. Business research is carried out in all the functional areas like marketing, accounting, finance, human resource management etc.

In marketing, research is carried out to study the market trends, identify new market opportunities, explore product related expectations of the people, measure advertising effectiveness, study consumer preferences etc.

In accounting, issues related to transfer pricing, tax management budget control systems, inventory costing methods are frequently researched. Times series behavior of quarterly earnings, cash recovery rates are some other areas of research.

In finance, capital budgeting, optimal financial ratios, mergers and acquisitions, sources of capital, inter corporate financing, exchange rate fluctuations and their impact on earnings, risk management are areas of research.

In human resource management, research can be carried out to study the employees' satisfaction, managerial leadership styles, performance appraisal systems, employee behavior such as employee absenteeism and turnover.

Large business houses may have a well established research department within the organizations. However they may also depend on external research agencies like the Centre for Monitoring India Economy (CMIE), Earnest & Young etc. for providing them with the relevant information on the internal or the external environment of the company or economy/.

Research in India

India is one of the fastest growing economies of the world, yet research in India had not really gained momentum until the turn of the century. A primary reason for this can be cited to the fact that, post independence, all energies were focused on making India a self-reliant economy. It was with the implementation of Intellectual Property Rights as laid down by the World Trade Organization, in its Uruguay Round that India woke up to the challenge of research. Today India has started investing in research in bio technology, agriculture research, medical research, nuclear research etc. The role of government, business and academics in promoting research is being duly recognized and efforts are being made to encourage research in all these sectors. In spite of these developments, research in India has met with certain roadblocks like:

1. Availability of qualified personnel. Research is a complex work demanding qualified personnel. In India large part of academic research is carried out by people who are unaware about research procedures and methods. India does not lack in qualified personnel but more often they get attracted by offers made by foreign research institutions. Whether it is the lack of technical infrastructure or better remuneration, the fact of the matter is that we are losing our own wealth of researches to foreign employers. India needs to develop as well as retain skilled researchers.
2. Poor liaison between academics and business. Presently the tie up between the research cells at various universities and business is very poor. With the exception of few premier academic institutes, research carried out by most other institutions finds little relevance in business or government circles. A proper liaison will not only give direction to research but will also address the problem of paucity of funds.
3. Paucity of funds. The third major problem is that very few funds find their way into research activities. It has been seen that except for few business houses, even the business organisations do not spend more than one percent of their sales revenue on research against a desired 10%. As mentioned earlier, if there is a well developed tie up between government and business on one side and universities on the other side we will find a larger number of sponsors for research.
4. Information data bank. In the age of computers it is very necessary that there should be centralized data bases available. Currently we find that lot of research is repetitive in nature because there is no centralized data bank where information on research work already done can be obtained. Further while doing research, a centralized

data bank of journals, magazines, international and national journals, books etc, help a researcher to save on time and cost factor. There are institutions like UGC, CMIE etc, which have maintained such a database of online reference material. Its use however has yet to gain momentum since it is available at select institutions and in certain cases at a high cost.

5. Ethics in research. A very serious allegation on many researches done in our country is that it is based on reverse engineering or the cut and paste method. It is necessary that we break free of such practices and show honesty in our research endeavors. A research involving collection of primary data should not be cooked up, but be genuinely collected from the field. It should be our endeavor to conduct authentic research surveys in an ethical manner.

Thus, it can be said that research in India has started figuring prominently in academic, government and business budgets. The hurdles are not insurmountable and the gains are definitely plentiful.

Theory Building and research

Research and theory building are closely associated, because many researches lead to building of new theory or generalizations. There exists two methods of reason – induction and deduction. The two opposite, although not contradictory approaches to theory building, use two different logical paths to building knowledge. While the induction approach to theory building develops a generalization based on the examination of a set of specifics, the deduction approach identifies an unknown or a new specific from a known set of general facts.

Induction theory

Research is a systematic process of identifying the problem, developing a hypothesis, collecting and analyzing data and interpreting the results. The inductive approach involves moving from a specific observation to broad generalization. The bottom up approach to theory building is shown in figure 2.1

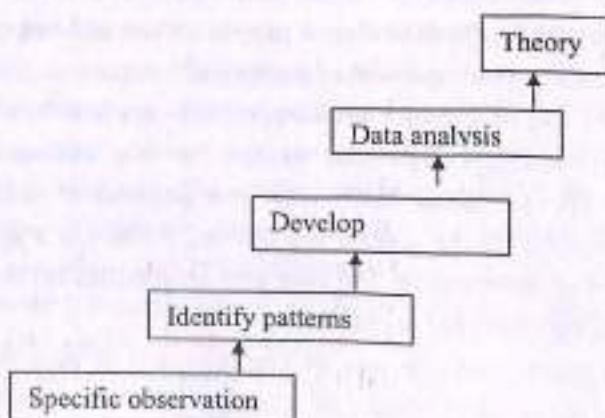


Fig 2.1 Inductive approach to theory building

An example of use of inductive approach is visible in the Newton's theory of Gravity when a simple observation of an apple falling from a tree led to the creation of a theory
Deduction theory

Deduction is defined by the Oxford English Dictionary as follows : " Drawing conclusion from something known or assumed to be known." The deduction approach is clearly visible and made popular by Sherlock Homes through the world famous statement - "Elementary, my dear Watson". The top down approach is followed in the deductive approach to theory building is shown in figure 2.2

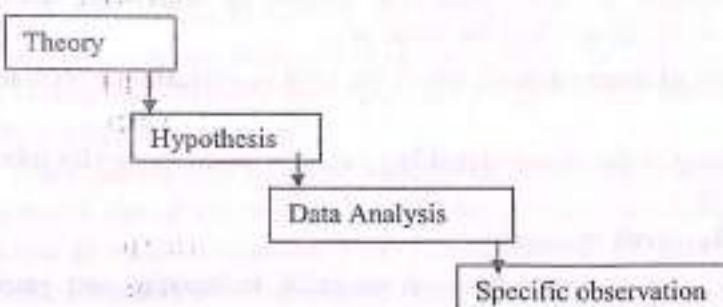


FIG 2.2 Deductive approach to theory building

The process starts with an existing theory from which we develop a specific hypothesis which we would like to test. Thereafter starts the process of collecting data, analyzing it in order to test the hypothesis framed earlier and thus establish if the hypothesis confirms to the theory or not.

Induction Vs Deduction

Induction and deduction are simply two aspects of the scientific method of research. As stated earlier they follow different logical paths which may not necessarily be contradictory. Debate has always been there regarding the applicability and superiority of the two methods in different fields of study. The deductive approach is more frequently used in marketing research whereas the inductive approach is visible in management and sociology research. The distinction between the two approaches is not very sharp. Many a times the difference comes in the way the argument is expressed. For example , listen to this conversation:

Aslam : I 've observed every time I hit the bag with force, it comes back towards me. So I guess the next time when I hit it, it will again come back."

Babu : "That is Newton's Law of Motion which says that every action has an equal and opposite reaction. And so if you hit the bag, it will come back towards you."

In the above arguments, Aslam has used an inductive approach where his observation regarding the behavior of the bag when hit, led him to conclude that it will come back with the same force. On the other hand, Babu's argument used deductive approach by using the Newton's second law of motion.

Research Methods and Methodology

It seems appropriate to explain the difference between research methods and research methodology. Research methods may be understood as all those methods or

techniques that are used for conducting research. Research methods or techniques refer to the methods the researchers use in performing research operations. In other words, all those methods which are used by the researcher during the course of studying his research problem are termed as research methods. Since the object of research, particularly applied research, is to arrive at a solution for a given problem, the available data and the unknown aspects of the problem have to be related to each other to make a solution possible. Keeping this in view, research methods can be put into following three groups

In the first group we include those methods which are concerned with the collection of data. These methods will be used where the data already available is not sufficient to arrive at the required solution.

The second group consists of those statistical techniques which are used for establishing relationships between the data and the unknowns.

The third group consists of those methods which are used to evaluate the accuracy of results obtained.

Research methods falling in the above stated last two groups are generally taken as the *Analytical tools of research*.

Research Techniques and Research Methods

At times, distinction is also made between research techniques and research methods. Research techniques refer to the behavior and instruments we use in performing research operations such as making observations, recording data, techniques of processing data and the like. Research methods refer to the behavior and instruments used in selecting and constructing research techniques. For instance, the difference between methods and techniques of data collection can be better understood from the details given in the following chart.

Methods	Techniques
Analysis of historical records	recording of notes, content analysis, Tape and Film analysis
Analysis of documents	statistical manipulations, abstract analysis
Direct observation	observation of behavioral scales, score cards
Mass observation	Noting behavior, score sheets
Mail questionnaire	Social and demographic analysis, economic analysis
Focus interview	Focused questions and response analysis
Personal interview	Schedule of questions, list of open ended questions
Telephone survey	Opinion analysis, response analysis
Case study	Audio visual recording, Observer recording

From the above, it can be seen that methods are more general in nature. It is the method which generates techniques. However, in practice, the two terms are taken as interchangeable and when talking of research methods, we do, by implication, include research techniques within their compass.

Research methodology is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically. In it we study the various steps that are generally adopted by a researcher in studying his research problem, along with the logic behind them. It is necessary for the researcher to know not only the research methods or techniques but also the methodology. Researcher not only need to know how to develop certain indices or conduct student's t test, but also how to

calculate the mean, the mode, the median or the standard deviation or value of Chi square, how to apply particular research techniques. They also need to know which of these methods or techniques are relevant and which are not, and what would they mean and indicate. Researcher also need to understand the assumptions underlying techniques and they need to know the criteria by which they can decide that certain techniques and procedures will be applicable to certain problems and other will not. All this means that it is necessary for the researcher to design a methodology of his problem as the same may differ from problem to problem. For example, an architect, who designs a building, has to consciously evaluate the basis of his decisions, i.e., he has to evaluate why and on what basis he selects particular size, number and location of doors, windows and ventilators, uses particular materials and not other and the like. Similarly, in research, the scientist has to expose the research decisions to evaluate before they are implemented. He has to specify very clearly and precisely what decisions he selects and why he selects them so that they can be evaluated by others also.

We can say that research methodology has many dimensions and research methods constitute a part of the research methodology. The scope of research methodology is wider than that of research methods. Thus, when we talk of research methodology, we not only talk of the research methods, but also consider the logic behind the methods we use on the context of our research study and explain why we are using a particular method or technique and why we are not using others so that research results are capable of being evaluated, by the researcher himself or by others. Why a research study has been undertaken, how the research problem has been defined, in what way and why the hypothesis has been formulated, what data have been collected and what particular method has been adopted, why particular technique of analyzing data has been used and a host of similar other questions are usually answered, when we talk about research methodology, concerning a research problem or study.

Significance in social sciences

Scientific research used by physical scientists has a long history. The social scientists used scientific method in researches only in the late 18th century. It was developed further in the present century.

Social science research deals with scientific investigations conducted in the field of social sciences and also in behavioral sciences. According to Pauline V Young, social research is "the systematic method of discovering new facts or verifying old facts, their sequences, their interrelationships, causal explanations and natural laws which govern them."

In the words of Herring "scientific research is a cumulative process. It is also an objective process, especially in the social sciences. The understanding can be advanced not only by gains in knowledge but also by discovering and rejecting worn out assumptions"

In the words of Slessinger and Stevenson, social science research aim to "extend, correct or verify knowledge, whether that knowledge aids in the construction of a theory or in the practice of an art"

Social research involves the application of scientific method for understanding, studying and analyzing social life in order to modify, correct or verify existing knowledge

as a system. It deals with human behavior and social problems. Thus social science research studies the social values, beliefs, traditions, events etc. It also tries to find out new facts and verifies old facts on the basis of applying certain tests on old facts.

Research in social sciences has gained importance and relevance, due to following aspects

1. Guide to social planning

Social research has an important role to play in guiding social planning. The success of social planning depends upon the available knowledge of social resources of people, their liabilities, their culture, needs and problems. Social research provides sufficient knowledge and information about the problems and resources of people.

2. Knowledge to control

Social research provides first hand information about the working of society, its organizations and other institutions. It gives more power to control the society as a whole.

3. Base for prediction

In social research prediction of social facts are possible. It helps as a guide to social growth on proper ways and leads to achieve the objectives.

4. Avoid social evil

Social research can disclose and identify social evils faced by the society. On the basis of such information, necessary action can be taken to eradicate such problems.

5. Remedial measures

Social research tries to provide remedial measures necessary to solve the societal problems. Only when problems are scrutinized, remedial measures can be suggested.

Review Questions

1. Why research is considered significant
2. How does research lead to better allocation of resources
3. Explain how research solves societal problems
4. Describe need for research
5. State the role of research in business
6. How is theory building related to research
7. Illustrate induction theory
8. What is deduction theory
9. Distinguish between induction and deduction
10. What are research methods
11. How research methods differ from research methodology
12. Distinguish research methods from research techniques
13. What is the significance of research in social sciences
14. Explain status of research in India

UNIT 3 CLASSIFICATION OF RESEARCH

There are numerous types of researches, because any organized, in-depth, problem based study is research. Researches can be classified according to nature, function, objective, method, application, sponsor etc.

For the sake of easily perceiving the concept of research, let us briefly classify various types of research from three perspectives:

1. Application of research study,
2. Objectives in undertaking research
3. Inquiry mode employed

These three classifications are not mutually exclusive, ie., a research study classified from the view point of application can also be classified from perspectives of objectives and inquiry mode employed. For example, a research project may be classified as pure or applied research, from the perspective of application, as descriptive, co relational, explanatory or exploratory, from the perspective of objectives and as qualitative or quantitative from the perspective of inquiry mode employed.

On the basis of Application

If you examine a research endeavor from the perspective of its application, there are two broad categories : pure and applied research. In the social sciences, according to Bailey, pure research involves developing and testing theories and hypotheses that are intellectually challenging to the researcher, but may or may not have practical application at the present time or in the future. Thus such work often involves the testing of hypotheses containing very abstract and specialized concepts.

Pure research is also concerned with the development, examination, verification and refinement of research methods, procedures techniques and tells that form the body of research methodology. Examples of pure research include developing a sampling technique that can be applied to a particular situation, developing a methodology to assess the validity of a procedure, developing an instrument; say to measure the stress level in people and finding the best way of measuring people's attitudes. The knowledge produced through pure research is sought in order to add to the existing body of knowledge of research methods.

Most of the research in the social sciences is applied. In other words, the research techniques, procedures and methods that form the body of research methodology are applied to the collection of information about various aspects of a situation, issue, problem or phenomenon so that information gathered can be used in other ways, such as for policy formulation, administration and the enhancement of understanding of a phenomenon.

On the basis of Objectives

If you examine a research study from the perspective of the objectives, broadly a research endeavor can be classified as descriptive, co relational, explanatory or exploratory.

A study classified as descriptive research attempts to describe systematically a situation, problem, phenomenon service or program or provides information about, say, the living conditions of community, or describes attitudes towards an issue. For example, it may attempt to describe the types of service provided by an organization, the administrative structure of an organization, the living conditions of original people in the outback, the needs of a community, what it means to go through a divorce, how a child feels living in a house with domestic violence, or the attitudes of employees towards management. The main purpose of such studies is to describe what is prevalent with respect to the issue or

problem under study, or to describe what is prevalent with respect to the issue or problem under study.

The main emphasis in a co relational research study is to discover or establish the existence of a relationship or association or interdependence between two or more aspects of a situation. What is the impact of an advertising campaign on the sale of product? What is the relationship between stressful living and the incidence of heart attack? What is the effect of a health service on the control of a disease, or the home environment on educational achievement? These studies examine whether there is a relationship between two or more aspect of a situation or phenomenon and, these are called co relational studies.

Explanatory research attempts to clarify why and how there is a relationship between two aspect of a situation or phenomenon. This type of research attempts to explain, for example, why stressful living results in heart attacks, why a decline in mortality is followed by fertility decline, or how the home environment affects children' level of academic achievement.

The fourth type of research, from the view point of the objectives of a study, is called exploratory research. That is, when a study is undertaken with the objective either to explore an area where little is known or to investigate the possibilities of undertaking a particular research study. When a study is carried out to determine its feasibility, it is also called a feasibility study or a pilot study. It is usually carried out when a researcher wants to explore areas about which he has little or no knowledge. A small scale study is undertaken to decide if it is worth carrying out a detailed investigation. On the basis of the assessment made during the exploratory study, a full research study may eventuate. Exploratory studies are also conducted to develop, refine and or test measurement tools and procedures.

Although theoretically, a research study can be classified in one of the above perspectives, in practice most studies are a combination of the first three categories, that is, they contain elements of descriptive, co relational and explanatory research. Table 3.1 shows types of research study from the view point of objectives.

Table 3.1 Certain Topics and type of research

Topic	Type of research
Socioeconomic characteristics of residents of a community	
Attitudes of students towards quality of teaching	
Types of service provided by an agency	
Needs of a community	
Sale of a product	
Attitudes of nurses towards death and dying	Descriptive research
Attitude of workers towards management	
Number of people living in a community	
Problems faced by new immigrants	
Extent of occupational mobility among immigrants	
Consumers' likes and dislikes with regard to a product	

Impact of a program Relationship between stressful living and heart attacks Impact of technology on employment impact of maternal health services on infant mortality Effectiveness of counseling on extent of marital problems Impact of advertising campaign on sale of a product Impact of incentives on productivity of workers Effectiveness of an immunization program in controlling fever	Co relational research
Why does stressful living result in heart attacks? How does technology create unemployment How does maternal health service affect infant mortality? Why do some people use a product while other do not? Why do prefer tea to coffee? Why do students go for self medication? Why do seniors rag junior students Why do teachers resort to more assignments	Explanatory research

On the basis of Inquiry Mode

The third perspective in our typology of research concerns the process or method you adopt to find answers to your research questions. Broadly, there are two approaches to inquiry:

Structured approach

The structured approach to inquiry is usually classified a quantitative research and unstructured as qualitative research. In the structured approach everything that forms the research process –objectives, design, sample and the questions that you plan to ask of respondents is predetermined.

Unstructured approach

The unstructured approach, by contrast, allows flexibility in all these aspects of the process. The structured approach is more appropriate to determine the extent of a problem, issue or phenomenon, the unstructured, to explore its nature. Both approaches have their place in research. Both have their strengths and weaknesses. Therefore, you should not lock yourself into solely quantitative or qualitative research. The choice of structured or unstructured approach, and of a quantitative or qualitative mode of inquiry, should depend upon :

Aim of enquiry – exploration, confirmation or quantification.

Use of findings – policy formulation or process of understanding

The distinction between quantitative and qualitative research, in addition to the structured or unstructured process of inquiry is also dependent upon some other considerations.

The study is classified as qualitative if the purpose of the study is primarily to describe a situation, phenomenon, problem or event. The information is gathered through

the use of variables measured on nominal or ordinal scales and if the analysis is done to establish the variation in the situation, phenomenon or problem without quantifying it. The description of an observed situation, the historical enumeration of events, an account of the different opinions people have about an issue, and a description of the living conditions of community are examples of qualitative research.

On the other hand, the study is classified as a quantitative study if you want to quantify the variation in a phenomenon, situation, problem or issue. If information is gathered using predominantly quantitative variables, and if the analysis is geared to ascertain the magnitude of the variation. Examples of quantitative aspects of research study are how many people have a particular problem or how many people hold a particular attitude?

The use of statistics is not an integral part of quantitative study. The main function of statistics is to act as a test to confirm or contradict the conclusions that you have drawn on the basis of your understanding analyzed date. Statistics among other things help you to quantify the magnitude of an association or relationship, provide an indication of the confidence you can place in your findings and help you to isolate the effect of different variables.

Both quantitative and qualitative approaches have their strengths and weaknesses and advantages and disadvantages. Neither one is markedly superior to the other in all aspects. The measurement and analysis of the variables about which information is obtained in a research study are dependent upon the purpose of the study. In many studies you need to combine both qualitative and quantitative approaches. For example, suppose you want to find out the types of service available to victims of domestic violence in a city and the extent of their utilization.

Types of service is the qualitative aspect of the study as finding out about them entails description of these services. The extent of utilization of the services is the quantitative aspect as it involves estimating the number of people who use the services and calculating other indicators that reflect the extent of utilization.

Another way of classifying researches is to divide them on the basis of nature of research and type of investigation. On the basis of nature of research, they may be broadly classified as fundamental research and applied research, or descriptive research and analytical research or quantitative research and qualitative research or conceptual research and empirical research.

Fundamental research and applied research

Fundamental research is mainly concerned with generalization with formulating of a theory. It is a research concerning principles or laws or rules. It aims at the achievement of knowledge and truth. Research studies concentrating on some natural phenomenon or relating to pure mathematics are examples of fundamental research. Basic research aims at some theoretical conclusions. It may verify the old theory or establish a new one. Need for such a research arises with the emergence of new problems and new situating in the society. Fundamental research aims at the attainment, increase and verification of knowledge.

Applied research is concerned with the solution of particular problem. It aims at finding a solution for an immediate problem facing a society or an industrial organization.

Applied research is empirical and practical. It is concerned with applied aspect of life. Research to identify social, economic or political trends that may affect a particular institution, or the marketing research, evaluation research etc are examples of applied research. In the field of family planning research aims at applications of some principles.

While fundamental research is historical, applied research is practically applied to solve the immediate problems. While basic research discovers principles and laws, applied research discovers their application in order to solve some social problems.

Descriptive research and analytical research

Descriptive research includes survey and fact finding enquiries of different kinds. The major purpose of descriptive research is the description of the state of affairs as it exists at present. The main characteristics of this type of research are that there searcher has no control over the variables. He can only report what has happened or what is happening.

In analytical research one has to use facts or information already available and analyze these to make a critical evaluation of the material.

Quantitative research and qualitative research

Quantitative research is applicable to phenomena that are measurable so that they can be expressed in terms of quantity.

Qualitative research is concerned with qualitative phenomena. Research designed to find out how people feel or what they think about a particular subject is qualitative research. Qualitative research is specially important in the behavioral sciences where the aim is to discover underlying motives of human behavior.

Conceptual research and empirical research

conceptual research is that related to some abstract ideas or theory, it is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones.

Empirical research relies on experience or observation alone. It is data based research coming up with conclusions capable of being verified by observations or experiments. It can be called experimental research. In empirical research, there searcher has to first set up a hypothesis or guess as to probable results. He then work out to get enough facts to prove or disprove his hypothesis.

Empirical studies have a great potential – they lead to inductions and deductions this research enables one to develop theories and principles and to arrive at generalizations. As research is based on observations and empirical evidences it improves knowledge and understanding as well as decision making skill and ability.

Based on the style of investigation, research may be classified under the following heads – Laboratory research, ex post facto research, action research, field investigation, evaluation research, library research diagnostic research etc.

Laboratory research

Also called experimental research, it is conducted in laboratory. For example researches in physics are laboratory research. Experimental researches are useful in social sciences also. The only problem in this type of research is to set up a social research laboratory and to create environmental conditions. In management science too, laboratory research can produce results.

Ex post facto research

This is the type of research which is based on analytical and scientific examinations of the relationship between independent and dependent variables. This is an empirical research and the researcher does not have any control over independent variables because they have already been manifested. Independent variables exist already and their relationship with the dependent variables is established in this type of research. For example, advertising expenditure is an independent variable while the sales value may be treated as a dependent variable on advertising expenditure.

Action research

Action research is the research conducted in order to enable him to achieve his purposes more effectively. Action research is conducted through direct action. The action research consists of a number of phases, like baseline survey, systematic action, periodical assessment etc. A teacher should conduct action research to improve his own teaching. A school administration conducts action research to improve its own teaching.

Action research helps a researcher to learn how to solve problems objectively and to study its relations and also to apply its solution in practical life. It helps teachers, principals and examiners of educational institutions to improve their own methods and to remove their defects on the basis of changes taking place daily.

Field research

Field experiments are conducted in real life situations. So the results to be obtained by this method are more realistic. This method is used not only in social sciences research but more extensively in managerial research.

Evaluation research

This is the type of study which is made use for the purpose of evaluating a project which has already been implemented or a work that has already been carried out. Performance evaluation, job evaluation, Programme Evaluation and Review technique etc are examples of this research method. This is a very useful method for the evaluation of actual operational efficiency.

Library research

This is type of research which is based on the data, documentation and material that are available in the library. The researcher depends on the material which can be collected from library sources and secondary sources. Many social science researchers now a day make use of this method mainly because of its convenience.

Technological research

Technological research consists largely of the application of the previously listed kinds of research to the immediate needs of business or industry.

Policy research

Policy research is basically a research with policy implications. Researches which are conducted for the specific purpose of application may be treated as policy researches. The results of such studies are used as indices for policy formulations and implications. Many management researches are policy researches.

Diagnostic research

Diagnostic research is similar to descriptive research but with a different focus. It aims at identifying the causes of a problem and the possible solutions for it.

Historical research

Historical research is a research based on historical data. Its objective is to draw explanations from the past in order to understand the present and to anticipate the future.

Other types of researches

We have also other types of research works on the basis of the purpose of research, environment in which research is done, and time required to accomplish research. From the point of view of time, we can think of research either as one-time research or longitudinal research. On the basis of the environment in which it is carried out, research can be field setting research or laboratory research. Research can be conclusion oriented or decision oriented. In a conclusion oriented research, a researcher is free to pick up the problem.

In decision oriented research the researcher is not free to embark upon research, according to their own inclination. Operations research is an example of decision oriented research, since it is scientific method of proving qualitative basis for decision regarding operation under their control.

Review Questions

1. What is pure research
2. Explain applied research
3. Describe descriptive research
4. What is exploratory research
5. Explain explanatory research
6. Differentiate between explanatory and exploratory research
7. What is co relational research
8. Explain structured research
9. How does structured approach differ from unstructured approach
10. Explain fundamental research
11. Distinguish between fundamental and applied research what is analytical research
12. Give a classification of research on the basis of application
13. What is empirical research
14. Explain Ex post facto research
15. What is action research
16. Explain diagnostic research
17. Describe policy research

UNIT 4 APPLICATIONS OF RESEARCH

Research today is no longer confined to the field of science alone. It has its application in the field of social sciences, education, agriculture, commerce, management, archaeology, anthropology and so on.

Social sciences

Social science research is the scientific investigation conducted in the field of social sciences and also in behavioral sciences. It involves the application of scientific method for

the understanding, studying, and analyzing social life in order to modify, correct or verify the existing knowledge as a system. The social science research is a systematic method of exploring, analyzing and conceptualizing social life in order to examine and, correct or verify knowledge whether that knowledge aids in the construction of a theory or in the practice of an art.

Social research throws light on human behavior and social problems. The main idea behind the social research is to discover new interactions, new knowledge, new facts and also to verify old ones. Such verification is necessary when the technique of analysis improves the concept itself or changes or new facts are available. Social research tries to find out the connection between various human activities and natural laws governing them by means of social and systematized methods. The main purpose of social research is to discover those laws which can be proper guidelines for studying human conduct and behavior.

Motivating factors of social research includes desire to understand cause and effect relationship of some social phenomena, desire to discover new theories, concepts and techniques in order to gain knowledge more efficiently, and curiosity and necessity to know the unknown.

Social research is valuable in procuring necessary information and data for social planning and promotion of social growth by studying social organizations and institutions. It can unfold and identify causes of social evils and thus help to take necessary action for eradicating social evils.

Educational research

Educational research is that activity which is directed towards development of a science of behavior in educational institutions. The ultimate aim of this research is to provide knowledge that will permit the educator to achieve his goals by most effective methods. So, educational research involves an application of main principles of scientific research to the solution of educational problems. Educational research is based upon faith in new methods. It results in improved educational practices. It promotes policies and plans that lead to progress.

Educational research would be classified as development, demonstration or operations research. Better education means better development and formulation of infrastructural aims, better motivation of pupils, better teaching methods, better evaluation and better supervision and administration. These are all activities of educational operations. Educational research can be applied to those operations.

Educationalists are searching for more effective methods of instruction, techniques of evaluation, learning materials etc. In fact, the purpose of research is progress and good life. Therefore need for research in educational practices and policies are being realized increasingly. Decisions based on systematic research in education save time, money and energy.

Educational research is inter disciplinary, since an educational problem has the characteristic of several disciplines, like psychology, sociology, history, economics etc.

Business research

Applications of research in the fields of business and commerce are wide, diverse and most dynamic. It may be noted that there was practically no attempt to make research studies in commerce till 1950s. It was from 1970 onwards that more scholars were attracted towards commerce. Commerce includes trade and aids to trade. Trade includes domestic trade and foreign trade. More than trade, aids to trade have occupied a prime place in commerce. The functions like transport, finance, insurance, advertising, warehousing etc are the aids to trade.

Research study in commerce will provide knowledge that will help in achieving the effective ways of improving trade and improving the function of those which aid the trade.

Business research covers trade, commerce and industry. It is conducted either within a firm by private research organizations and trade associations or in colleges and universities. Market research is an important segment of business research.

Areas of business research

Business research is a systematic process of gathering information for aid in making business decisions and development of business standards and processes. It covers all areas of business such as accounting, finance, Management Information System, market, human resource management and operations management. Some major areas are stated as under:

- **Market research**

It addresses issues pertaining to advertising, product image, packaging and pricing, after-sale services and demand analysis. Market research is widening through application of econometrics to marketing and innovative marketing techniques. Market research is broader in scope and examines all aspects of a business environment. It asks questions about competitors, market structure, government regulations, economic environment. Sometimes the term refers more particularly to the financial analysis of companies, industries or sectors. In this case, financial analysts usually carry out the research and provide the results to investment advisors and potential investors.

- **Product research**

This looks at what products can be produced with available technology and what product innovations near-future technology can develop.

- **Advertising research**

Advertising research is a specialized form of marketing research conducted to improve the effectiveness of advertising. Copy testing, also known as pretesting, is a form of customized research that predicts in market performance of an advertisement before it airs, by analyzing audience levels of attention, brand linking, motivation, entertainment and communication, as well as breaking down the advertisement's flow of attention and flow of emotion. Pre-testing is also used on advertisements still in rough form.

Management research

Significance of management has gone up to day because of the increased acceptance of industrial organizations. Management is now spread to the specialized areas. We have now management concepts as applied to finance, production, personnel and marketing. A

survey of research so far conducted in the field of management, indicates that research work is concentrated only on few areas like working capital, financial analysis and capital market.

There are many other fields in management, where research work if done, can help studying problems in the management and finding their solutions. Thus for developing efficient management impact in the fields of production, business and marketing, research programs help considerably.

Management research can be defined as the process of systematic investigation of any management problems. Considerable numbers of business problems are now given quantitative treatment with some degree of accuracy, with the help of operations research.

Research in management is broadly a systematic activity directed towards investigating managerial or business problems and results in inventing or discovery of management tools for solving problems and making decisions.

Managerial research has unlimited scope in a business organization. Research facilitates formulation of standard formula enabling the executive to rely moderately on personal judgment, especially at the middle and lower levels.

Management research maybe performed with the objectives of decision making relating to , project implementation, policy making, controlling, scanning of business environment, market study, product development, innovation , devising business strategies etc.

Significance of management research

Managers are faced with the critical task of decision making. The manager has the option of taking decisions either intuitively or randomly, or under the directions of an authority or through rational analysis. The best approach is to rationally analyse the problem which requires that the manager should have access to the right kind of information. This information could have been gathered by others at some time in the past. Or it could be gathered by the manager himself specifically with the objective of solving the problem at hand. This information is gathered by a manager thorough research only.

Human Resource Management Research

Human resources and organizational behavior is an area which involves basic for fundamental research as a lot of academic, macro level research may be adapted and implemented by organizations into their policies and programmes. Applied human resource research by contrast is more predictive and solution oriented. Though there are a number o academic and organizational areas in which research is conducted , yet some key contemporary areas which seem to attract more research are as follows:

- Performance management – leadership analysis development and evaluation, organizational climate and work environment studies, talent and aptitude analysis and management, organizational change implementation, management and effectiveness analysis.
- Employee selection and staffing – this includes pre and on the job employee assessment and analysis, staffing studies
- Organizational planning and development – culture assessment, study of merged culture analysis for mergers and acquisitions, manpower planning and development

- Incentive and benefit studies – these include job analysis and performance appraisal studies, recognition and reward studies, hierarchical compensation analysis, employee benefits and reward analysis etc.
 - Training and development – these include training need gap analysis, training development schedules, monitoring and assessing impact and effectiveness of training.
- Critical success factor analysis and employer branding are some emerging areas in which human resource research is being carried out. The first is a participative form of management technique, developed by Rockart (1981) in which the employees of an organization identify their critical success factors and help in customizing and incorporating them in developing the mission and vision of their organization.

Financial and Accounting Research

The areas of financial and accounting research is so vast that it is difficult to provide a pen sketch of the research areas. In this section, we are providing just a brief overview of some research topics:

- Asset pricing, corporate finance and capital markets. The focus here is on stock market response to corporate actions, financial reporting and the impact of factors on returns.
- Financial derivatives and interest rate and credit risk modeling. This includes analyzing interest rate derivatives, development and validation of corporate credit rating models and associated derivatives, analyzing corporate decision making and invest risk appraisal.
- Market based accounting research . Analysis of corporate financial reporting behavior, accounting based valuations, evaluation and usage of accounting information by investors and evaluation of management compensations schemes.
- Auditing and accountability. This includes both private and public sector accounting studies, analysis of audit regulation, analysis of different methodologies, governance and accountability of audit committees.

Considering the nature of the decision required in this area, the research is a mix of historical and empirical research. Behavioral finance is a new and contemporary area in which, probably, for the first time , subjective and perceptual variables are being studied for their predictive value in determining consumer sentiments.

Research in practice

Very little research in the field is pure in nature. That is , very few people do research in research methodology per se. Most research is applied research, which has wide application in many disciplines. Every profession uses research methods in varying amounts in many areas. They use the methods and procedures developed by research methodologists in order to increase understanding in their own profession and to advance professional knowledge base. It is through the application of research methodology that they strengthen and advance their own profession. Examine your own field. You will find that its professional practice follows procedures and practices tested and developed by others over a long period of time. It is in this testing process that you need research skills, the development of which, fall in the category of pure research. As a matter of fact, the

validity of your findings entirely depends upon the soundness of the research methodology adopted.

Research techniques applied entirely in nature are used primarily for professional consolidation, understanding, development and advancement.

As just mentioned, the questions that can be raised about any profession where you directly or indirectly provide a service – health, nursing, occupational therapy, physiotherapy, community health, health promotion, public health, education, town planning, library studies, psychology, business studies, social work – can be considered from four different perspectives.

1. The service provider
2. The service administrator, manager
3. The service consumer, and
4. The professional

These perspectives are summarized below. It is impossible to list all the issues in every discipline but this framework can be applied to most disciplines and situations in the humanities and the social sciences to identify, from the view point of the above perspectives, the possible issues in your own academic field.

Issues for business research

Major participants in business field include service providers, administrator or manager, consumers and professionals. Following are certain model issues faced by such businessmen.

Service provider

To answer questions such as :

- How many people are using the service or the product?
- Why do some people use the service or product while other do not?
- How effective is the service or product?
- How can the service or product be improved?
- Which type of people use or do not use the service or product?
- How satisfied or dissatisfied are the consumer of the service or product?
- What are the problems with the service or product?

Administrator or manager

To answer questions such as :

- What are the needs of the community
- What types of service or product are needed by the community
- How many service providers are needed?
- What are the training needs of the staff?
- How many cases can a worker handle in a day?
- How can the effectiveness of each worker be evaluated?
- How can the service or product be made more popular?

Consumer

To answer questions like:

- Am I, as a consumer, getting value for money?
- How good are the services or product?

- What are long term effects of the product I am using?
- Are my consumer rights protected fully?
- What are the remedies, if my rights are violated?
- What are consumer grievances
- What are Redressal Forums?

Professional

To answer questions such as :

- Which is the most effective intervention for a particular problem or issue?
- What is the relationship between X and Y?
- How valid is a particular theory in the present conditions?
- What is the best way of measuring attitudes
- What is the process through which people decide to adopt a program?
- What are the causes for a particular effect?

And so on

Review questions

1. What is social science research
2. State the objectives of social science research
3. Describe educational research
4. For what purpose, business research is employed
5. What is the significance of management research
6. What problems face a service provider
7. What are applications areas of management research
8. What are the objectives of marketing research
9. What is the significance of consumer research
10. What are the application areas of social science research

UNIT 5 METHODS OF RESEARCH

Research has a wide variety of tools and methods which facilitate successful completion of journey for new knowledge. Such popular methods include case study method, survey method, experimental method, statistical method, historical method etc.

Case Study Method

Empirical social research methods may broadly be divided into two types namely statistical survey method and case study approach. The statistical enquiries are based on large scale collection of facts while case study is based on intensive study of comparatively fewer cases, sometimes confined to very small number of cases or even a single case only. The case study is thus more intensive in nature. By observing a man, only some of his experience can be learnt. But to understand his behavior fully and intimately an intensive study of all his behavior and attitude is to be made. A fairly exhaustive study of a person or a group is called a case study.

Case study is a method of exploring and analyzing the life of a social unit, the social unit may be a person or a family or an institution or an organization or even a community. Case study is a method of collecting information and its analysis. It is a qualitative study. It

is one of the best research designs for exploratory studies. Through case study one can know precisely the factors and causes of a particular phenomenon.

Case study is a method, a technique or an approach to social reality and a mode of organizing data in terms of some chosen units. It is a way of organization of social data so as to preserve the unitary character of the social object, being studied.

How to perform a case study

In case study method, the researcher tries to find out how experiences at different stages of life have influenced the individual's life or behavior. Therefore data are collected in chronological order. Several types of data techniques can be used for collecting data. Depth interview is one such technique. The respondent is asked to review his life from the past to present or from the present to the past. The collection of information starts from the childhood of the respondent.

Following facts come up for study –

1. Respondent's relationship with the members of his family
2. The opinion, attitude, behavior and adjustment with the members of his society and their influence
3. Respondent's experience in schools, religious groups, friend's circle and their influence.

Using these important experiences and events of an individual's past life the researcher can reconstruct the developments of the present behavioral pattern. Besides the individual, his childhood friends, school teachers, family members etc can also be interviewed. Observation method can also be used for the case study. Documents like letters and diaries are the other sources from which information can be collected.

It is impossible for a researcher to record the complete case study of a person. So he selects, from the mass of facts available, only those which are pertinent and essentially to be included. These collected information are then organized and summarized in order to explain the present state of affairs.

Sources of case study

Sources from which information are collected in a case study are:

1. Personal documents – many people keep diaries and write their autobiographies or memories. Many are used to write letters. These personal documents contain the description of remarkable events of the life of the narrator as well as his reactions towards them. Therefore from these personal documents, one can study the writer's personality, social relations and philosophy of life. So they are helpful in studying the personality of the writer. A personal diary may contain even hidden secrets concerning the life of an individual
2. Life history – life history of a person is a source for case study. Through interviews with a respondent, his life history can be known. This is an objective study in which various events of respondent's life are studied with an attempt to find their significance for the society.

Apart from these two sources, there are other source also that help in case study method. They are books, magazines, literature, government documents, communication etc. These contain information about the case under study.

Features of case study method

1. A particular individual or a family or any other social unit is studied.
2. The study of the units selected is made intensively and in minute details.
3. The study under this method extends over a long period of time to ascertain the natural history of the unit. All complexities of facts connected with the social factors are brought under the study. Therefore case study gives us a clear insight into the social unit studied.
4. The study is made qualitatively. Therefore qualitative information is collected.
5. The study makes an effort to know the mutual interrelationship of causal factors.
6. The behavior pattern of the social unit is studied directly.
7. Case study results in fruitful hypothesis and necessary data for testing this hypothesis is also obtained.

Assumptions of case study

A case study is made under the following assumptions or beliefs:

1. The single unit selected for study is representative of the total.
2. The total phenomenon is much complex. So few cases are considered adequate and useful for study
3. A social phenomenon is influenced by the time factor. Soc natural history of the unit is studied for any case study.
4. All human being are fundamentally active. They share common nature.
5. Social conditions and circumstances tend to recur from time to time with marginal changes.

Applications of case study

Case study method is used now a days in several disciplines. Its applications are increasing day by day.

Case study method is widely used in sociology. It is considered to be a systematic research technique in sociology. This method is used for comparative study of different clusters. Anthropologists use the case study method to study the problems relating to their areas of interest. Management experts also use this method for solution to their management problems. It can be seen that physicians, social workers, anthropologists, psychologists and many other persons make use of this study method for extending this knowledge. Case study method can be used not only for collection of data but also to supplement other research methods.

Merits of case study

1. it adds to our knowledge by focused study.
2. it preserves integrity of units under study. It offers insight to guide further research
3. It is flexible. It can be conducted in any kind of social setting
4. it is useful for formulating and testing hypothesis.
5. It is suitable for exploring completely new fields, especially when we have no idea about categories of variables.
6. this method is a best for the study of subjective aspects.

Demerits

1. It may lead to overconfidence and thus, errors.

2. Much of collected data for a case study may turn out to be unimportant.
3. It depends on other experiences regarding past events, thus resulting in inefficiency.
4. It is time consuming and expensive.
5. There is lack of objectivity in case study.

Survey Method

Survey literally means "seeing over a particular thing from a high place." In natural science, survey refers to measuring things while in social science it indicates the investigation of social problems. It is the technique of collection of data through observation, interview, questionnaire, library etc. Now a days the survey method is not just a way of collecting data only. It is also of analyzing the results.

Types of survey

Survey investigations may be classified on the basis of the subject matter, field of study, the purpose, technique etc. On the basis of subject matter under study, surveys can be general and specialized. General survey involves the study of the entire community in a general way. A specialized survey means a study of some particular aspect of the community. Surveys can also be divided as regular or adhoc, census or sample, primary or secondary.

Regular survey is conducted after regular intervals. Adhoc surveys are conducted for certain purposes and are undertaken once for all. In a census survey, the entire population is studied. In a sample survey, a part of the population only is studied. Primary survey is conducted in order to acquire directly the relevant facts. Secondary survey is conducted after the primary survey has been completed. According to the field of study there are social surveys, commercial surveys, community surveys, educational surveys etc.

Social survey

One of the most frequently applied survey technique is social survey. Social survey is a process of collecting and analyzing quantitative facts about social problems, conditions, structures etc of a social group or community occupying definite geographical limits, with a view to bring about changes for betterment to that social group or community.

Objects of social survey

1. The primary aim of a social survey is the collection of facts about certain definite aspects of a community in order to obtain scientific and well ordered information.
2. A social survey aims at the study of the conditions and factors of the social retardation and backwardness of the society. After knowing the cause of backwardness, plans are formulated to remove these conditions.
3. The problems of disease, epidemics, unemployment, unhygienic conditions of living, moral depravity etc are acute among socially and economically backward classes of people. These problems are interrelated. Social survey aims at studying the interrelationship between these factors.
4. The motivation of social survey is practical and utilitarian. The social survey has immediate and practical objective also. As a result of social surveys many remedial measures are taken in the society. Government passes various laws to protect the backward classes of people in the society. This is the effect of social surveys.

Steps in survey

- Identification of the problem to be investigated
- Determining the aim and purpose of the survey
- Formulation of the problem under study
- Identification of various elements or variables involved in the problem.
- Determination of time available for the survey.
- Assessing sources of data.
- Determining units to be surveyed.
- Deciding the techniques for conducting survey.

Merits of survey method

- A survey permits the researcher to come in direct contact with the people.
- It permits greater objectivity. The data are not influenced by any one man's view or belief.
- General survey brings to light a number of problems and propositions which is not possible by pure theoretical analysis.
- A survey allows full emotional impact of social situations.
- Survey method is the most scientific method for providing reliable data whose experimentation is not easily possible.

Limitations of survey method

- Survey method does not take into account historical perspective of the problem under study. It is only a quick and handy approach to a problem requiring urgent remedial measures.
- Surveys are handicapped due to lack of hypothesis. Also they do not always lead to any hypothesis.
- Surveys are usually biased by the preconceived notions of the authority conducting the survey. So surveys may lack clarity and reliability.
- They provide answers to immediate social problems under a frame work. The significance of study ends as soon as the problem ends.
- If the workers collecting data are not efficient and dependable, data collected may not be reliable.

Experiment method

Experiment is the basis of scientific research. An experiment is based on a hypothesis. A hypothesis is the product of rational thinking. Experiments are conducted to test the hypothesis in the expectation of confirming or refuting them. Experimental research can be used to focus attention on the nature and behavior of social groups or institutions.

In natural sciences, everything is studied on the basis of experiments. In recent years, the experiment plays an important role in various social sciences also. But experiments that are conducted in the field of social sciences are different from those conducted in the field of natural sciences. Experimental studies assume that causes can be postulated in advance and that the problem in research is simply one of determining their conditions and efficiency.

Three main features of experimental techniques are ;

- Isolating of factors
- Replication of the experiment
- Quantitative measurement of results

Types of Experimental methods

Experiment is a general term and can be conducted as per several methods:

1. Trial and Error Method – in this method, the researcher formulates a hypothesis and tries to test on actual conditions. For example, a Finance Minister can experiment with a hypothesis that a moderate tax rate structure may entail greater tax compliance
2. Natural or uncontrolled observation method – this type of experiment is conducted in the natural setting or circumstances. It is also called field observation. In this type of experiment certain things are done and their effect is studied in the background of natural setting.
3. Controlled Observation Method – in this method the observation or study is made under controlled conditions. Controlled experiment may be defined as a procedure for testing cause and effect relationship within a setting that permits maximum control over extraneous variations and allows the experimenter to observe the effect of one variable on another.
4. Expost facto experiment – in this type of experiment, the researcher selects two groups with equal number of members. One group is under the influence of one factor and the other is under the influence of another but identical factor. The researcher can judge their positions and he can conclude that the difference in this position is due to the difference in the two factors.
5. Laboratory method of experiment – a laboratory experiment may be defined as one in which the investigator creates an artificial situation with the exact conditions he wishes to have and in which he controls some and manipulates other variables. That is, in this method of experimentation, the researcher creates conditions in which he wants to make his study through manipulation of variables. Experimentation can be done effectively only in laboratory conditions.

In the modern method of social research, a laboratory is designed for carrying out social research. This is generally known as sociological laboratory. In this type of laboratory, the experiments relating to social problems are processed. In a sociological laboratory, the data are available for those who do research. There are trained staff in the laboratory who train others in the research method. There are also various specialists in the laboratory who co operate with the research project.

Aims of experiments

- To study the effect of particular factor on a system. By considering other affecting factors as irrelevant.
- To test a hypothesis about any phenomenon
- To confirm a known law of nature or to suggest a new one.

Merits of experiment method

- Experiment method enables study of the cause and effect of particular factor on a system, by isolating other factors.
- It is possible to formulate hypothesis, test them and reach conclusions, easily, through experiments.

- It is useful in situations where control of variables is often necessary.
- Survey leads to confirmation of known law of nature or to suggest a new one.

Demerits of experiment method

- In social events, many causes are at work. So there is complexity of social events. Hence it is very difficult to study the effect of a single independent variable on a dependent variable.
- The subject matter of social sciences is so heterogeneous that it is very difficult to apply experimentation method in social sciences.
- The experiment method will be fruitful only if subject matter under study is controlled. But the subject matter of social sciences is conscious and fluid and it cannot be easily controlled.
- Since the subject matter of social sciences are fluid and conscious, they cannot be easily observed. For example, thoughts of a man cannot be rightly observed. So experimental method finds it difficult to tackle such situations.

Survey Vs experiment

Survey method and experiment method – both are widely applied in academic as well as social science researches. A comparison between the two reveal following distinctions.

- Survey method is applied for descriptive research studies while experiments are a part of experimental research studies.
- For survey methods large samples are usually required while for experimental studies only small samples are sufficient.
- Survey methods are concerned with present and at times consider past events. In surveys variables that exist and have already occurred are selected and observed.
- Experimental research provides a systematic and logical method by which deliberate manipulations is done in respect of the effect of certain variables. In an experiment, the researcher measures the effects of an experiment which he conducts intentionally.

Statistical Method

Statistical method is a system of procedures and techniques of analysis applied to quantitative data. It consists of a system of statistical techniques applicable to numerical data. Statistical method is also called analytical study method.

Statistical methods are a mechanical process especially designed to facilitate the consideration and the analysis of the large body of quantitative data. The aim of statistical method is to facilitate comparison, study relationships between the two phenomena and to interpret the data for the purpose of analysis.

Steps in statistical methods

1. Collection of material – data must be collected in a systematic manner. The data should be reliable, purposeful and adequate. The data can be collected by library method, experiment method observation method or questionnaire method.

2. Presettation of data – after collection, data should be suitably classified. Classification is the process of arranging data with some purpose. The classified data should be presented in a neat manner with the help of tables, graphs and diagrams.
3. Analysis of data – the presented data should be properly analyzed with the help of statistical tools like averages, measures of dispersion, coefficient of correlation etc. By analysis, we mean making a study about the data using statistical measures.
4. Interpretation of data – the analysis of data helps to interpret what the data tells. Interpretation is the process of translating data into easily understandable language to common people.

Uses of statistical methods

1. Statistical methods help comparison between present and the past
2. It is useful to make projection and to make predictions
3. Relationships can be studied
4. It helps to form probable inferences
5. It describes facts through presentation of data.
6. Its approach is quantitative and hence definite and reliable

Historical method

Historical method of research is a study of past records and other information sources with a view to reconstructing the original and development of an institution or a movement or a system and discovering the trends in the past.

Historical method depends upon inferences and logical analysis of recorded data and indirect evidences rather than upon direct observation.

The objective of historical method is to draw explanations and generalizations from the past trends in order to understand the present and to anticipate the future. It enables us to know our relationship with the past and to plan more intelligently for the future.

Sources of data for historical method

The source s of data for historical research are :

1. Eye witness or a participant in an event
2. Oral testimony of elders
3. Records and other documentary materials
4. Relics.

These sources provide data of past but may be scattered or discontinuous. Therefore they should be studied with caution. The origin and genuineness of the sources should be examined thoroughly. Their authenticity should also be tested.

Steps in historical method

1. The feasibility of the study should be examined. It must be ensured that the problem selected is amenable to historical analysis
2. The selected problem should be formulated and the plan of study should be designed.
3. Sources of data should be located and determined.
4. Genuineness of the sources and the validity of the facts contained in them should be tested.

5. Relevant facts should be collected from the authentic sources and they should be checked.
6. The facts should be arranged into a logical sequence.
7. Adopting induction process, meaningful interpretations and generalizations should be made.

Limitations of historical method

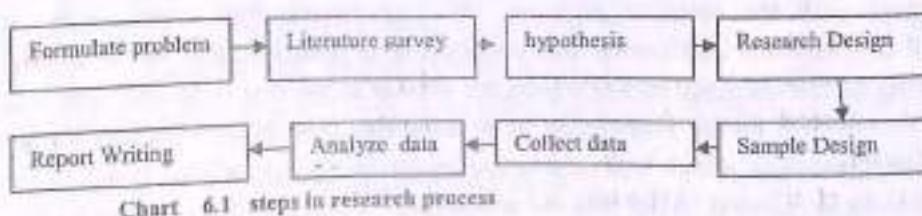
- When the past is much remote, it is very difficult to get the facts.
- It is difficult to test genuineness and authenticity of source and the data available.
- It is difficult to arrange the data in time order of events.
- As the record available is not kept in order, the inference drawn may not be full or accurate.
- Researcher may not succeed in locating all documents.
- It is difficult to know the real significance of data when they belong to remote past.
- Precise measurement, verification and replication are not possible.

Review questions

1. Describe various types of research methods
2. Distinguish between research techniques and research methods
3. Give a classification of different research methods
4. What are features of case study
5. State the problems in the case study
6. Explain the sources of case study
7. State the procedure in case study
8. What are the assumptions in case study method
9. Briefly explain survey method
10. What is social survey
11. What are the objects of social survey method
12. Compare survey method with experiment method
13. State steps in statistical method
14. What are the merits of historical method.

UNIT 6 RESEARCH PROCESS

Before embarking on the details of research methodology and techniques, it seems appropriate to present a brief overview of the research process. Research process consists of series actions or steps necessary to effectively carry out research. The chart shown in figure 6.1 illustrates research process.



The chart indicates that the research process consists of number of closely related activities, as shown through I to VII. But such activities overlap continuously rather than

following a strictly prescribed sequence. At times, the first step determines the nature of the last step to be undertaken. If subsequent procedures have not been taken into account in the early stages, serious difficulties may arise which may even prevent the completion of the study. One should remember that neither various steps involved in a research process are mutually exclusive; nor they are separate and distinct. They do not necessarily follow each other in any specific order and the researcher has to be constantly anticipating at each step in the research process the requirements of the subsequent steps. However the following order concerning various steps provides a useful procedural guideline regarding the research process: (i) formulating the research problem; (ii) extensive literature survey; (iii) developing the hypothesis; (iv) preparing the research design; (v) determining sample design; (vi) collecting the data; (vii) execution of the project; (viii) analysis of data; (ix) hypothesis testing; (x) generalization and interpretation; and (xi) preparation of the report or presentation of the results; i.e., formal write-up of conclusions reached.

A brief description of the above stated steps will be helpful.

Formulating the research problems:

The foremost step in research process is formulation of the problem to be researched. There are two types of research problems - viz. those which relate to states of nature and those which relate to relationships between variables. At the very outset, the researcher must single out the problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject-matter that he would like to inquire into. Initially, the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved. Then, the feasibility of particular solution has to be considered before a working formulation of the problem can be set up. The formulation of a general topic into a specific research problem, thus, constitutes the first step a scientific enquiry. Essentially two steps are involved in formulating the research problem, viz., understanding the problem thoroughly, and rephrasing the same into meaningful terms from an analytical point of view.

The best way of understanding the problem is to discuss it with one's own colleagues or with those having some expertise in the matter. Often, a problem is put in general terms and it is up to the researcher to narrow it down and phrase the problem in operational terms. In private business units or in governmental organizations, the problem is usually earmarked by the administrative agencies with whom the researcher can discuss as to how the problem originally came out and what considerations are involved in its possible solutions.

The researcher must at the same time examine all available literature to get himself acquainted with the selected problem. He may review two types of literature – the conceptual literature concerning the concepts and theories, and the empirical literature consisting of studies made earlier which are similar to the one proposed. The basic outcome of the review will be the knowledge as to what data and other materials are available for operational purposes which will enable the researcher to specify his own research problem in a meaningful context. After this the researcher rephrases the problem into analytical or operational terms i.e., to put the problem in as specific terms as possible. This task of formulating, or defining a research problem is a step of greatest importance in the entire

research process. The problem to be investigated must be defined unambiguously for that will help discriminating relevant data from irrelevant ones. Care must, however, be taken to verify the objectivity and validity of the background facts concerning the problem. The statement of the objective is of basic importance because it determines the data which is to be collected, the characteristics of the data which is relevant, relations which are to be explored, the choice of techniques to be used in these explorations and the form of the final report. If there are certain pertinent terms, the same should be clearly defined along with the task of formulating the problem. In fact, formulations of the problem often follows a sequential pattern where a number of formulations are set up, each formulation more specific than the preceding one, each one phrased in more analytical terms, and each more realistic in terms of the available data and resources.

Literature survey:

Once the problem is formulated, a brief summary of it should be written down. At this juncture the researcher should undertake extensive literature survey connected with the problem. For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals conference proceeding, government reports, book etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another. The earlier studies, if any, which are similar to the study in hand should be carefully studied. A good library will be of a great help to the researcher at this stage. In this era of internet, it has become quite easy to articles. One can use some search engine, like 'Google' or specially 'Google scholars' to search scholarly articles.

Development of hypotheses:

After extensive literature survey, researcher should state in clear terms the assumptions leading to the problem under study. It takes the shape of working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research. They also affect the manner in which tests must be conducted in the analysis of data and indirectly the quality of data which is required for the analysis. Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested. The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facts of the problem. It also indicates the type of data required and the type of methods of data analysis to be used.

How does one go about developing working hypotheses? The answer is by using the following approach:

1. Discussions with colleagues and experts about the problem, its origin and the objectives in seeking a solution;
2. Examinations of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;
3. Review of similar studies in the area or of the studies on similar problems; and

4. Exploratory personal investigation which involves original field interviews on a limited scale with interested parties and individuals with a view to secure insight into the practical aspects of the problem.

Thus, working hypotheses arise as a result of priority thinking about the subject, examination of the available data and material including related studies and the counsel of experts and interested parties. Working hypotheses are more useful when stated in precise clearly defined terms. It may as well be remembered that occasionally we may encounter a problem where we do not need working hypotheses, especially in the case of exploratory or formulative researches which do not aim at testing the hypothesis. But as a general rule, specification of working hypotheses is another basic step of the research process in many research problems.

Preparing Research Design:

The research problem having been formulated in clear terms, the researcher will be required to prepare a research design, i.e., he will have to state the conceptual structure within which research would be conducted. The preparation of such design facilitates research to be as efficient as possible yielding maximal information. In other words, the function of research design is to provide for the collection of relevant evidence with optimum effort, time and expenditure.

But how all these can be achieved depends mainly on the research purpose. Research purposes may be grouped into four categories, viz., Exploration, Description, Diagnosis, and Experimentation.

A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose happens to be an accurate description of a situation or of an association between variables, the suitable design will be one that minimizes bias and maximizes the reliability of the data collected and analyses.

There are several research designs, such as, experimental and non-experimental hypothesis testing. Experimental designs can be either informal designs (such as before-and-after without control, after-only with control, before-and-after with control) or formal designs, simple and complex factorial designs), out of which the researcher must select one for his own project.

Factors in research design

There are several types of research designs. But, the preparation of any research design, appropriate for a particular research problem, involves usually, consideration of the following factors:

1. Means of obtaining the information;
2. Availability and skills of the research and his staff (if any);
3. Explanation of the way in which selected means of obtaining information will be organized and the reasoning leading to the selection;
4. Time available for research; and
5. Cost factor relating to research, i.e., the finance available for the purpose.

Determining Sample Design:

All the items under consideration in any field of inquiry constitute a 'universe' or 'population'. A complete enumeration of all the items in the all 'populations' is known as a census inquiry. It can be presumed that in such an inquiry when the items are covered , no element of chance is left and highest accuracy is obtained. But in practice this may not be true. Even the slightest element of bias in such an inquiry will get larger and larger as the number of observations increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks, Besides, this type of inquiry involves a great deal of time ,money and energy. Not only this , census inquiry is not possible in practice under many circumstances. For instance, blood testing is done only on sample basis. Hence, quite often we select only a few items from the universe for our study purposes. The items so selected constitute what is technically called a sample.

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data is actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city's 200 drugstores in a certain way constitutes a sample design. Samples can be either probability or non – probability samples. With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Sample design includes determination of sampling techniques such as probability sampling ,cluster sampling, stratified sampling, or systematic sampling etc, whereas non-probability samples are those based on convenience sampling , judgment sampling and quota sampling techniques. A brief mention of the important sample designs is as follows:

Random sampling techniques

These are sampling techniques where the items are selected on probability theory. Each unit of the population has the opportunity of being selected into the sample. Random sampling techniques include simple random sampling, stratified sampling, systematic sampling, cluster sampling etc.

Simple random sampling: This type of sampling is also known as chance sampling or probability sampling where each and every item in the population has an equal chance of inclusion in the sample and each one of the possible samples, in case finite universe ,has the same probability of being selected. For example ,if we have to select a sample of 300 items from a universe of 15,000 items, then we can put the names or numbers of all the 15,000 items on slips of paper and conduct a lottery . Using the random number tables is another method of random sampling. To select the sample , each items is assigned a number from 1 to 15000. Then ,300 five digit random numbers are selected from the table. To do this we selected some random starting point and then a systematic pattern is used proceeding through the table. We might start in the 4th row, second column and proceed down the column to the bottom of the table and then move to the top of the numbers in the frame, in our case over 15,000 , it is simply passed over and the next number selected that does fall within the relevant range since the numbers were placed in the table in a completely random fashion, the resulting sample is random ,this procedure gives each item an equal probability

of being selected. In sample is random. This procedure gives each item an equal probability of being selected. In case of infinite population, the selection of each item in a random. This procedure gives each item in a random sample is controlled by the same probability and that successive selections are independent of one another.

Stratified sampling

If the field of enquiry is not homogenous and contains variety y of items, this type of sampling is adopted. In this method, we first divide the population into different sub populations known as stratum, such that items in each stratum are homogeneous. From each stratum, items are selected by simple random sample method. For example, if we want to study the consumption pattern of people in Kerala, the state of Kerala may be divided into a number of zones. From each zone, a sample may be selected at random this method has the advantages of simple random sample method. Stratified sampling method reduces time and expense to a greater extent.

Systematic sampling:

In some instances the most practical way of sampling is to go select every 15th name on list, every 10th house on one side of a street and so on. Sampling of this type is known as systematic sampling. An element of randomness is usually introduced into this kind of sampling by using random numbers to pick up the unit with which to start. This procedure is useful when sampling frame is available in the front of a list. In such a design the selection process start by picking some random points in the list and then every nth element is selected until the desired number is secured.

Cluster sampling

In many surveys, there would be no reliable list of the units of the population or it would be very expensive to collect such a list. In such situations, cluster sampling can be adopted. Cluster sampling consists in forming suitable clusters of units and surveying all the units in a sample of clusters selected. Cluster sampling is used in many surveys as it is very convenient and less costly. Tribal survey can be done by cluster sampling technique. Tribes are generally settled in particular areas. Simple random sampling method will fail here, in such surveys, few settlements are selected as clusters and surveying all the units of the selected cluster may be less time consuming and less costly.

Multi stage sampling

It is further development of the principle of cluster sampling. Suppose we want to investigate the working efficiency of nationalized banks in India and we want to take a sample of few banks for this purpose. The first stage is to select large primary sampling unit such as state or a country. Then we may select certain districts and interview all banks in the chosen district. This would represent a two stage sampling design with the ultimate sampling units being clusters of district.

Ordinarily multi stage sampling is applied in big inquiries extending to a considerable large geographical area, say, the entire country. There are two advantages of this sampling design – viz., it is easier to administer than most single stage designs and large number of units can be sampled for a given cost.

Non random sampling techniques

Samples may be selected not on based on probability, but considering convenience and easiness of selecting sample units. Such non random sampling techniques include deliberate sampling, judgment sampling, convenience sampling, quota sampling etc.

Deliberate sampling:

Deliberate sampling is also known as purposive sampling. When population elements are selected for inclusions in the sample based on the ease of access, it can be called convenience sampling. If a researcher wishes to secure data from, say , gasoline buyers, he may select a fixed number of petrol stations and may conduct interviews at these stations. This would be an example of convenience sample of gasoline buyers. At times such a procedure may give very biased results particularly when the population is not homogeneous.

Judgment sampling

In judgment sampling the researcher's judgment is used for selecting items which he considers as secure reactions to a new method of teaching .Judgment sampling is used quite frequently in qualitative research where the desire happens to develop hypothesis rather than to larger population.

Convenience sampling

The primary consideration in convenience sampling is the convenience of the researcher as to who can provide the best information to achieve the objectives of the study. The researcher only goes to those people who in his opinion are likely to have the required information and be willing to share it. This type of sampling is extremely useful when you want to construct a historical reality, describe a phenomenon or develop something about which only a little is known.

Quota sampling

The main consideration directing quota sampling is the researcher's ease of access to the sample population. In addition to convenience, he is guided by some visible characteristic, such as gender or race, of the study population, that is of interest to him. The sample is selected from a location convenient to the researcher, and whenever a person with this visible relevant characteristic is seen, that person is asked to participate in the study. The process continues until the researcher has been able to contact the required number of respondents (quota)

Collecting the data :

In dealing with any real life problem it is often found that data at hand is inadequate, and hence ,it becomes necessary to collect data that is appropriate. There are several ways of collecting the appropriate data which differ considerably in context of cost , time and other resources at the disposal of the researcher.

Primary data can be collected either through experiment or through survey. If the researcher conducts an experiment, he observes some quantitative measurements, or the data ,with the help of which he examines the truth contained in his hypothesis. But in the case of a survey, data can be collected by any one or more of the following ways.

By observation:

This method implies the collection of information by way investigator's own observation, without interviewing the respondents. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes of respondent. This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

Through personal interview :

The investigator follows a rigid procedure and seeks answer to a set of pre-conceived questions through personal interviews. The method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

Through telephone interviews :

This method of collecting information involves contacting the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in developed regions, particularly, when the survey has to be accomplished in a very limited time.

By mailing of questionnaires:

The researcher and the respondents do come in contact with each other if this method of survey is adopted. Questionnaires are mailed to the respondent with a request to return after completing the same. It is the most extensively used method in various economic and business surveys. Before applying this method, usually a pilot study for testing the conducted which reveals the weakness, if any, of the questionnaire. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information.

Through schedules:

Under this method the enumerators are appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondent with these schedules. Data is collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depend upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

The researcher should select one of these methods of collecting the data taking in to consideration the nature of investigation, objective and scope of the inquiry, financial resources, available time and desired degree of accuracy. Through he should pay attention to all these factors but much depends upon the ability and experience of the researcher. In this context Dr. A. L. Bowley very aptly remarks that in collection of statistical data commonsense is the chief requisite and experience of the chief teacher.

Analysis of data:

After the data has been collected, the researcher turns to the task of analysing them. The analysis of data requires a number of closely related operation such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. The unwieldy data should necessarily be conducted into

a few manageable groups and tables for further analysis. Thus, researcher should classify the raw data into some purposeful and usable categories. Coding operation is usually done at this stage through which the categories of data is transformed into symbols that may be tabulated and counted. Editing is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. Tabulation is a part of technical procedure wherein the classified data is put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, specially in large inquiries, is tabulated by computers. Computers not only save time but also make it possible to study large number of variables affecting a problem simultaneously.

Analysis work after tabulation is generally based on the computation of various percentages, coefficients, etc., by applying various well defined statistical formulate. In the process of analysis, relationships or difference supporting or conflicting with original or new hypothesis should be subjected to tests of significance to determine with what validity data can be said to indicate any conclusion(s). For instance, if there are two samples of weekly wages; each sample being drawn from factories in different parts of the same city, giving two different mean values, then our problem may be whether the two mean values are significantly different or the difference is just a matter of chance. Through the use of statistical tests we can establish whether such a difference is real one or is the result of random fluctuations. If the difference happens to be real, the inference will be that the two samples come from different universe and if the difference is due to chance, the conclusion would be that the two samples belong to the same universe. Similarly, the technique of analysis of variance can help us in analyzing whether three or more varieties of seeds grown on certain fields yield significantly different result or not. In brief, the researcher can analyze the collected data with the help of various statistical measures.

Hypothesis-testing:

After analysing the data as stated above, the researcher is in a position to test the hypothesis, if any, he had formulated earlier. Do the facts support the hypothesis or they happen to be contrary? This is the usual question which should be answered while testing hypothesis. Various tests, such as Chi square test, t-test, F-test, have been developed by statisticians for the purpose. These tests are discussed in detail later in the book. The hypothesis may be tested through the use of one or more of such tests, depending upon the nature and object of research inquiry. Hypothesis-testing will result in either accepting the hypothesis or in rejecting it. If the researcher had no hypothesis to start with, generalizations established on the basis of data may be stated as hypothesis to be tested by subsequent researches in times to come.

Generalization and interpretation:

If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalization, i.e., to build a theory. As a matter of fact, the real value of research lies in its ability to arrive at certain generalizations. If the researcher had no hypothesis to start with, he might seek to explain his findings on the basis of some theory. It is known as interpretation. The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

Preparation of the report or the thesis:

Finally, the researcher has to prepare the report of what has been done by him. Writing of report must be done with great care keeping in view the following factors:

1. The layout of the report should be as follows: the preliminary pages; the main text ,and the end matter.
- 2.In this preliminary pages the report should carry title and date followed by acknowledgements and foreword. Then there should be a table of contents followed by a list of tables and list of illustrations, if any ,given in the report.
- 3.The main text of the report should have the following parts:

Introduction: It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.

Summary of findings: After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarized.

Main report: Then main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.

Conclusion: Towards the end of the main text , researcher should again put down the results of his research clearly and precisely. In fact, it is the final summing up.

At the end of the report, appendices should be enlisted in respect of all technical data. Bibliography, i.e., list of books, journals, reports etc, consulted, should also be given at the end. Index should also be given specially in a published research report.

Report should be written in a concise and objective style in simple language avoiding vague expressions such as 'it seems', 'there may be', and the like.

Charts and illustrations in the main report should be used only if they present the information more clearly and forcibly.

Calculated confidence limits must be mentioned and the various constraints experienced in conducting research operations may as well be stated.

Review questions

1. Explain research process
2. Give a flow chart of research process
3. Explain formulation of research problem
4. What are the steps in formulating research problem
5. Explain the sources of research problem
6. What is review of literature
7. Why literature review is conducted
8. Explain the term hypothesis
9. Distinguish between null hypothesis and alternate hypothesis
10. Enlist steps in formulating hypothesis
11. What is a research design
12. What is the significance of research design
13. What is sample design
14. What factors influence research design

15. What are the objectives of sample design
16. Discuss different sampling techniques

UNIT 7 RESEARCH PROBLEM

In research process , the first and foremost step happens to be that of selecting and properly defining a research problem. A researcher must find the problem and formulate it so that becomes susceptible to research. Like a medical doctor , a researcher must examine all the symptoms (presented to him or observed by him) concerning a problem before he can diagnose correctly . To define a problem correctly, a researcher must know: what a problem is?

What is a research problem?

A research problem, in general, refers to some difficulty which a researcher experiences in the context of either a theoretical or practical situation and wants to obtain a solution for the same. Usually we say that a research problem does exist in the following conditions are met with:

- (i) There must be an individual (or a group or an organization), let us call it 'I', to whom the problem can be attributed. The individual or the organization , as the case may be ,occupies an environment, say 'N' which is defined by values of the uncontrolled variables, Y.
- (ii) There must be at least two courses of action , say C and C ,to be pursued. A course of action is defined by one or more values of the controlled variables. For example, the number of items purchased at specified time is said to be one course of action.
- (iii)There must be at least two possible outcomes, say O and O ,of course of action ,of which one should be preferable to the other. In other words, this means that there must be at least one outcome that the researcher wants , i.e., an objective.
- (iv) The courses of action available must provide some chance of obtaining the objective, but they cannot provide the same chance, otherwise the choice would not matter. Thus ,if $P(O_1/C, N) \neq P(O_2/C, N)$. In simple words , we can say that the choices must have unequal efficiencies for the desired outcomes.

Over and above these conditions , the individual or the organisation can be said to have the problem only if 'I' does not know what courses of action best , i.e., ' I' , must be in doubt about the solution. Thus , an individual or a group of persons can be said to have a problem which can be technically described as a research problem, if they (individual or the group) , having one or more desired outcomes , are confronted with two or more courses of action that have some but not equal efficiency for the desired objective(s) and are in doubt about which course of action is best.

We can, thus, state the components of a research problem as under:

Essential components of research problem

1. There must be an individual or a group which has some difficulty or the problem.

2. There must be some objective(s) to be attained at . If one wants nothing , one cannot have a problem.
3. There must be alternative means or the courses of action, for obtaining the objective , one wishes to attain. This means that there must be at least two means available to a researcher for if he has no choice of means, he cannot have a problem.
4. There must remain some doubt in the mind of a researcher with regard to the selection of alternatives. This means that research must answer the question concerning the relative efficiency of the possible alternatives.

Thus , a research problem is one which requires a researcher to find out the best solution for the given problem ,i.e., to find out by which course of action the objective can be attained optimally in the context of a given environment. There are several factors which may result in making the problem complicated. For instance, the environment may change affecting the efficiencies of the courses of action or the values of the outcomes; the number of alternative courses of action may be very large; persons not involved in making the decision may be affected by it and react to it favorably or unfavorably ,and similar other factors. All such elements , or at least the important ones, may be thought of in context of a research problem.

Selecting the problem

The research problem undertaken for study must be carefully selected. The task is difficult one, although it may not appear to be so. Help may be taken from a research guide in this connection. Nevertheless, every researcher must find out his own salvation for research problems cannot be borrowed. A problem must spring from the researcher's mind like a plant springing from its own seed. If our eyes need glasses, it is not the optician alone who decides about the number of the lens we require. We have to see ourselves and enable him to prescribe for us the right number by cooperating with him. Thus , a research guide can at the most only help a researcher choose a subject. However , the following points may be observed by a researcher in selecting a research problem or a subject for research:

- (i) Subject which is overdone should not be normally chosen, for it will be a difficult task to throw any new light in such a case.
- (ii) Controversial subject should not become the choice of an average researcher.
- (iii) Too narrow or too vague problems should be avoided.
- (iv) The subjects selected for research should be familiar and feasible so that the related research materials or sources of research are within one's reach.

Even then it is quite difficult to supply definitive ideas concerning how a researcher should obtain ideas for his research. For this purpose, a researcher should contact an expert or a professor in the University who is already engaged in research. He may as well read articles published in current literature available on the subject and may think how the techniques and ideas discussed therein might be applied to the solution of other problems. He may discuss with others what he has in mind concerning a problem. In this way he should make all possible efforts in selecting a problem.

The importance of the subject , the qualifications and the training of a researcher, the costs involved, the time factor are few other criteria that must also be considered in selecting a problem . In other words, before the final selection of a problem is done, a researcher must ask himself the following questions:

- Whether he is well equipped in terms of his background to carry out the research?
- Whether the study falls within the budget he can afford?
- Whether necessary cooperation can be obtained from those who must participate in research as subjects?

If the answers to all these questions are in the affirmative, one may become sure so far as the practicability of the study is concerned.

If the subject for research is selected properly by observing the above mentioned points, the research will not be a boring drudgery ,rather it will be love's labour. In fact, zest for work is a must. The subject or the problem selected must involve the researcher and must have an upper most place in his mind so that he may undertake all pains needed for the study.

Necessity of defining the problem

Quite often we all hear that a problem clearly stated is a problem half solved. This statement signifies the need for defining a research problem .The problem to be investigated must be defined unambiguously for that will help to discriminate relevant data from the irrelevant ones. A proper definition of research problem will enable the researcher to be on the track whereas an ill-defined problem may create hurdles. Questions like: What data is to be collected? What characteristic of data are relevant and need to be studied? What relation is to be explored? What techniques are to be used for the purpose? And similar other questions crop up in the mind of the researcher who can well plan his strategy and find answers to all such questions only when the research problem has been well defined. Thus, defining a research problem properly is a prerequisite for any study and is a step of the highest importance. In fact, formulation of a problem is often more essential than its solution. It is only on the careful detailing of the research problem that we can work out the research design and can smoothly carry on all the consequential steps involved while doing research.

Technique involved in defining a problem

Let us start with the question: What does one mean when he/she wants to define a researcher problem? The answer may be that one wants to state the problem along with the bounds within which it is to be studied. In other words, defining a problem involves the task of laying down boundaries within which a researcher shall study the problem with a pre-determined objective in view.

How to define a research problem is undoubtedly a herculean task. However, it is a task that must be tackled intelligently to avoid the perplexity encountered in a research operation. The usual approach is that the researcher should himself pose a question (or in case someone else who wants the researcher to carry on research ,the concerned individual ,organisation or an authority should pose the question to the researcher) and set-up techniques and procedures for throwing light on the question concerned for formulating or defining the research problem. But such an approach generally does not produce definitive

results because the question phrased in such a fashion is usually in broad general terms and as such may not be in a form suitable for testing.

Defining a research problem properly and clearly is a crucial part of a research study and must in no case be accomplished hurriedly. However, in practice this is frequently overlooked which causes a lot of problems later on. Hence, the research problem should be defined in a systematic manner, giving due weightage to all relating points. The technique for the purpose involves the undertaking of the following steps generally one after the other

- : (i) Statement of the problem in a general way
- (ii) Understanding the nature of the problem ;
- (iii) Surveying the available literature;
- iv) Developing the ideas through discussion ; and
- (v) Rephrasing the research problem into a working proposition.

A brief description of all these points will be helpful.

Statement of the problem in a general way

First of all the problem should be stated in a broad general way, keeping in view either some practical concern or some scientific or intellectual interest. For this purpose, the researcher must immerse himself thoroughly in the subject matter concerning which he wishes to pose a problem. In case of social research, it is considered advisable to do some field observation and as such the researcher may undertake some sort of preliminary survey or what is often called pilot survey. Then the researcher can himself state the problem or he can seek the guidance of the guide or the subject expert in accomplishing this task. Often, the guide puts forth the problem in general terms, and it is then up to the researcher to narrow it down and phrase the problem in operational terms. In case there is some directive from an organizational authority, the problem then can be stated accordingly. The problem stated in a broad general way may contain various ambiguities which must be resolved by cool thinking and rethinking over the problem. At the same time the feasibility of a particular solution has to be considered and the same should be kept in view while stating the problem.

Understanding the nature of the problem

The next step in defining the problem is to understand its origin and nature clearly. The best way of understanding the problem is to discuss it with those who first raised it in order to find out how the problem originally came about and with what objectives in view. If the researcher has stated the problem him self, he should consider once again all those points that induced him to make a general statement concerning the problem. For a better understanding of the nature of the problem involved, he can enter into discussion with those who have a good knowledge of the problem concerned or similar other problems. The researcher should also kept in view the environment within which the problem is to be studied and understood.

Surveying the available literature

All available literature concerning the problem at hand must necessarily be surveyed and examined before a definition of the research problem is given. This means that the researcher must be well-conversant with relevant theories in the field, reports and records as also all other relevant literature. He must devote sufficient time in reviewing of

research already undertaken on related problem. This is done to find out what data and other materials, if any, is available for operational purposes. "Knowing what data is available often serves to narrow the problem itself as well as the technique that might be used." This would also help a researcher to know if there are certain gaps in the theories, or whether the existing theories applicable to the problem under study are inconsistent with each other, or whether the finding of the different studies do not follow a pattern consistent with the theoretical expectations and so on. All this will enable a researcher to take new strides in the field for furtherance of knowledge i.e., he can move up starting from the existing premise. Studies on related problems are useful for indicating the type of difficulties that may be encountered in the present study as also the possible analytical shortcomings. At times such studies may also suggest useful and even new lines of approach to the present problem.

Developing the ideas through discussion

Discussion concerning a problem often produces useful information. Various new ideas can be developed through such an exercise. Hence, a researcher must discuss his problem with his colleagues and others who have enough experience in the same area or in working on similar problem. This is quite often known as an experience survey. People with rich experience are in a position to enlighten the researcher on different aspects of his proposed study and their advice and comments are usually invaluable to the researcher. They help him sharpen his focus of attention on specific aspects within the field. Discussion with such persons should not only be confined to the formulation of the specific problem at hand, but should also be concerned with the general approach to the given problem, techniques that might be used, possible solutions, etc.

Rephrasing the research problem

Finally, the researcher must strive to rephrase the research problem into a working proposition. Once the nature of the problem has been clearly understood, the environment (within which the problem has got to be studied) has been defined, discussions over the problem have taken place and the available literature has been surveyed and examined. Rephrasing the problem into analytical or operational terms is not a difficult task. Through rephrasing puts the research problem in as specific terms as possible so that it may become operationally viable and may help in the development of working hypothesis.

An illustration

The technique of defining a problem outlined above can be illustrated for better understanding by taking an example as under:

Let us suppose that a research problem in a broad general way is as follows:

"Why is productivity in Japan so much higher than in India?"

In this problem, the question has a number of ambiguities such as : What sort of productivity is being referred to ? With what industries the same is related? With what period of time the productivity is being talked about? In view of all such ambiguities the given statement or the question is much so too general to be amenable to analysis. Rethinking and discussions about the problem may result in narrowing down the question

"What factors were responsible for the higher labour productivity of Japan's manufacturing industries during the decade 1971 to 1980 relative to India's manufacturing industries?".

This latter version of the problem is definitely an improvement over its earlier version for the various ambiguities have been removed to the extent possible. Further rethinking and rephrasing might place the problem on a still better operational basis as shown below:" To what extent did labour productivity in 1971 to 1980 in Japan exceed that of India in respect of 15 selected manufacturing industries? What factors were responsible for the productivity differentials between the two countries by industries?

With this sort of formulation, the various term involved such as 'labour productivity', productivity differentials', etc. must be explained clearly. The researcher must also see that the necessary data is available. In case the data for one or more industries selected is not available for the concerning time-period, then the said industry or industries will have to be substituted by other industry or industries. The suitability of the time-period must also be examined. Thus, all relevant factors must be considered by a researcher before finally defining a research problem.

We may concluded by saying that the task of defining a research problem, very often, follows a sequential pattern—the problem is stated in a general way, the ambiguities are resolved, thinking and rethinking process result in more specific formulation of the problem so that it may be a realistic one in terms of the available data and resources and is also analytically meaningful. All this results in a well defined research problem that is not only meaningful from an operational point of view, but is equally capable of paving the way for the development of working hypotheses and for means of solving the problem itself.

Review Questions

1. Describe briefly the techniques of defining a research problem.
2. What is research problem
3. Define the main issues which should receive the attention of the researcher in formulating the research
4. How do you define a research problem
5. Give three examples of research problems
6. What is the necessity of defining research problem? Explain.
7. Write short notes on :
 - Experience survey;
 - Pilot survey;
 - Component of a research problem;
 - Rephrasing the research problem.
8. "The task of defining the research problem often follows a sequential pattern". Explain.
9. "Knowing what data are available often serves to narrow down the problem itself as well as the technique that might be used". Explain the underlying idea in this statement in the context of defining a research problem.
10. Write a comprehensive note on the "Task of defining a research problem".
11. Specify essential components of a research problem

12. What are the steps in selecting a research problem
13. Why literature to be surveyed for problem formulation
14. What are the qualities of researchable problem
15. "A well formulated problem is half solved" Explain

18.1.4

18.1.5

UNIT 8 LITERATURE REVIEW

Introduction

One of the essential preliminary tasks when you undertake a research study is to go through the existing literature in order to acquaint yourself with the available body of knowledge in your area of interest. The literature review is an integral part of the entire research process and makes a valuable contribution to almost every operational step.

Review of literature

Review of literature refers to the process of scanning past literature, journal articles, earlier researches, and all other published sources, for assessing what has been done so far, on the topic or related topic.

It has value even before the first step; that is, when you are merely thinking about a research question that you may want to find answers to through your research journey. In the initial stages of research it helps you to establish the theoretical roots of your study, clarify your ideas and develop your methodology, but later on the literature review serves to enhance and consolidate your knowledge base and helps you to integrate your finding with the existing body of knowledge. Since an important responsibility in research is to compare your findings with those of others, is it here that the literature review plays an extremely important role. During the write-up of your report it helps you to integrate your findings with existing knowledge – that is, to either support or contradict earlier research. The higher the academic level of your research. The higher the academic level of your research, the more important a thorough integration of your findings with exiting literature becomes.

Significance of literature review in research

Reviewing literature can be time-consuming, daunting and frustrating, but it is also rewarding. A literature review has a number of functions:

1. It provides a theoretical background to your study.

The literature review involves a paradox. On the one hand, you cannot effectively undertake a literature search without some idea of the problem you wish to investigate. On the other hand, the literature review can play an extremely important role in shaping your research problem because the process of reviewing the literature helps you to understand the subject area better and thus helps you to understand the relationship between your research problem and the body of knowledge in the area.

2. It Improves your methodology

It reviews the means by which you establish the links between what you are proposing to examine and what has already been studied. In other words, it helps you to refine your research methodology. Going through the literature acquaints you with the methodologies that have been used by others to find answers to research questions similar

to the one you are investigating. A literature review tells you if others have used procedures and methods similar to the ones that you are proposing , which procedures and methods have worked well for them, and what problem they have faced with them. By becoming aware of any problems and pitfalls, you will be better positioned to select a methodology that is capable of providing valid answers to your research questions. This will increase your confidence in the methodology you plan to use and will equip you to defend its use

3. It contributes to stock of knowledge.

Through the literature review you are able to show how your findings have contributed to the existing body of knowledge in your profession. By conducting research it is expected to acquire new knowledge or modify existing knowledge. It is through literature review, that one can assess, what is current knowledge, and what is lacking.

4. It enables you to contextualize your findings.

Obtaining answers to your research questions is comparatively easy; the difficult part is examining how your practical findings fit into the existing theory and concepts. How do answers to your research questions compare with what others have found? How are your findings different from those of others? For you to be able to answer these questions you need to go back to your literature review. It is important to place your findings in the context of what is already known in your field of inquiry.

5. It broadens your knowledge base in your research area

The most important function of the literature review is to ensure you read widely around the subject area in which you intend to conduct your research study. It is important that you know what other researchers have found in regard to the same or similar questions , what theories have been put forward and what gaps exist in the relevant body of knowledge. When you undertake a research project for a higher degree(that is, an MA or a PhD) you are expected to be an expert in your area of study. A thorough literature review helps to ensure that you fulfil this expectation. Another important reason for doing a literature review is that it helps you to understand how the findings of your study fit into the existing body of knowledge (Martin 1985:30)

6. It enhances clarity and focus.

It also helps you to bring clarity and focus to your research problem. It improves your methodology, broaden your knowledge base in your research area and bring proper direction to research . It helps you to understand new relationships between your research problem and the body of knowledge in the area.

Procedure for reviewing the literature

If you do not have a specific research problem, you should review the literature in your broad area of interest with the aim of gradually narrowing down to what you want to find out about. After that the literature review should be focused around your research problem. There is a danger in reviewing the literature without having a reasonably specific idea of what you want to study. It can condition your thinking about your study and the methodology you might use, resulting in a less innovative choice of research problem and methodology than otherwise would have been the case. Hence ,you should try to conceptualize your research problem before undertaking your major literature review.

Steps in review of literature

There are four steps involved in conducting a literature review: Search for existing literature in your area of study, review the literature selected, develop a theoretical framework and develop a conceptual framework.

The skills required for these tasks are different. Developing theoretical and conceptual frameworks are more difficult than the other tasks.

Search for existing literature

To effectively search for literature in your field of inquiry, it is imperative that you have in mind at least some idea of the broad subject area and of the problem you wish to investigate, in order to set parameters for your search. Next compile a bibliography for this broad area. There are two sources that you can use to prepare a bibliography: books and journals.

The best way to search for a book is to look at your library catalogues. When librarians catalogue a book they also assign to it subject headings that usually are based on Library of Congress Subject Heading. If you are not sure, ask your librarian to help you to find the best subject heading for your area. This can save you a lot of time. Publications such as Book Review Index can help you to locate books of interest.

There are several sources designed to make your search for journals easier and these can save you enormous time. They are:

- 1 Indices of journals(e.g Humanities Index);
- 2 Abstract of articles(e.g. ERIC);
- 3 Citation indices (e.g. Social Sciences Citation Index).

All the above indexing, abstracting and citation services are available in print or on CD-ROM, or are stored on a mainframe computer accessible through the Internet, a world-wide electronic communication system.

In most libraries, information on books, journals, abstracts and so on is stored on computers and CD-ROMs. In each case the information is classified by subject, author and title. You may also have the keywords option (author/keyword; title/keyword; expert/keyword; or just keywords). What system you use depends upon what is available in your library and what you are familiar with.

There are specially prepared electronic databases in a number of disciplines. These can also be helpful in preparing a bibliography. For example, most libraries carry the electronic data bases shown in Table

Books

Books, though a central part of any bibliography, have their advantages as well as disadvantages. The main advantage is that the material published in books is usually important and of good quality, and the findings are 'integrated with other research to form a coherent body of knowledge' (Martin 1985: 33). The main disadvantage is that the material is not completely up to date, as it can take a few years between the completion of a work and its publication in the form of a book.

The best way to identify books is to use a computer catalogue. Use the subject catalogue or keywords option to search for book in your area of interest. Narrow the subject area searched by selecting the appropriate keywords. Look through these titles carefully and

identify the books you think are likely to be of interest to you. If you think the title seem appropriate to your topic, print them down on a piece of paper. Be aware that sometimes a title does not provide enough information to decide if a book is going to be of use. To make this decision you may have to search for such books in the library and examine their contents.

When you have selected 10-15 books that you think are appropriate for your topic, examine the bibliography of each. It will save time to photocopy their bibliographies. Go through these bibliographies carefully to identify the books common to several of them. If a book has been referenced by a number of authors, you should include it in your reading list. Prepare a final list of books that you consider essential reading.

Having prepared your reading list, locate these books in your library or borrow them from other sources. Examine their contents to double-check that they really are relevant to your topic. If you find that a book is not relevant to your research, delete it from your reading list. If you find that something in a book's contents is relevant to your topic, make an annotated bibliography. An annotated bibliography contains a brief abstract of the aspects covered in a book and your own notes of its relevance. Be careful to keep track of your references. To do this you can prepare your own card index or use a computer program such as Endnotes or Pro-Cite.

Journals

In the same way, you need to go the journals in your research area. Journals provide you with the most up-to-date information, even though there is often a gap of two to three years between the completion of a research project and its publication in a journal. You should select as many journals as you possibly can, though the number of journals available depends upon the field of study- certain fields have more journals than others. As with books, you need to prepare a list of the journals you want to examine for identifying literature relevant to your study. This can be done in a number of ways. You can :

1. locate the hard copies of the journals that are appropriate to your study ;
2. look at citation or abstract indices to identify and/or read the abstracts of such articles;
- search electronic databases;
3. use the Internet

Whichever method you choose, first identify the journals that you want to look at in more detail for your review of the literature. The next step is to make preparations to go through them.

If you have been able to identify any useful journals and articles, prepare a list of those you want to examine, by journal. Select one of these journals and, starting with the latest issue, examine its contents page to see if there is an article of relevance to your research topic. If you feel that a particular article is of interest to you, read its abstract. If you think you are likely to use it, depending upon your financial resources, either photocopy it, or prepare a summary and record its reference for later use.

Review the literature selected

Now that you have identified several books and as useful, the next step is to start reading them critically to pull together themes and issue that are associated. If you do not

have theoretical frame work of themes in mind to start with use separate sheets of paper for each article or book. Once you develop a rough framework, using a separate sheet of paper sheet of paper for each theme of that framework . As you read further , go on slotting the information where it logically belongs under the themes so far developed. You may need to add more themes as you go. In doing so, read critically with particular reference to the following aspects:

1. Note whether the knowledge relevant to your theoretical framework has been confirmed beyond doubt.
2. Note the theories put forward, the criticism of these and their basis, the methodologies adopted(study design, sample size and its characteristics, measurement procedures, etc.) and the criticisms of them.
3. Examine to what extent the findings can be generalized to other situations.
4. Notice where there are significant differences of opinion of opinion among researcher and give your opinion about the validity of these differences.
5. Ascertain the areas in which little or nothing is known- the gaps that exist in the body of knowledge.

Develop a theoretical frame

Examining the literature can be a never-ending task, but as you have limited time it is important to set parameters by reviewing the literature in relation to some main themes pertinent to your research topic. As you start reading the literature, you will soon discover that the problem you wish to investigate has its roots in a number of theories that have been developed from different perspectives. The information obtained from different books and journals now needs to be sorted under the main themes and theories, highlighting agreements and disagreements among the authors and identifying the unanswered questions or gaps. You will also realize that the literature deals with a number of aspects that have a direct or indirect bearing on your research topic. Use these aspects as a basis for developing your theoretical frame work. Unless you review the literature inn relation to this frame work, you will not be able to develop a focus in your literature search: that is, your theoretical framework provides you with a guide as you read. This brings us to the paradox mentioned previously: until you go through the literature you cannot develop a theoretical framework and until you have developed a theoretical framework, you cannot effectively review the literature. The solution is to read some of the literature then attempt to develop a framework, even a loose one, within which you can organize the rest of the literature you read. As you read more about the area, you are likely to change the framework. However, without it, you will get bogged down in a great deal of unnecessary reading and note-taking that may not be relevant to your study.

Literature pertinent to your study may deal with two types of information:

- 1 Universal (global trends)
- 2 Specific (local trends or a specific program).

In writing about such information you should start with the general information, gradually narrowing it down to the specific.

Look at the example in Figure 3.1.

Figure 3.1 developing a theoretical framework-the relationship between mortality and fertility

If you want to study the relationship between mortality and fertility, you should review literature about:

Fertility – trends, theories, some of the indices and critiques of them, factors affecting fertility, methods of controlling fertility, factors affecting acceptance of contraceptive , and so on;

Mortality – factors affecting mortality, mortality indices and their sensitivity in measuring change in mortality levels of a population, trends in mortality, and so on; and , most importantly

The relationship between fertility and mortality –theories that have been put forward to explain the relationship, and implications of the relationship.

Out of this literature review you need to develop the theoretical framework for your study. Primarily this should revolve around theories that have been put forward about the relationship. For example, it has been explained from economic, religious, medical and psychological perspectives. Within each perspective several theories have been put forward : 'insurance theory', 'fear of non-survival', 'replacement theory', 'price theory', 'utility theory', 'extra' or 'hoarding theory' and 'risk theory'.

Your literature review should be written under the following headings, with most of the review involving the examination of the relationships between fertility and mortality:

Fertility theories;

- 1.The theory of demographic transition;
- 2.Trends in fertility global , and then narrow it to national and local levels
- 3.Methods of contraception , their acceptance and effectiveness
- 4.Factors affecting mortality;
- 5.Trends in mortality and their implications
- 6.Measurement of mortality indices and their sensitivity
- 7.Relationship between fertility and mortality - different theories such as 'insurance', 'fear of non-survival', 'replacement', 'price' 'utility', 'risk'

Develop a conceptual framework

The conceptual framework stems from the theoretical frame work and concentrates, usually, on one section of that theoretical framework which becomes the basis of your study. The latter consist of the theories or issues in which your study is embedded, whereas the former describes the aspects your selected framework, to become the basis of your inquiry. The conceptual framework is the basis of your research problem. For instance, in the example cited in Figure 3.1, the theoretical framework includes all the theories that have been put forward to explain the relationship between fertility and mortality . However, out of these, you may be planning to test only one, say, the fear of non-survival. Hence the conceptual framework grows out of the theoretical framework and relates to the specific research problem concerning the fear of non-survival theory.

Writing up the literature reviewed

Now all ,that remains to be done is to write about the literature you have reviewed. As mentioned in the beginning of this chapter, the broad two functions of a literature review are

1. To provide a theoretical background to your study and
2. To enable you to contextualize your findings in relation to the existing body of knowledge in addition to refining your methodology.

The content of your literature review reflects these two purposes. In order to fulfill the first purpose, you identify and describe various theories relevant to your field; and specify gaps in existing knowledge in the area, recent advances in the area of study , current trends and so on. In order to comply with the second function you integrate your results with specific and relevant findings from the existing literature by comparing the two for confirmation or contradiction.

While reading the literature for theoretical background back ground of your study , you will realize that certain themes have emerged. List the main ones, converting them into subheadings. These subheadings should be precise, descriptive of the theme in question, and follow a logical progression. Now, under each subheading, record the main findings with respect to the theme in question, highlighting the reasons for and against an argument if they exist, and identifying gaps and issues.

Some people write up the entire literature review', without subheadings. The author strongly suggests that you write your literature review under subheadings.

The second broad function of the literature review-contextualizing the findings of your study-requires you to very systematically compare your findings with those made by others. Quote from these studies to show how your findings contradict, confirm or add to them. It places your findings in the context of what other have found out. This function is undertaken when writing about your findings, that is, after analysis of your data.

Almost every great step in the history of science has been made by the 'anticipation of nature' , that is , by the invention of hypotheses which, though verifiable , often had very little foundation to start with .

Review Questions

1. What is meant by review of literature
2. What is the purpose of literature review
3. State the significance of review of literature
4. Enlist the procedure in literature review
5. How does literature review broaden your knowledge
6. How does literature review improve research methodology
7. What is the need for literature review
8. Is Internet helpful in literature review
9. What are the aspects to be noted in literature survey
10. What is the role of journal in literature review
11. State the procedure in writing the literature reviewed
12. How does literature review bring clarity and focus to research

NIT 9 DEVELOPMENT OF HYPOTHESIS

Introduction

The second important consideration in the formulation of a research problem is the construction of hypotheses. Hypotheses bring clarity, specificity and focus to a research problem, but are not essential for a study. You can conduct a valid investigation without constructing a single formal hypothesis. On the other hand, within the context of a research study, you can construct as many hypotheses as you consider to be appropriate. Some believe that one must formulate a hypothesis to undertake an investigation.

Hypotheses primarily arise from a set of 'hunches' or assumptions that are tested through a study and one can conduct a perfectly valid study without having these hunches or speculations. However, in epidemiological studies, to narrow the field of investigation, it is important to formulate hypotheses.

Importance of hypothesis

The importance of hypotheses lies in their ability to bring direction, specificity and focus to a research study. They tell a researcher what specific information to collect, and thereby provide greater focus.

Let us imagine you are at the races and you place a bet. You bet on a hunch that a particular horse will win. You will only know if your hunch was right after the race. Take another example. Suppose you have a hunch that there are more smokers than non-smokers in your class. To test your hunch, you ask either all or just some of the class if they are smokers. You can then conclude whether your hunch was right or wrong.

Now let us take a slightly different example. Suppose you work in the area of public health. Your clinical impression is that a higher rate of a particular condition prevails among people coming from a specific population subgroup. You want to find out the probable cause of this condition. There could be many causes. To explore every conceivable possibility would require an enormous amount of time and resources. Hence, to narrow the choice, based on your knowledge of the field, you could identify what you assume to be the most probable cause. You could then design a study to collect the information needed to verify your hunch. If on verification you were able to conclude that the assumed cause was the real cause of the condition, your assumption would have been right.

In these examples, you started with a superficial hunch or assumption. In one case (horse racing) you waited for the event to take place and in the other two instances you designed a study to assess the validity of your assumption, and only after careful investigation did you reach a conclusion about the validity of your assumptions.

Hypotheses are based upon similar logic. As a researcher, you did not know about a phenomenon, a situation, the prevalence of a condition in a population or about the outcome of a program, but you do have a hunch to form the basis of certain assumptions or guesses. You test these by collecting information that will enable you to conclude if your hunch was right. The verification process can have one of the three outcomes. Your hunch may prove to be : right , partially right , or wrong.

Without this process of verification, you cannot conclude anything about the validity of your assumption.

Definition of hypothesis

Hence, a hypothesis is a "hunch, assumption, suspicion, assertion or an idea about a phenomenon, relationship or situation, the reality or truth of which you do not know". A researcher calls these assumptions, assertions, statements or hunches hypotheses and they become the basis of an inquiry. In most studies the hypothesis will be based upon either previous studies or on your own or someone else's observation.

There are many definitions of a hypothesis. According to Kerlinger, 'A hypothesis is a conjectural statement of the relationship between two or more variables'.

Webster's New International Dictionary of English Language defines a hypothesis as :

"proposition, condition, or principle which is assumed, perhaps without belief, in order to draw out its logical consequences and by this method to test its accord with facts which are known or may be determined."

Black and Champion define a hypothesis as, 'a tentative statement about something, the validity of which is usually unknown'. In another definition, Bailey defines a hypothesis as :

"A proposition that is stated in a testable form and that predicts a particular relationship between two or more variables."

In other words, if we think that a relationship exists between two or more variables to be studied, we first state it as a hypothesis and then test the hypothesis in the field (1978:35)

According to Grinnell and Stothers, "A hypothesis is written in such a way that it can be proven or disproven by valid and reliable data—it is in order to obtain these data that we perform our study. From the above definitions it is apparent that a hypothesis has certain characteristics:

- 1 It is a tentative proposition.
- 2 Its validity is unknown.
- 3 In most cases, it specifies a relationship between two or more variables.

Functions of hypothesis

While some researchers believe that to conduct a study requires a hypothesis, having a hypothesis is not essential as already mentioned. However, a hypothesis is important in terms of bringing clarity to the research problem.

Specifically, a hypothesis serves the following functions.

1. Focus -The formulation of a hypothesis provides a study with focus. It tells you what specific aspects of a research problem to investigate.
2. Discretion - A hypothesis tells you what data to collect and what not to collect, thereby providing discretion to the study.
3. Objectivity - it provides a direction and objective for the study. The construction of a hypothesis enhances objectivity in a study.
4. A hypothesis may enable you to add to the formulation of theory. It enables you to specifically conclude what is true or what is false.'

Characteristics of hypothesis

There are a number of considerations to keep in mind when constructing a hypothesis, as they are important for valid verification.

1. A hypothesis should be simple, specific and conceptually clear.

There is no place for ambiguity in the construction of a hypothesis, as ambiguity will make the verification of your hypothesis almost impossible. It should be 'one-dimensional'- that is, it should rest only one relationship or hunch at a time. To be able to develop a good hypothesis you must be familiar with the subject area. The literature review is of immense help. The more insight you have into a problem, the easier it is to construct a hypothesis.

For example: "The average age of the male students in a class is higher than that of the female students."

The above hypothesis is clear, specific and easy to test. It tells you what you are attempting to compare. - Average age of this class, which population groups are being compared -female and male students, What you want to establish higher average age of the male students etc.

Let us take another example : 'Suicide rates vary inversely with social cohesion'

This hypothesis is clear and specific, but a lot more difficult to test. There are three aspects of this hypothesis: 'suicide rates'; 'vary inversely', which stipulates the direction of the relationship; and 'social cohesion'. To find out the suicide rates and to establish the relationship is inverse or otherwise are comparatively easy, but to ascertain social cohesion is a lot more difficult. What determines social cohesion? How can it be measured? This problem makes it more difficult to test this hypothesis.

2.A hypothesis should be capable of verification.

Methods and techniques must be available for data collection and analysis. There is no point in formulating a hypothesis if it cannot be subjected to verification because there are no techniques to verify it. However, this does not necessarily mean that you should not formulate a hypothesis for which there are no methods of verification. You might, in the process of doing your research, develop new techniques to verify it.

3.A hypothesis should be related to the existing body of knowledge.

It is important that your hypothesis emerges from the existing body of knowledge, and that it adds to it, as this is an important function of research. This can only be achieved if the hypothesis has its roots in the existing body of knowledge.

4.A hypothesis should be operationalisable.

This means that it can be expressed in terms that can be measured. If it cannot be measured, it cannot be tested and, hence, no conclusions can be drawn. It should be practicable , valid and stable.

Types of hypothesis

As explained, any assumption that you seek to validate through an inquiry is called a hypothesis. Hence, theoretically there should be only one type of hypothesis, that is, the research hypothesis-the basis of your investigation. However , because of the convention in scientific inquiries and because of the wording used in the construction of a hypothesis, hypotheses can be classified into several types. Broadly, there are two categories of hypothesis:

1 Research hypothesis;

Ordinarily, a hypothesis is simply an assumption to be proved. But for a research hypothesis is a formal question that he intends to resolve. Quite often a research hypothesis is a predictive statement, capable of being tested by scientific methods. These are hypotheses which can be objectively and statistically verified and tested. Generally, research hypothesis are accepted when proved right.

2 Alternate hypothesis.

The formulation of an alternate hypothesis is a convention in scientific circles. Its main function is to explicitly specify the relationship that will be considered as true in case the research hypothesis proves to be wrong. In a way, an alternate hypothesis is the opposite of the research hypothesis. Again, conventionally, a null hypothesis, or hypothesis of no difference, is formulated as an alternate hypothesis.

Let us take an example. Suppose you want to test the effect of different combinations of maternal and child health services (MCH) and nutrition supplements (NS) have on the infant mortality rate. To test this, an alternate hypothesis may be formed as "Lack of Nutrition leads to higher infant mortality rate."

Formulating working hypothesis

Working hypothesis means hypothesis formulated for specific testing and study, as a part of the research process. Such working hypothesis may be Null hypothesis, Hypothesis of difference, Hypothesis of point prevalence, or Hypothesis of association. There are several ways of formulating practical hypotheses. For example:

1. There will be no difference in the level of infant mortality among the different treatment modalities.
2. The MCH and NS treatment group will register a greater decline in infant mortality than the MCH, The NS treatment or the control group.
3. Infant mortality in the MCH treatment group will reach a level of 30/1000 over five years.
4. Decline in the infant mortality rate will be three times greater in the MCH treatment group than in the NS one over five years.

Let us take another example. Suppose you want to study the smoking pattern in a community in relation to gender differentials. The following hypotheses could be constructed.

1. There is no significant difference in the proportion of male and female smokers in the study population.
2. A greater proportion of females than males are smokers in the study population.
3. A total of 60 per cent of females than males are smokers in the study population.
4. There are twice as many female smokers as male smokers in the study population.

In both sets of examples, the way the first hypothesis has been formulated indicates that there is no differences either in the extent of the impact of different treatment modalities on the infant mortality rate or in the proportion of male and female smokers. When you construct a hypothesis stipulating that there is no difference between two situations, groups, outcomes ,or the prevalence of a condition or phenomenon, this is called a null hypothesis and is usually written as H_0 .

The second hypothesis in each example implies that there is a difference either in the extent of the impact of different treatment modalities on infant mortality or in the proportion of male and female smokers among the population, though the extent of the difference is not specified. A hypothesis which specifies its magnitude is called a **hypothesis of difference**.

A researcher may have enough knowledge about the smoking behavior of the community or the treatment program and its likely outcomes to speculate almost the exact prevalence of the situation or the outcomes of a treatment program in quantitative units. Examine the third hypothesis in both sets of examples: the level infant mortality is 30/1000 and the proportion of female and male smokers is 60 and 30 per cent respectively. This type of hypothesis is known as a **hypothesis of point-prevalence**.

The fourth hypothesis in both sets of examples speculates a relationship between the impact of different combinations of MCH and NS programs on the dependent variable (infant mortality) or the relationship between the relevance of a phenomenon(smoking) among different populations(male and female). This type of hypothesis stipulates the extent of the relationship in terms of the effect of different treatment groups on the dependent variable('three times greater in the MCH treatment group than in the NS on over five years') or the prevalence of a phenomenon in different population groups('twice as many female as male smokers'). This type of hypothesis is called a **hypothesis of association**.

There may be some confusion about null and research hypotheses, as in figure 6.3 the null hypothesis is also classified as hypothesis of no difference under 'research hypothesis'. Any type of hypothesis, including a null hypothesis, can become the basis of an inquiry. When a null hypothesis becomes the basis of an investigation, it becomes a research hypothesis.

Procedure in hypothesis testing

To test a hypothesis means to tell whether or not the hypothesis seems to be valid. In hypothesis testing the basic question is whether to reject the null hypothesis or not to reject null hypothesis. Various steps in hypothesis testing are :

1. Setting up hypothesis

This consists of making a formal statement of null hypothesis and also of alternate hypothesis. Setting up of hypothesis is an important step which must be followed in accordance with the object and nature of the problem.

2. Selecting significance level

Hypotheses are tested on a predetermined level of significance and as such the same should be specified. Generally, in practice, either 5% or 1% level is adopted for the purpose. Level of significance is determined considering size of sample, variability of measurements, specificity of the measurement, nature of problem etc.

3. Selecting test statistic

Hypotheses are examined based on means, proportions, variances and standard deviations etc. Test statistic is selected considering the values available and accuracy level desired. Test statistic is obtained using sample observation, selected by the researcher, and hypothetical parameter value stated under null hypothesis.

4. Critical value

Using the sample distribution, test statistic and level of significance, we obtain critical value, with which we compare the difference between two tested concepts – usually sample value and hypothetical population value.

5. Decision

Comparing the value of test statistic and critical value, we make decision about rejecting or not rejecting null hypothesis. We reject null hypothesis when value of test statistic is lower than critical value. When it is greater we will accept null hypothesis.

Errors in testing a hypothesis

As already mentioned, a hypothesis is an assumption that may prove to be either correct or incorrect. It is possible to arrive at an incorrect conclusion about a hypothesis for a variety of reasons. Incorrect conclusions about the validity of a hypothesis may be drawn if:

1. The study design selected is faulty;
2. The sampling procedure adopted is faulty;
3. The method of data collection is inaccurate;
4. The analysis is wrong;
5. The statistical procedures applied are inappropriate; or
6. The conclusions drawn are incorrect.

Any, some or all of these aspects of the research process could be responsible for the inadvertent introduction of error in your study, making conclusions misleading. Hence, in the testing of a hypothesis there is always the possibility of errors attributable to the reasons identified above. Figure 6.4 shows the types of error that can result in the testing of a hypothesis.

Review Questions

1. Briefly explain hypotheses
2. How does hypothesis bring clarity to research problem
3. How does hypothesis arise
4. Define hypothesis
5. Give a classification of hypothesis
6. Explain significance of hypothesis
7. Enlist functions of hypothesis
8. What are the characteristics of hypothesis
9. Give two examples of hypothesis
10. What is research hypothesis
11. What is null hypothesis
12. What is alternate hypothesis
13. Explain formulation of hypothesis
14. What is hypothesis of point prevalence
15. State procedure in hypothesis formulation
16. State steps in hypothesis testing
17. Explain errors in hypothesis testing
- 18.

UNIT 10 RESEARCH DESIGN

Introduction

The formidable problem that follows the task of defining the research problem is the preparation of a design of the research project, popularly known as "research design". A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure. In fact, the research design is the conceptual structure within which research conducted; it constitutes the blue print for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to the final analysis of data. More explicitly, the design decisions happen to be in respect of :

- What is the study about?
- Why is the study being made?
- Where will the study be carried out?
- What type of data is required?
- Where can the required data be found?
- What periods of time will the study include?
- What will be the sample design?
- What techniques of data collection will be used?
- How will the data be analysed?
- In what style will the report be prepared?

Components of research design

A research design has no formal format. But almost all research designs involve following components. Keeping in view the above stated design decisions, one may split the overall research design into the following parts:

1. The sampling design which deals with the method of selecting items to be observed for the given study.
 2. The observational design which relates to the conditions under which the observations are to be made.
 3. The statistical design which concerns the question of how many items are to be observed and how the information and data gathered is to be analysed.
 4. The operational design which deals with the techniques by which the procedures specified in the sampling, statistical and observational design can be carried out.
- From what has been stated above, we can state the important features of a research design as under:

Features of Research Design

Research Design is to be characterized by the following aspects.

1. Research Design is a plan that specifies the sources and types of information relevant to the research problem.
2. It is a strategy which specifies the sources and types of information relevant to the research problem.
3. It sets out the time and cost budgets since most studies are done under these two constraints.

4. It must clearly make a statement of the research problem.
5. It must specify the population to be studied and the procedures of sampling, if any.
6. It must hint about methods to be used in processing and analysing data.

Need for Research Design

Practically research process can be undertaken and completed without a formal research design. But in all researches, research design appropriate to the study is employed due to following aspects.

- Research design is needed because it facilitates the smooth sailing of the various research operations, thereby making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money. Just as better, economical and attractive construction of a house, we need a blueprint (or what is commonly called the map of the house) well thought out and prepared by an expert architect.
- Similarly, research design or a plan is unavoidable, in advance of data collection and analysis for our research project. Research design stands for advance planning of the methods to be adopted for collecting the relevant data and the techniques to be used in their analysis, keeping in view the objective of the research and the availability of staff, time and money.
- Preparation of the research design should be done with great care as any error in it may upset the entire project.
- Research design, in fact, has a great bearing on the reliability of the results arrived at and as such constitutes the firm foundation of the entire edifice of the research work.

Sometimes, the need for a well thought out research design is, at times is not realised by many. The importance which this problem deserves is not given to it. As a result, many researches do not serve the purpose for which they are undertaken. In fact, they may even give misleading conclusions. Thoughtlessness in designing the research project may result in rendering the research exercise futile. It is, therefore imperative that an efficient and appropriate design must be prepared before starting research operations. The design helps the researcher to organize his ideas in a form whereby it will be possible for him to look for flaws and inadequacies. Such a design can even be given to others for their comments and critical evaluation. In the absence of such a course of action, it will be difficult for the critic to provide a comprehensive review of the proposed study.

Features of a good Research Design

1. A good design is often characterised by adjectives like flexible, appropriate, efficient, economical and so on.
2. Generally, the design which minimises bias and reliability of the data collected and analysed is considered a good design.
3. The design which gives the smallest experimental error is supposed to be the best design.

4. Similarly ,a design which yields maximal information and provides an opportunity for considering many different aspects of a problem is considered most appropriate and efficient.
5. The question of good design is related to the purpose or objective of the research problem and also with the nature of the problem to be studied.

However, it is to be noted that a design may be quite suitable in one case, but may be found wanting in one respect or the other in the context of some other research problem. One single design cannot serve the purpose of all types of research problems.

Factors in determining Research Design

A research design appropriate for a particular research problem, usually involves the consideration of the following factors:

- (i) The methods of obtaining information;
- (ii) The availability and skills of the researcher and his Staff, if any;
- (iii) The objective of the problem to be studied;
- (iv) The nature of the problem to be studied ; and
- (v) The availability of time and money for research work.

Appropriateness of Research Design

If the research study happens to be an exploratory or a formulative one, wherein the major emphasis is on discovery of ideas and insights, the research design most appropriate must be flexible enough to permit the consideration of many different aspects of a phenomenon.

But when the purpose of a study is accurate description of situation or of an association between variables(or in what are called the descriptive studies), accuracy becomes a major consideration and a research design which minimises bias and maximizes the reliability of the evidence collected is considered a good design.

Studies involving the testing of a hypothesis of a causal relationship between variables require a design which will permit inferences about causality in addition to the minimisation of bias and maximization of reliability.

But in practice it is the most difficult task to put a particular study in a particular group, for a given research may have in it elements of two or more of the functions of different studies. It is only on the basis of its primary function that a study can be categorised either as an exploratory or descriptive or hypothesis -testing study and accordingly the choice of a research design may be made in case of a particular study. Besides , the availability of time ,money ,skills of the research staff and the means of obtaining the information must be given due weight age while working out relevant details of the research design such as experimental design, survey design ,sample design and the like.

Important concepts relating to research designs

Before describing the different research design, it will be appropriate to explain the various concepts relating to designs so that these may be better and easily understood.

Dependent and Independent Variables

A concept which can take on different quantitative values is called a variable. As such the concepts like weight, height, income are all examples of variables. Qualitative phenomena (or the attributes) are also quantified on the basis of the presence or absence of the concerning attribute(s). Phenomena which can take on quantitatively different values even in decimal points are called 'continuous variables'. But all variables are not continuous. If they can only be expressed in integral values, they are non-continuous variables or in statistical language 'discrete variables'. Age is an example of continuous variable, but the number of children is an example of non-continuous variable. If one variable depends upon or is a consequence of the other variable it is termed as an independent variable, and the variable that is antecedent to the dependent variable it is termed as an independent variable. For instance, if we say that height depends upon age, then height is a dependent variable and age is an independent variable. Further, if in addition to being dependent upon age, height also depends upon the individual's sex, then height is a dependent variable and age and sex are independent variables. Similarly, readymade films and lectures are examples of independent variables. Behavioral changes, occurring as a result of environment changes are of dependent variables.

Extraneous variable

Independent variables that are not related to the purpose of the study, but may affect the dependent variable are termed as extraneous variables. Suppose the researcher wants to test the hypothesis that there is a relationship between children's gains in social studies achievement and their self concepts. In this case self-concept is an independent variable and social studies achievement is a dependent variable. Intelligence may as well affect the social studies achievement, but since it is not related to the purpose of the study undertaken by the researcher, it will be termed as an extraneous variable. Whatever effect is noticed on dependent variable as a result of extraneous variable(s) is technically described as an 'experimental error'. A study must always be so designed that the effect upon the dependent variable is attributed entirely to the independent variable(s), and not to some extraneous variable or variables.

Control

One important characteristic of a good research design is to minimise the influence or effect of extraneous variable(s). The technical term 'control' is used when we design the study minimising the effects of extraneous independent variables. In experimental, the term 'control' is used to refer to restrain experimental conditions.

Confounded Relationship

When the dependent variable is not free from the influence of extraneous variable(s), the relationship between the dependent and independent variables is said to be confounded by an extraneous variable(s).

Research Hypothesis

When a prediction or a hypothesized relationship is to be tested by scientific methods, it is termed as research hypothesis. The hypothesis is a predictive statement that relates an independent variable to a dependent variable. Usually, a research hypothesis must contain at least one independent and one dependent variable. Predictive statements which

are not to be objectively verified or the relationship that are assumed but not to be tested, are not termed research hypothesis.

Experimental and Non-Experimental Hypothesis –Testing Research

When the purpose of research is to test a research hypothesis, it is termed as hypothesis –testing research. It can be of the experimental design or of the on-experimental design. Research in which the independent variable is manipulated is termed 'experimental hypothesis-testing research' and a research in which an independent variable is not manipulated is called 'non-experimental hypothesis-testing research'. For instance, suppose a researcher wants to study whether intelligence affects reading ability for a group of students and for this purpose he randomly selects 50 students and tests their intelligence and reading ability by calculating the efficient of correction between the two sets of scores. This is an example of non-experimental hypothesis-testing research because herein the independent variable ,intelligence ,is not manipulated. But now suppose that our researcher randomly selects 50 students from a group of students who are to take a course in statistics and then divides them into two groups by randomly assigning 25 to Group A ,the usual studies programme, and 25 to Group B , the special studies programme. At the end of the course , he administers a test to each group in order to judge the effectiveness of the training programme on the student's performance level. This is an example of experimental hypothesis-testing research because in this case the independent variable, viz., the type of training programme, is manipulated.

Experimental and Control Groups

In an experimental hypothesis-testing research when a group is exposed to usual conditions, it is termed a 'control group', but when the group is exposed to some or special condition ,it is termed as 'experimental group'. In the above illustration, the Group A can be called a control group and the Group B an experimental group. If both groups A and B are exposed to special studies programme ,then both groups would be termed 'experimental groups'. It is possible to design studies which include only experimental groups or studies.

Treatments

The different conditions under which experimental and control groups are put are usually referred to as 'treatments'. In the illustration taken above ,the two treatments are the usual studies programme and the special studies programme. Similarly ,if we want to determine through an experiment the comparative impact of three varieties of fertilizers on the yield of wheat, in that case the three varieties of fertilizers will be treated as three treatments.

Experiment

The process examining the truth of a statistical hypothesis ,relating to some research problem ,is known as an experiment . For example , we can conduct an experiment to examine the usefulness of a certain newly developed drug. Experiments can be of two types viz., absolute experiment and comparative experiment. If we want to determine the impact of a fertilizer on the yield of a crop, it is a case of absolute experiment ;but if we want to determine the impact of one fertilizer as compared to the impact of some other fertilizer ,our

experiment then will be termed as a comparative experiments. Often, we undertake comparative experiments which we call *comparative designs of experiments*.

Experimental Units

The pre-determined plots or the blocks ,where different treatments are used, are known as experimental units. Such experimental units must be selected or defined very carefully.

Types of Research Designs

Research Designs will vary according to nature of research ,type of research and objectives of research. Different research designs can be categorized as below:

We take up each category separately.

Research Designs in Case of Exploratory Research Studies

Exploratory research studies are also termed as formulative research studies. The main purpose such studies is that of formulating a problem for more precise investigation or of developing the working hypothesis from an operational point of view. The major emphasis in such studies is on the discovery of ideas and insights. As such the research design appropriate for such studies must be flexible enough to provide opportunity for considering different aspects of a problem under study . In built flexibility in research design is needed because the research problem ,broadly defined initially is transformed into one with more precise meaning in exploratory studies, which infact may necessitate changes in the research procedure for gathering relevant data. Generally ,the following three methods I the context or research design for such studies are talked about: (a) the survey of concerning literature ;(b) the experience survey and(c) the analysis of 'insight -stimulating' examples.

The survey of concerning literature happens to be the most simple and fruitful method of formulating precisely the research problem or developing hypothesis. Hypothesis stated by earlier workers may be reviewed and their usefulness be evaluated as a basis for further research. It may also be considered whether the already stated hypothesis. In this way the researcher should review and build upon the work already done by others ,but in cases where hypothesis have not yet been formulated , his task is to review the available material for deriving the relevant hypothesis from it.

Besides , the bibliographical survey of studies ,already made in one's area of interest may as well be made by the researcher for precisely formulating the problem. He should also make an attempt to apply concepts and develop theories in different research contexts to the area in which he is himself working. Sometimes the works of creative writers also provide a fertile ground for hypothesis-formulation and as such may be looked into by the researcher.

Experience survey means the survey of people who have had practical experience with the problem to be studied. The object of such a survey is to obtain insight in to the relationship between variables and new ideas relating to the research problem. For such a survey people who are competent and can contribute new ideas may be carefully selected as respondents

to ensure a representation of different types of experience. The respondents so selected may then be interviewed by the investigator. The researcher must prepare an interview schedule for the systematic questioning of informants. But the interview must ensure flexibility in the sense that the respondents should be allowed to raise issues and questions which the investigator has not previously considered. Generally, the experience-collecting interview is likely to be long and may last for few hours. Hence, it is often considered desirable to send a copy of the questions to be discussed to the respondents for doing some advance thinking over the various issues involved so that, at the time of interview, they may be able to contribute effectively. Thus, an experience survey may enable the researcher to define the problem more concisely and help in the formulation of the research hypothesis. This survey may as well provide information about the practical possibilities for doing different types of research.

Analysis of 'insight-stimulating' examples is also a fruitful method for suggesting hypothesis for research. It is particularly suitable in areas where there is little experience to serve as a guide. This method consists of the intensive study of selected instances of the phenomenon in which one is interested. For this purpose the existing records, if any, may be examined, the unstructured interviewing may take place, or some other approach may be adopted. Attitude of the investigator, the intensity of the study and the ability of the researcher to draw together diverse information into a unified interpretation are the main features which make this method an appropriate procedure for evoking insights.

Research Design in Case of descriptive and Diagnostic Research Studies

Descriptive research studies are those studies which are concerned with describing the characteristics of particular individual, or of a group, whereas research studies determine the frequency with which something occurs or its association with something else. The studies concerning whether certain variables are associated are examples of diagnostic research studies. As against this, studies concerned with specific predictions, with narration of facts and characteristic concerning individual, group or situation are all examples of descriptive research studies. Most of the social research comes under this category. From the point of view of the research design, the descriptive as well as diagnostic studies share common requirements and as such we may group together these two types of research studies. In descriptive as well as in diagnostic studies, the researcher must be able to define clearly, what he wants to measure and must find adequate methods for measuring it along with a clear cut definition of 'population' he wants to study. Since the aim is to obtain complete and accurate information in the said studies, the procedure to be used must be carefully planned. The research design must make enough provision for protection against bias and must maximise reliability with due concern for the economical completion of the research study. The design in such studies must be rigid and not flexible and must focus attention on the following:

- Formulating the objective of the study (what the study is about and why is it being made?)
- Designing the methods of data collection (what techniques of gathering data will be adopted?)

- Selecting the sample (how much material will be needed?)
- Collection the data (where can the required data be found)
- Processing and analyzing the data.
- Reporting the finding.

In a descriptive / diagnostic study the first step is to specify the objectives with sufficient precision to ensure that the data collected is relevant. If this is not done carefully, the study may not provide the desired information.

Research Design in Case of Hypothesis-Testing Research Studies

Hypothesis-testing research studies (generally known as experimental studies) are those where the researcher tests the hypothesis of causal relationship between variables. Such studies require procedure that will not only reduce bias and increase reliability ,but will permit drawing inferences about causality . Usually experiments meet this requirement. Hence, when we talk of research design in such studies, we often mean the design of experiments.

Professor R.A .Fisher's name is associated with experimental designs. Beginning of such design was made by him when he was working at Rothamsted Experimental Station(Centre for Agricultural Research in England). As such the study of experimental design has its origin in agricultural research. Professor Fisher found that by dividing agricultural fields or plots into different blocks and then by conducting experiments in each of these blocks, whatever information is collected and inferences drawn from them, happens to be more reliable. This fact inspired him to develop certain experimental design for testing hypothesis concerning scientific investigations. Today, the experimental designs are being used in researches relating to phenomena of several disciplines . Since experimental designs originated in the context of st operations , we still use, thought in a technical sense, several terms of agriculture (such as treatment, yield ,plot ,block etc.) in experimental designs.

Basic principles of experimental designs

Professor Fisher has enumerated three principles of experimental designs: The principles of Replication; The principle of Randomization ; and The Principle Local Control.

According to the principle of Replication , the experiment should be repeated more than once Thus, each treatment is applied in many experimental units instead of one. By doing so the statistical accuracy of the experiments is increased. For example, suppose we are to examine the effect of two varieties of rice. For this purpose we may divide the field into two parts and grow one variety in one part and the other variety in the other part. We can then compare the yield of the two parts and draw conclusion on that basis. But if we are to apply the principle of replication to this experiment, then we first divide the field into several parts, grow one variety in half of these parts and draw conclusion by comparing the same. The result so obtained will be more reliable in comparison to the conclusion by comparing the same. The result so obtained will be more reliable in comparison to the conclusion we draw without applying the principle of replication. The entire experiment can even be repeated several times for better results. Conceptually replication does not present any difficulty, but computationally it does. For example , if an experiment requiring a two-way

analysis of variance since replication itself may be a source of variation in the data . However ,it should be remembered that replication is introduced in order to increase the precision of a study ; that is to say , to increase the accuracy with which the main effects and interaction can be estimated.

The Principle of Randomization provides protection when we conduct an experiment against the effect of extraneous factors by randomization. In other words , this principle indicates that we should design or plan the experiment in such a way that the variations caused by extraneous factors can all be combined under the general heading of "chance". For instance, if we grow one variety of rice, say, in the first half of the parts of a field and the other variety is grown in the other half , then it is just possible that the soil fertility may be different in the first half in comparison to the other half. If this is so ,our results would not be realistic. In such a situation , we may assign the variety of rice to be grown in different parts of the field on the basis of some random sampling technique.i.e., we may apply randomization principles and protect ourselves against the effects of the extraneous factors(soil fertility difference in the given case). As such ,through the application of the application of the principle of randomization , we can have a better estimate of the experimental error.

The Principle of Local Control is another important principle of experimental designs. Under it the extraneous factor, the known source of variability , is made to vary deliberately over as wide range as necessary and this needs to be done in such a way that the variability it causes can be measured and hence eliminated from the experimental error. This means that we should plan the experiment in a manner that we can perform a two-way analysis of variance, in which the total variability of the data is divided into three components attributed to treatments(varieties of rice in our case), the extraneous factor (soil fertility in our case) and experimental error. In other words , according to the principle of local control, we first divide the field into several homogeneous parts, known as blocks, and then each such block is divided into parts equal to the number of treatments. Then the treatments are randomly assigned to these parts of a block. Dividing the field into several homogeneous parts is known as 'blocking'. In general , blocks are the levels at which we hold an extraneous factor fixed ,so that we can measure its contribution to the total variability of the data by means of a two-way analysis of variance. In brief ,through the principle of local control we can eliminate the variability due to extraneous factor(s) from the experimental error.

Important experimental designs

Experimental design refers to the framework or structure of an experimental and as such there are several experimental designs. We can classify experimental design into two broad categories, viz., informal experimental designs. Informal experimental designs are those design that normally use a less sophisticated form of analysis based on difference in magnitudes, whereas formal experimental designs offer relatively more control and use precise statistical procedure for analysis .Important experimental designs are as follows:

Informal experimental designs:

- After Only Design
- Before-and-after without control design.
- After-only with control design.
- Before-and-after with control design.
- Ex Post Facto design

Formal experimental designs:

- Completely randomized design(C.R.Design).
- Randomized block design (R.B.Design)
- Latin square design(L.S.Design).
- Factorial designs

We may briefly deal with each of the above stated informal as well as formal experimental designs.

After Only Design

This is the simplest of all experimental designs. This design consists of applying the experimental variable to an experimental group and measuring the dependent variable after and only after the application of the experimental variable. For instance, a bulb manufacturer wants to know the market response to his product. For this, he may have given advertisements in news papers. Now, he wants to measure the impact of the advertisement on sales. He can compare the sales of the product after advertisement and before and may conclude about the impact of advertisement on his sale. But this is a crude method of finding the impact of advertisement, because sales of the product might have gone up due to some other factors also.

Before-and-After Without Control Design

In such a design a single test group or area is selected and the dependent variable is measured before the introduction of the treatment. The treatment is then introduced and the dependent variable is measured again after the treatment has been introduced. The effect of the treatment would be equal to the level of the phenomenon after the treatment minus the level of the phenomenon before the treatment. The design can be represented thus:

After-Only with Control Design

In this design two groups or areas(test area and control area) are selected and the treatment is introduced into the test area only. The dependent variable is then measured in both the areas at the same time. Treatment impact is assessed by subtracting the value of the dependent variable in the control area from its value in the test area. This can be exhibited in the following form:

The basic assumption in such a design is that the two areas are identical with respect to their behavior towards the phenomenon considered. If this assumption is not true, there is the possibility of extraneous variation entering into the treatment effect. However, data can be collected in such a design without the introduction of problems with the passage of time. In this respect the design is superior to before-and-after without control design.

Before-and-After with Control Design

In this design two areas are selected and the dependent variable is measured in both the areas for an identical time-period before the treatment. The treatment is then introduced into the test area only, and the dependent variable is measured in both for an identical time-period after the introduction of the treatment. The treatment effect is determined by subtracting the change in the dependent variable in the control area from the change in the dependent variable in test area. This design can be shown in this way:

This design is superior to the above two designs for the simple reason that it avoids extraneous variation resulting both from the passage of time and from non-comparability of the test and control areas. But at times, due to lack of historical data, time or a comparable control area, we should prefer to select one of the first two informal design stated above.

Ex Post facto design

In this design, the researcher is required to select two or more groups for this study. Out of these two groups, in one group, the event has already taken place and the other is free from the happening of the event in social research, there are many problems which can be studied only with the help of adequate historical background. For instance, we are interested in finding out why revolutions take place in a particular country. In this case, the researcher has to select a country where a revolution has taken place and the other where there is no revolution. The countries otherwise should be broadly similar. Through a comparative study of the conditions of the two countries, the researcher may be able to find out the cause of the revolution.

Completely Randomized Design (C.R. Design)

Involves only two principles viz., the principle of replication and the principle of randomization of experimental designs. It is the simplest possible design and its procedure of analysis is also easier. The essential characteristic of the design is that subjects are randomly assigned to experimental treatments (or vice-versa). For instance, if we have 10 subjects and if we wish to test 5 under treatment A and 5 under treatment B, the randomization process gives every possible group of 5 subjects selected from a set of 10 an equal opportunity of being assigned to treatment A and treatment B. One-way analysis of variance (or one-way ANOVA) is used to analyse such a design. Even unequal replications can also work in this design. It provides maximum number of degrees of freedom to the error. Such a design is generally used when experimental areas happen to be homogeneous. Technically, when all the variations due to uncontrolled extraneous factors are included under the heading of chance variation, we refer to the design of experiment as C.R. design.

Randomized Block Design (R.B. Design)

Randomized Block Design is an improvement over the C.R. design. In the R.B. design the principle of local control can be applied along with the other two principles of experimental designs. In the R.B. design, subjects are first divided into groups, known as blocks, such that within each group the subjects are relatively homogeneous in respect to some selected variable. The variable selected for grouping the subjects is one that is believed to be related to the measures to be equal to the number of treatments and one subject in each block would be randomly assigned to each treatment. In general, blocks are the levels at which we hold the extraneous factor fixed, so that its contribution to the total

variability of data can be measured. The main feature of the same number of the R.B. design is that in this treatment appears the same number of times in each block. The R.B. design is analysed by the two-way analysis of variance technique , discussed latter.

Random replications design:

The limitation of the two -group randomized design is usually eliminated within the random replications design . In the illustration just cited above ,the teacher differences on the dependent variable were ignored, I.e. the extraneous variable was not controlled . But in a random replications design, the effect of such differences are minimised for (or reduced) by providing a number of repetitions for each treatment. Each repetition is technically called a 'replication'. Random replication design serves two purposes viz., it provides controls for the differential effects of the extraneous independent variables and secondly, it randomizes any individual differences among those conducting the treatments. Diagrammatically we can illustrate the random replications design thus:

Latin Square Design(L.S. Design)

This is an experimental design very frequently used in agricultural research. The conditions under which agricultural investigations are carried out are different from those in other studies for nature plays an important role in agriculture. For instance, an experiment has to be made through which the effects of five different varieties of fertilizers on the yield of a certain crop, say wheat ,is to be judged. In such a case the varying fertility of the soil in different blocks in which the experiment has to be performed must be taken into consideration; is the results obtained may not be very dependable because the output happens to be the effect of not only fertilizers, but it may also be the effect of fertility of soil. Similarly , there may be impact of varying seeds on the yield. To overcome such difficulties, the L.S. design is used when there are two major extraneous factors such as the varying soil fertility and varying seeds.

The Latin-square designs is one wherein each fertilizer , in our example ,appears five times but is used only once in each row and in each column of the design . In other words, the treatments in a L.S. design are so allocated among the plots that no treatment occurs more than once in any one row or any one column. The two blocking factors may be represented through rows and columns (one through rows and the other through columns). The following is a diagrammatic from of such a design in respects of ,say, five types of fertilizers, viz., A,B,C,D and E and the two blocking factors viz., the varying soil fertility and the varying seeds:

The above diagram clearly shows that in a L.S. design the field is divided into as many blocks as there are varieties of fertilizers and then each block is again divided into as many parts as there are varieties of fertilizers in such a way that each of the fertilizer variety is used in each of the block (whether column-wise or row-wise) only once. The analysis of the L.S . design is very similar to the two-way ANOVA technique.

The merit of this experimental design is that it enables differences in fertility gradients in the field to be eliminated in comparison to the effects of different varieties of fertilizers on the yield of the crop. But this design suffers from one limitation ,and it is that although each row and each column represents all fertilizer varieties equally , there may be

considerable differences in the row and column means both up and across the field. This, in other words, means that in L.S. design we must assume that there is no interaction between treatments and blocking factors. This defect can, however, be removed by taking the means of rows and column equal to the field mean by adjusting the results. Another limitation of the design is that it requires number of rows, columns and treatments to be equal. This reduces the utility of the design. In case of (2x2) L.S. design, there are no degrees of freedom available for the mean square error and hence the design cannot be used. If treatments are 10 or more, than each row and each column will be larger in size to that rows and columns may not be homogeneous. This may make the application of the principle of local control ineffective. Therefore, L.S. design of orders (5x5) to (9x9) are generally used.

Factorial Designs

Factorial designs are used in experiments where the effects of varying more than one factor are to be determined. They are specially important in several economic and social phenomena where usually a large number of factors affect a particular problem. Factorial Designs can be two types (i) simple factorial designs and (ii) complex factorial designs. We take them separately.

Simple factorial designs :

In case of simple factorial designs, we consider the effects of varying two factors on the dependent variable, but when an experiment is done with more than two factors, we use complex factorial designs. Simple factorial designs is also termed as a 'two-factorial design', whereas complex factorial design is known as 'multi-factor-factorial design'. Simple factorial design may either be a 2x2 simple factorial design, or it may be, say, 3x4 or 5x3 or the like type of simple factorial design.

Complex factorial designs

Experiments with more than two factors at a time involve the use of complex factorial designs. A design which considers three or more independent variables simultaneously is called complex factorial design. A complex factorial design is known as multi-factor factorial design. In the case of three factors with one experimental variable having two treatments and two control variables, each one of which having two levels, the design used will be termed 2 x 2 x 2 complex factorial design.

Research Designs for studies in commerce and management

Much of the research in commerce and management in India is exploratory in character. Researchers in these disciplines are now trying to have better insights into the problem areas and are in the process of testing vague hypothesis for clear conclusion. In this sense, we can say that the stage of research in these disciplines in India is primary. Because of the various problems of collecting data from primary sources, many researchers are depending on the secondary data. Mainly researchers in India use Government or statutory organisation's publications for basing their hypothesis. Surveying the individuals is also common. Studies in the areas like financial management, personal management, banking industry are generally based on the secondary type of data and hence are mainly exploratory in nature.

Descriptive and diagnostic studies occupy the next important place in research in commerce and management. Diagnostic studies are undertaken as a step to identify the factors contributing the phenomenon. Studies are undertaken to find out the reason why profitability of company or industry is falling? Similarly, studies are conducted to find out the reason for fluctuations in the prices of shares.

Classical experimental research design

Suppose there are two cases. Let in one of them observation X can be made while in the other it cannot be. Let the factor A occurs when observation X is made and does not occur when X is not made. Then it can be asserted that there is causal relationship between A and X. such relations are studied by way of constructing an experimental design defining controlled groups or series. This is classical experimental research design.

Review Questions

1. Explain research designs
2. State the purpose of preparing research designs
3. What are the contents of research designs
4. State the functions of research designs
5. What is the need for research designs
6. State the qualities of good research design
7. State factors to be considered in determining research design
8. What are dependent variables
9. What are independent variables
10. Explain extraneous variable
11. What is meant by control in research
12. Classify research designs
13. What are treatments in research
14. State the basic principles in research designs
15. What are formal and informal research designs
16. What is Before and After research design
17. What is Ex Post Facto research design
18. Explain Randomized Block research design

UNIT 11 POPULATION AND SAMPLE STUDY

Introduction

In any statistical investigation or research we are interested in studying the various characteristics of individuals or items of a particular group. Theoretically all the individuals or items of the group must be taken into consideration. the sum total of the items or individuals is called the population.

But in practice, the study is limited to a sub group called sample. The findings on the basis of sample study are applicable to the population. This group of individuals under study is known as the population or a universe.

Let us take very simple example to explain the concept of sampling . Suppose you want to estimate the average age of the students in your class. There are two ways of doing this. The first method is to contact all students in the class, find out their ages, add them up

and then divide this by the number of students. The second method is to select a few students from the class, ask them their ages add them up and then divide by the number of students you have asked. From this you can make an estimate of the average age of the class. In researches, population, sample, sampling and sample designs are critical factors.

Population

In statistics and researches, population is the aggregate of objects, animate or inanimate under study. It is the collection of all the items about which the information is desired. For example, if we want to measure the average cell phone bill of the people in a particular city in a particular month, our population would consist of the specified month's cell phone bills of all the people in that city. The population or universe can be finite or infinite, or existent or hypothetical.

Finite and infinite population

A population containing a finite number of objects or items is known as finite population. For example, the students in a college, the population of city etc. On the other hand, a population having an infinite number of objects or with the number of objects so large as to appear practically infinite is termed as infinite population. For example, the population of temperatures at various points of the atmosphere, the population of stars in the sky etc are infinite.

Existent and hypothetical population

Populations may be existent or hypothetical. A population consisting of concrete objects is known as existent population. Example, the population of books in a library. On the other hand, if the population consists of imaginary objects then it is called hypothetical population. For instance, the population of the throws of a die thrown infinite number of times is hypothetical population.

Census method and sample survey

Census method and Sample survey are two methods of collection of data. For certain processes, census method is employed whereas in most researches sample survey is made use of.

Census Method

Census is a method of collecting data in which information are collected from every individuals of the population. For example, if it is desired to study the physical stature of the students of a particular college, data may be obtained from all students. The total student strength of the college is referred to as universe and each student as a unit. So under census enquiry, a study is made of all the units of the universe. The population census conducted in India once in every ten years is an example of census enquiry.

Advantages of census method

Even though census method is very rarely applied, it has certain inherent advantages, as noted below.

1. Data are obtained from each and every unit of the population.
2. The results obtained are likely to be more accurate and reliable.
3. The possibility of bias and prejudice is minimum.
4. It is an appropriate method in the case of rare events.
5. It has no substitute for certain processes like national population survey.

Disadvantage of census method

Following are some of the reasons why census method is no so commonly used.

1. The effort, money and time required for carrying out complete enumeration, generally, is very much.
2. If the study of an item destroys that unit, then the census method is inappropriate.
3. If the population is infinite, census method fails to cover all items
4. Incomplete coverage, faulty observations and incorrect tabulation may give inaccurate information.

Sample survey (or sampling)

Sampling may be defined as the process of obtaining information about an entire population by examining only a part of it. In any investigation if data are collected only from a representative part of the universe we say that the data are collected by sampling. The representative part is called a sample. The study of the sample reveals the characteristic of the universe. This is based on the Law of statistical regularity. For example: if we want to know the average height of students of a particular college, it is sufficient if the required measurements are taken from a few students selected at random. The average height revealed by sample enquiry will not be materially different from the average height revealed by the census enquiry.

Aim of Sampling

The basic objective of sampling is to draw inferences about the population. It is a tool which helps to know the characteristics of the population. Although much of the developments in the theory of sampling has taken place only recent years, sampling techniques were in use even during very old days. To examine the quality of grain in a bag, only a handful of grain examined. A house wife examines only two or three boiling rice to know whether the rice in the pot is ready or not. Similarly in many other fields knowingly or unknowingly people make use of sampling techniques'. The process of sampling involves three elements (a) selecting the sample (b) collecting the information from the sample (c) making an inference about the population.

Merits of Sampling

1. Less time and effort : Since in sampling only a part of the population is studied, considerable time and labour can be saved. Time is saved not only in collecting data but in processing it.
2. Less cost : The total financial burden of sample survey is generally less than that of a complete census , since the coverage of items is less.
3. More reliable results : In this method we have the advantage of detailed and intensive study and so the results are generally more accurate and reliable. The results obtained from sampling are generally more reliable than that obtained from a census method because (a) It is always possible to determine the extent of sampling errors. (b) The errors due to inaccuracy of information, incompleteness of returns etc. are less in a sample survey. Effective precautions can be taken to avoid errors.
4. Administrative convenience : It is possible to avail of the service of expert and to impart through training to the enumerators since the area of study is small.

5. More detailed and accurate information : Since only a part of population is studied, we can collect detailed information from all those items. Sampling gives accurate result since it is conducted by trained and experienced investigators.
6. Destructive nature of certain enquires : In many cases sample enquiry is the practical means of obtaining the required information when the quality of a thing can be ascertained only by destroying the articles For example : In testing the life of bulbs. Here sampling is the only practical way of assessing the quality of the whole lot. Again only a few chalks out of a certain lot need be tested for ascertaining the breaking strength of all chalks.
7. Sampling of method is best suited at times : When the universe is infinite or hypothetical census method is not practicable. So in those enquiries sampling survey can be adopted.
8. Limitations of Sampling
 1. A sample survey must be carefully planned and executed. Otherwise , the results obtained may be inaccurate and misleading. So sampling procedure must be perfect.
 2. Sampling requires the service of experts. In the absence of qualified and experienced persons, the information obtained from sample cannot be relied upon. It is likely that bias and prejudices will creep into this method.
 3. It is not easy to select a sample which is perfectly representing the universe in many situations.
 4. If information is required for each and every unit, sampling technique cannot be used.

Importance of Sampling

The sampling method has acquired an important place in the field of research. The causes for the importance of this method are:

1. Only representative units are studied : When sampling technique is adopted , we need not waste our time and energy for studying the items. We are only studying representative units.
2. Large area can be covered : Since we are selecting only few items from the universe, we can extend our enquiry to large area.
3. Scrutiny of all results available is possible : Since in sampling we are only studying representative units, it is easy to scrutinize all items selected for study.
4. Intensive study is possible : Since we are concentrating on few selected items, intensive study of those items is possible.
5. Collecting of Information is easy: Since in a sampling , entire universe is not studied. It is easy to collect information from the desired unit.
6. Required results are attained : Since we are selecting representative items the results obtained from the sample are same as those of the population.

Need for sampling (Why sampling?)

1. It saves money.
2. It saves time.
3. It gives detailed and accurate measurements.

4. It is necessary when the population is infinite.
5. It is necessary when the study of an item causes its destruction.
6. It is necessary when the area of survey is wide.

Assumption in sampling (Basis of sampling)

The sample is a small part of the population. This small body is to represent the population. Then only the sample will have the same characteristics of the population. The selection of sample is based on certain assumptions. They are

1. Homogeneity in general :

Each unit in a population differs from another in the same population. But still we find an overall uniformity among all items together. Therefore samples are drawn on the assumption that the items in the population have common characters.

2. Representative nature :

If we select a large number of units from the sample, the sample selected represents the universe generally.

3. Sample results are true:

In statistical enquiries, it is not possible to expect perfect accuracy. Therefore the results of sample study is presumed to be sufficiently accurate even though they are not perfectly accurate. Sample results are true in general.

4. Sample values are close to the population values.

Sample values can be used to estimate probable population values. It is universally accepted, due to two popular sampling theories.

Factors which decide whether Census method or Sample survey is suitable (Criteria for choosing sampling method or census method)

1. Field of enquiry : If the area to be covered is large , then census method is impracticable. In such a case sample method can be adopted so that some selected area from the field of enquiry can be subjected to study.
2. Availability of time : The time for the completion of the enquiry is a criterion for deciding which method is to be adopted. Census method is time consuming while sample survey requires less time.
3. Availability of finance : Census method is very expensive, while sample survey is less costly. Therefore depending upon the money available for spending, one can choose one of the two methods.
4. The degree of accuracy desired : Since every item of the universe is examined in a census method, the results obtained are reliable and accurate. In a sample survey only representative part is studied. So the results obtained need not fully convey the information about the population. In cases where individual information are necessary census method is to be followed.
5. When census method fails : When the population is infinite or when a study of a unit of the population destroys that unit or when the population is hypothetical, census method is not practicable.

Distinction between census method and sample survey

1. In census method, we collect information from all the units of the population while in sample survey we collect information only from a representative part of the population.
2. Census method of data collection is impossible in certain situations (Eg. When the population is infinite or when the study of an item causes destruction of that item etc.). But a sample survey is possible in all situations.
3. Census method makes a study of all individuals items and therefore it has the merit of accuracy and adequacy. Error due to incomplete coverage is also avoided. In a sample survey only few individuals are studied and therefore the information obtained has only a representative nature. Hence those information may not be accurate and adequate in some situations.
4. Census method is meant to study the population while sample survey is a method for drawing conclusions about the population.
5. Census method is usually time consuming and costly while sampling survey is less time consuming.
6. In census survey, a detailed study of all items is not possible when there are many items. In sample survey detailed study of the items is possible as the number of items selected is not much.

Theoretical basis of sampling

There are two important principles on which the theory of sampling is based. (1) Principle of statistical regularity (2) Principle of inertia of large numbers.

Law of statistical regularity:

This principle is derived from the mathematical theory of probability. In the words of King, 'The law of statistical regularity' lays down that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of that group. In other words the principle states that if a sample is taken at random from a population , it is likely to possess almost the same characteristic as that of the population.

Law of inertia of large numbers:

This principle is corollary to the principle of statistical regularity. This law states that large numbers are relatively more stable than small numbers. The difference in the aggregate result is likely to be significant, when the size of the sample is large , because the variations balance each other.

This principle does not state that large numbers do not permit change. But the change would be more regular.

Eg. The production of wheat in India may vary slightly over a period of years. But the total world production of wheat remains relatively stable for decades. Thus large numbers have great inertia.

Concept of probability

The idea of probability arises when something is not sure of taking place. When there is uncertainty about the future course of events and about their prediction we express the possibility of the occurrence of the event in terms of the times the event takes place in a

long series of observations during the past. This proportion is the probability. Therefor probability of an event is the numerical expression of the possibility of the occurrence of that event. It is the proportion of times an event is expected to occur out of the total number of repetitions of the experiment. The concept of probability is very useful in sampling theory. For example: one may be interested in knowing the probability for the given sample to belong to the given population and so on.

Review Questions

1. What is meant by population
2. What is a sample
3. What are the objectives of sampling
4. What are finite and infinite population
5. What are existent population
6. Explain hypothetical population
7. What is census
8. Explain sample survey
9. Distinguish between sample survey and census method
10. State merits and demerits of census
11. What are the merits of sample survey
12. Explain demerits of sample survey
13. State the significance of sampling in research
14. State assumptions behind sampling
15. Explain factors in choosing sampling
16. State theoretic basis of sampling
17. what is the role of probability in sampling
18. Why probabilistic sampling is preferred

UNIT 12 SAMPLE DESIGN

A sample design is a definite plan for obtaining a sample from a given population. It refers to the procedure, adopted by a researcher for selecting items for a sample. That is, it gives an idea about the size of the sample, number of samples and method of sampling to be adopted. Accordingly, there are many sample designs. A researcher can choose the appropriate sample design.

Objectives of sample design

Sample design is determined before data are collected, with specific objectives.

1. To estimate the size and number of samples to be selected for the study purpose.
2. To achieve maximum precision in the estimates within a given sample size.
3. To avoid bias in the selection of samples.
4. To determine the appropriate technique of drawing required number of samples.

Step in sample design

Following are the steps to be followed in developing a sample design.

1. Determining the type of population

The first step in developing any sample design is to clearly define the universe. The universe can be finite or infinite, existent (real) or hypothetical etc.

A population containing a finite number of items is known as finite population. For example, the students in a college, the population of a city etc. On the other hand, a population having an infinite number of objects or with the number of objects so large as to appear practically infinite, is termed as an infinite population. For example, the population of temperatures at various points of the atmosphere, the population of stars in the sky etc. The population consisting of concrete objects is known as existent population. Eg. the population of books in a library. On the other hand, if the population consists of imaginary objects then it is called hypothetical population. For instance, the population of the throws of a die thrown infinite number of times is hypothetical population.

Therefore in developing a sample design, a researcher has to define the objects to be studied. That is, he has to define whether the population under study is finite or infinite, existent or hypothetical etc.

2. Deciding the Sampling units

Before we proceed to study or select samples, we shall have to decide the units of the sample. The sampling units may be (1) geographical units like a state, a region, a district, a city etc. (2) structural units like a building, house, flat etc., (3) social groups like a family, a school, club, church etc., (4) an individual.

A sampling unit should be defined clearly. The unit should be a standard one. The unit selected should be easily accessible. A researcher has to decide the sampling units before selecting a sample.

3. Selecting Sampling frame (Source list)

A sampling frame is the list containing all items in the population. A researcher has to prepare a sampling frame. It is from this sampling frame that the items are selected to form the sample items are selected to form the sample.

The sampling frame should be exhaustive. That is, all the items of the universe must be there in a sampling frame. The list should be up to date and valid. It should be reliable and complete. It should be suitable for the study. That is, it should be relevant and properly drawn up.

4. Determining the sampling size

For a logical and meaningful sample survey, the sample size should be adequate. So the researcher should fix sample size before conducting sample survey.

5. Determining the parameters.

While developing a sample design, a researcher has to consider the parameters of the population also. Parameters may be proportion, average, measure of variability etc. The sample design should be so developed that these parameters are properly estimated from the samples.

6. Fixing the finance for the survey

The cost of selection of items will decide the size of the sample and type of the sample. Therefore the amount available for the conduct of the survey should be fixed before fixing the survey.

7. Selecting a sampling technique

A researcher has to use a proper technique in selecting the items for the sample. There are several sampling techniques available and the researcher should choose that which causes smaller sampling errors for a given sample size and cost.

Sampling Techniques

Various methods available for selecting samples can be grouped under two major heads : (1) Random sampling (probability sampling) and (ii) Non random sampling (Non-probability sampling).

Probability sampling (Random sampling)

Probability sampling design is based on probability for the selection of each item. Probability sampling is also known as random sampling or chance sampling. In this method each item has its own chance for being selected. Probability sampling may be simple random sampling and complex random sampling . Stratified sampling, systematic sampling and cluster sampling are known as complex random sampling designs.

In probability sampling we can measure the errors of estimation. Also it ensures the law of statistical regularity. This is why probability sampling is considered as the best technique of selecting representative sample.

Non-Probability sampling

Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability for each item to be included in the sample. Deliberate sampling, judgement sampling etc., are non-probability sampling. In non probability sampling generally the person selecting the sample purposively selects a sample which is suitable for his study.

Random sampling when the population is finite

In a finite population, the number of units in the population is finite (that is exhaustively countable). So it is possible to assign probability for each item's selection. Each item has equal probability.

Random sampling may be done 'with replacement' or 'without re-placement'. In a random sample with replacement the units are selected in such a way that the units once selected is returned before the selection of next unit. In random sampling ,it cannot appear in the sample again. Usually sampling without replacement is adopted in random sample selection.

If there are 'N' item in a population and 'n' items are to be selected in the sample, then it is possible to select samples in ${}^N C_n$ ways. To be more clear , say we have a population with 8 items, and we want to select certain samples from this population to have only 4 items each. Then there are ${}^8 C_4$ possibilities of selecting samples. = 70. That is 70 choices are there for selecting samples. The probability for each item to be selected is 1/70. Therefore if we select a sample in such a way that each sample has the probability of 1/70 of being selected then the sample selected is called random sample.

Random sampling from an infinite universe

It is difficult to explain the concept of random sample when the population is infinite. For example, when a die is thrown indefinite times, the population consists of

results of all the possible throws of the die. This is infinite. In this case the sample is the result of a particular number, of independent throws of the die,

Simple random sampling

A simple random sample is a sample selected from a population in such a way that every member of the population has an equal chance of being selected and the selection of any individual does not influence the selection of any other. The selection purely depends on chance. So the personal bias of the investigator will not be present in the sample selected. Random samples may be selected (a) by Lottery method (b) from table of random numbers.

(a) Lottery method :

Under this method all items of the universe are numbered or named on separate slips of paper of identical size and shape. These slips are then folded and mixed up in a container. A blind fold selection is then made of the number of slips required to constitute the desired size of sample. While preparing slips it must be seen that slips are of identical size, shade, colour etc. Otherwise there is possibility of selecting a particular slip. The selection of items can be made by some mechanical process also.

(b) Table of Random numbers:

Using random number table we can select random samples in an easy manner. Several standard tables of random numbers are available.(1) Tippett's random number tables(2) Fisher and Yates available(3) Kendall and BB Smith tables(4) C.R Rao's tables etc. Tippett's table of random numbers contains series of random digits arranged in rows and columns.

Random numbers avoid any bias in selecting the sample. The random numbers are so prepared that they fulfill the criterion of complete randomness. The use of the tables of random numbers involves the following steps(1) each number of the population is assigned a number .(2) From some random point of the table of random numbers, the random numbers are read out. The item whose number is identical with the random number, read out ,is selected.(3) The selection of the items is continued till the desired number of items are selected.

Since the selection of items in a simple random sample depends entirely on chance, there is no possibility of personal bias. If size of the sample is increased it will represent the population in a better manner.

Assumption in a simple random sample

Following assumption is made in the selection of a simple random sample.

1. All the items in population have the same chance of being included in the sample.
2. The selection of a particular item in a draw has no influence on the probabilities of selection of any other draw.
3. All the items of sample are selected independently of one another
4. The selection of each item is in a unbiased manner.

Merits and demerits of simple random sampling

Simple random sampling has both advantages and disadvantages.

Merits

1. There is no possibility of personal bias.
2. The simple random sample usually represent population particularly when the sample size is large.
3. The accuracy of the estimates based on the simple random sampling can easily be assessed since it is possible to measure the errors of estimation in random sampling.
4. Random sampling ensures the law of statistical regularity which states that the sample will have the same characteristic of the population if the sample is selected at random.

Demerits

1. It is often difficult to have an up-to-date list of all the items of the population to be sampled.
2. If the size of the sample is small, the results may not be reliable.
3. If the item in the universe are not homogeneous selection of a sample by simple random sample method cannot be adopted.

Illustration

Explain how you will select a sample of 100 units from population consisting of 4212 units, using random numbers.

Ans: Here $n=4212$. Identify all the units of the population by a number from 1 to 4212.

Choose four figure random numbers table and read out the numbers continuously, vertically or horizontally, starting from any random point. If the number read out is less than or equal to 4212 (but a four digit number) then divide it by 4212 and take the remainder. The unit with the number equal to the remainder may be selected. For example : if a number selected is less than 4212(403) select it. If the number selected is more than 4212 say 5150 then divide the number 5150 by 4212 and the remainder is 938. Then the unit 938 is selected. The number once selected may be discarded if it appears again. Thus we select 100 units from the population, which constitute the sample.

Example

A set of random numbers (4 digit) drawn from Tippet's random number table is given below.

6446	1545	6641	2370	3170	1300	1089	8126
4233	9792	1396	0560	7979	2693	5246	9143
2952	6913	5356	2754	5911	3408	1112	1405
4167	77691	8816	7483	7203	2762	6107	9025
9524	5624	8776	3992	6111	3563	6008	7002

Complex Random Sampling Designs

Probability sampling under restricted sampling technique may result in complex random sampling designs. These are also known as mixed sampling designs since both probability and non probability procedures are applied in this method. Some of the popular complex random sampling designs are (i) stratified sampling (ii) systematic sampling (iii) cluster sampling (iv) multistage sampling (v) sequential sampling etc.

Stratified sampling

If the field of enquiry is not homogeneous and contains variety of items, this type of sampling is adopted. In this method we first divide the population into different sub population known as strata, such that items in each stratum are homogeneous. Various strata, such that items in each stratum are homogeneous in character and strata between themselves are heterogeneous. From each stratum, items are selected by simple random sample method. For example: if we want to study the consumption pattern of the people in Kerala, Kerala may be divided into a number of zones. From each zone a sample may be taken at random. This method has the advantage of a simple random sample method and also the different groups in the population get representation. Stratified sampling method reduces time and expense to a great extent. There is little possibility of any essential group of the population being completely excluded. As the item in each stratum are homogeneous, stratified sampling greater accuracy. Stratified sampling is a mixed sampling since it consists in itself both purposive sampling and random sampling. Strata are purposively formed and are usually based on past experience and personal judgment of the researcher.

Kinds of Stratified sampling

1. Proportional stratified Sampling (where number of items taken from each stratum is on the basis of size of each stratum)
2. Disproportionate stratified sampling(where equal number of items are selected from each stratum, irrespective of its size).
3. Stratified weighted Sampling (where equal number of items are selected from each stratum, but weights are given to stratum on the basis of its size.)

Allotting sampling size of different strata

Generally number of units selected from each stratum is proportional to the number of units of that stratum in the population. For example, A population has 10000 units and a sample of 1000 units is to be selected. Then let the population be divided into 4 strata which has 1000,2000,3000 and 4000 units of the universe. Then the number of units to be selected for the sample will be as shown below:

$$\text{From I Stratum} : \frac{1000}{10000} \times 1000 = 100 \text{ units}$$

$$\text{From II Stratum} : = \frac{2000}{10000} \times 1000 = 200 \text{ units}$$

$$\text{From III Stratum} = \frac{3000}{10000} \times 1000 = 300 \text{ units}$$

$$\text{From IV Stratum} : \frac{4000}{10000} \times 1000 = 400 \text{ units}$$

Merits and demerits of stratified sampling

Merits:

1. It is more representative than simple random sample. There is very little possibility of any essential group of the population being completely excluded.
2. It ensures accuracy since each stratum consists of homogeneous items.

3. Stratified samples can be more concentrated geographically. Units from the different strata may be selected in such a way that all of them are located in one geographical area.
4. This will reduce the time and cost.

Demerits

If proper stratification is not done, the sample will have the effect of bias. If the units selected in a stratum are not homogeneous the results may not be reliable.

In the absence of random selection of items from each stratum which cannot be ensured in many cases, the results are not reliable.

Systematic sampling

This method is popularly used in those cases where complete list of the population from which sample is to be drawn is available. Under this method the items in the population are included in intervals of magnitude K. From every interval selected an item by random sample method. Therefore between two items selected, there is an interval 'K'. For example : when there are 1000 items in the population and we want to draw a sample of 200 items, then the interval = 5. From the first interval select an item by random sample method. Thereafter the items selected will be with a gap of 5. If first selected is 3rd item then the items in the sample will be 3rd, 8th etc.

Systematic sampling is relatively simple technique and may be more efficient than simple random sampling. The time and works involved in sampling by this method are relatively smaller. This method will not give a representative sample if some hidden periodicities are existing in the universe.

For example : If every 50th item produced by a certain production process is defective then the systematic sample will contain either all defective items or not defective items.

Systematic sampling is an easier and less costlier method of sampling and can be conveniently used even in case of large populations.

Cluster sampling

Cluster sampling is a random sampling procedure in which the elements for the sample are chosen from the population in groups or clusters rather than singly. Cluster sampling consists in forming suitable clusters of units. All the units in the sample of clusters selected are surveyed. The clusters used are pre existing natural or administrative groups of the population such as schools, Colleges, factories etc.

Cluster sampling can be adopted when there would be no reliable list of the units of the population or when it would be very expensive to collect such a list.

Cluster sampling is used in many surveys as it is very convenient and less costly. Tribal surveys can be done by cluster sampling technique. Tribes are generally selected in particular areas. Simple random sampling will fail here. In such surveys few settlements are selected as clusters and all the units of the selected clusters are surveyed.

Multistage sampling

This is another random sampling procedure carried out in several stages. The population is regarded as made up of a number of first stage sampling units each of which is made of second stage units etc. At first, the first stage units are sampled by some suitable

method like simple random sampling. Then a sample of second stage units is selected from each of the selected first stage units by some suitable method. This procedure is extended. The method has the advantage of being flexible.

Usually multistage sampling is applied in big enquiries which cover large geographical area. The advantages of this method are (1) easier to administer (2) large number of units can be sampled for a given cost.

Multistage sampling is a further development of the principle of cluster sampling. For example, if we want to study the working of cooperative institutions in the country, we can select certain states for study. This is first stage. Then we can select and study the functioning of the institution there. This represents a two stage sampling design with ultimate sampling units being clusters of districts. Suppose we select certain town from the select districts and study all institutions there, it is third stage sampling.

Multiphase sampling

When we want to organise a survey to study the total expenditure incurred by families in a town, we may have only a list of households in the town. So we have to collect data on consumer expenditure from a sample of families selected. If the variation among the households is very large we may have to select a large sample. The cost for such enquiry may therefore be very large.

Therefore to be more economic, another alternative procedure can be adopted. For this, as a first phase collect data on some simple characteristics related to consumer expenditure from a sample first selected. Use this information for arrangement, stratification and selection of the second phase sample of the households for the collection of the data on consumer expenditure. This is two phase sampling.

In multiphase sampling, the first step is to select samples as first phase and from it to collect data on some suitable characteristics (like household size). If the main survey is selected in two or more phases then it is described as multiphase sampling.

Sequential Sampling

This type of sampling is adopted in acceptance sampling plans in statistical Quality Control. When a particular lot is to be accepted or rejected on the basis of a single sample, it is known as single sampling. When the decision is to be taken on the basis of two samples, it is known as double sampling. When it is based on more than two samples it is multiple sampling. Thus in this method samples are taken one after another as desired. The ultimate size of the sample under this technique is not fixed in advance. This type of sampling design is called sequential sampling.

Purposive Sampling (Judgment sampling) (deliberate sampling)

Under purposive sampling the investigator exercises his discretion in the matter of selecting the items that are to be included in the sample. He deliberately picks up those items which he thinks are the representative of the population. Therefore, in this method items are selected on the judgment of the researcher. For instance, to study the economic conditions of people in a country, few towns or villages can be deliberately selected which the researcher feels that will represent the country.

Under purposive sampling, therefore the selection of items to be included in the sample entirely depends upon the direction of the investigator and no formula or principles

is followed. If the investigator is wise and unbiased, the sample selected in this manner will be representative. But the chances of personal prejudice and bias are always there. This sampling design is rarely used in large enquiries. This method is not at all scientific. If the researcher is impartial, unbiased and experienced, the results obtained from a purposive sampling will be tolerable.

The basic difference between purposive and random sampling, therefore, is that in the former the selection of items to be included in the sample, entirely depends upon the discretion of the investigator. In the latter, it is not so, since the chance alone is allowed to determine which items from the population are to be selected. So the personal bias of the investigator will not be present in the sample selected. Purposive sampling is subjective whereas random sampling is entirely objective.

Merits of purposive sampling

1. Representative character: It is possible to make the selection of items as representative as possible.
2. Specific objective : In a purposive sampling , there is specific purpose for which the study is made.
3. Free from bias : It is possible for the researcher to keep himself free from bias and prejudice.
4. Merits of Purposive sampling
5. If the researcher avoids bias and prejudices, even a small sample can be representative.
6. The researcher can select useful units. Therefore the sample will not include irrelevant items.

Demerits

1. It is not possible for the researcher to have complete idea about the universe. So he may not make a proper selection.
2. If the researcher is biased , the sample results can not be depended on

Convenience Sampling

A convenience sample is obtained by selecting a convenient population. For example, for the study of the spending habits of students, if we select a sample of students from a nearby college the sample is convenience sample.

Quota Sampling

This method of sampling is adopted in making surveys of public opinion. Only definite quotas of persons in different social classes, different age groups, different age groups, different regions etc. are selected for interviews. Convenience sampling methods are sometimes not representative since they are biased and unsatisfactory.

Under quota sampling the interviewers are simply given quotas to be filled from the different strata with some restriction they are to be filled. This type of sampling is convenient and less expensive.

Snowball sampling

Snowball sampling is the process of selecting a sample using networks. To start with, a few individuals in a group or organization are selected and the required information is collected from them. They are then asked to identify other people in the group or

organization, and the people selected by them become a part of the sample. Information is collected from them, and these people are asked to identify others members of the group and, in turn, those identified become the basis of further data collection. This process is continued until the required number or a saturation point has been reached, in terms of the information being sought.

Interpreting validity of sample result

A sample is selected to find out the characteristics of the universe. Therefore a sample selected from the universe should be reliable. The reliability of the sample may be tested on the basis of the following aspects:

1. Size of the sample : The size of the sample determines the representative character. So it should be adequate and appropriate. The accuracy of the results derived through the sample can be judged on the basis of the size of the sample selected.
2. Representativeness of the sample : The sample selected for study should represent, by and large, the population in all respects. The results obtained from a sample which is not a true representative, cannot be used to generalise the characteristics of the population.
3. Homogeneity of the samples : Samples selected for the study of population must be homogeneous. Otherwise the sample cannot be relied upon.
4. Parallel Samples : To examine the reliability of a sample, another sample from the same universe can be drawn and a comparison between the two samples can be made.
5. Unbiasedness in the selection of a sample : Sample is to be selected in an unbiased manner. The researcher has to take certain precautions before selecting a sample so that the sample selection is free from bias and prejudice.
6. Sampling within sample : To test the reliability of a sample drawn from the universe, we draw a sub sample from the main sample. An intensive study of the sub sample can be made. The findings of this study can be compared with the findings of the study of the main sample. This helps the researcher to detect the errors in conclusions.
7. Comparison of measurements of sample with known measures of population: If any measure of the population (for example, average) is known then it can be compared with the similar measure obtained from the sample. If there is difference in two measures the reliability of the sample is to be doubted.

Sampling and Non sampling errors

Sampling is just a short cut method , but scientific , for studying a phenomenon. It has several inherent mistakes or errors, which must be considered by a researcher. Such sampling errors involved in the collection of data are classified into sampling and non sampling errors.

Sampling errors

The errors arising due to drawing inference about the population on the basis of sample is termed sampling error. Even if most care has been taken in selecting a sample the

results derived from a sample study may not be exactly equal to the true value in the population. Hence sampling gives rise to certain errors known as sampling errors. Therefore sampling error is nonexistent in complete enumeration survey. Sampling errors may be biased or unbiased. Biased errors are due to chance difference between values included and not included in the sample.

The sampling errors occur randomly and they are compensatory type. Therefore their expected value is zero. When the sample size increase, sampling errors decrease. If the population is homogeneous the sampling error becomes smaller. The measure of sampling error is called precision of the sampling plan. Therefore by increasing the population size, precision of the sampling plan can be improved. But increasing the size of the sample results into increase of cost. Therefore effective way of increasing precision is to select a better sampling design which has a smaller sampling error.

Non Sampling errors

The errors arising mainly at the stages of ascertaining and processing of data are termed non sampling errors. They occur both in complete enumeration and sample survey. Non sampling error can occur at every stage of planning and execution of the census survey. Non sampling errors may arise from one more of the following factors.

1. Data collected are inadequate and not consistent with the objective of the surveys.
2. Inappropriate statistical unit.
3. Lack of trained and experienced investigators.
4. Errors in presentation and summarization of data.

Systematic bias

A systematic bias results from the errors in the sampling procedures. It can be reduced by increasing the sample size. The causes for this type of bias (or errors) can be detected and corrected. The systematic bias is caused by the following factors.

i) Inappropriate sampling frame:

If the sampling frame does not represent the universe properly, it will result in systematic error.

ii) Defective measuring device :

If the devices used for measuring values are defective, they will cause systematic error.

iii) indeterminacy

If individuals are kept under observation , they act in a manner which is different from if they are not kept under observation. Therefore indeterminacy principle may also cause systematic bias.

Factors which decide (or criteria for) selection of a sampling technique

In sampling analysis , two kinds of costs are involved, one the cost of collecting the data and the other - costs of an incorrect inference resulting from the data. The cause of incorrect inference are systematic bias and sampling error. Systematic bias results from errors in sampling procurers. Sampling errors are the random variation in the sample estimates derived for the population parameters.

While selecting a sampling technique, a researcher must ensure that:

- The technique adopted causes a small sampling error such as inference error.
- It controls the systematic bias in a better way, and
- Reduces the cost involved in sampling process.

Biased and Unbiased errors

Error is the difference between the observed value and the true value. Errors may be classified as biased errors and unbiased errors. Bias errors are also called Systematic errors. They arise due to bias of the informants or enumerators or defective measurements. Biased errors are either positive or negative. They are cumulative in nature. The unbiased errors occur in the regular course. They are positive as well as negative. They are compensating also.

Measurement of Errors

Measurement of errors may be absolute or relative.

Absolute Error:

Difference between actual value and the estimated value of an item is called absolute error.

Relative Error :

Relative error is the ratio of the absolute error to the estimated value. Relative error when multiplied by 100 is called percentage error.

Review Questions

1. What is meant by population
2. What is sampling
3. What are the essential qualities of good sample?
4. Distinguish between census and sampling methods of collecting data.
5. Why sampling necessary in many statistical enquiries?
6. What are the merits and demerits of sample survey?
7. Distinguish between Population and Sample
8. What is Simple Random Sampling
9. Explain Stratified Sampling
10. What is and Complex Random Sampling.
11. What is multiphase sampling
12. Define simple random sampling and explain a method of drawing such a sample from population.
13. What are random sampling numbers? How do you use them for the selection of a sample from a given population?
14. Explain the law of statistical regularity and law of inertia of large numbers?
15. What is systematic sampling?
16. What is meant by cluster sampling?
17. Explain sequential sampling?
18. Explain the criteria for the selection of a sampling procedure.
19. What is a sample design?

20. What is a Purposive Sampling?
21. Distinguish between Random sampling and Deliberate sampling.
22. What is sampling error?
23. What is non sampling error?

UNIT 13 DETERMINATION OF SAMPLE SIZE

One of the critical factors influencing research and statistical investigation is sampling and sample size. Sampling should be sufficiently reliable to estimate and examine certain values of the whole population. A housewife takes a few boiling rice, from a boiling pot, to check its cooking. She is ensuring the cooking of the whole pot, ie, the whole population on the basis of a few sample grains. Thus , emerged the concept of sampling.

Sampling is a tool which helps to know the characteristics of the population. Sampling is defined as the process of drawing representative number of items for collecting information to infer about the population. This unit focuses on principles governing sample, sample size, factors influencing sample size and approaches to determination of sample size.

Principles of sampling

Sampling is reliable, because it is based on two universally accepted principles or theories.

Law of statistical regularity

The principle states that if samples are drawn at random from a population, it is likely to possess the characteristics of the population. In other words samples will be statistically regular, if samples are regular.

Law of inertia of large numbers

According to this principle, large numbers are relatively more stable than small numbers. It is difficult to move or change large numbers. Large numbers have consistency.

Sample

A finite subset of a population, selected from it with the objective of investigating its properties is called a sample of that population. A sample is representative part of the population. It ideally purports to be a minute model or replica of the population. For example, when we want to study the life of electric bulbs produced by a company we select some bulbs and study their length of life. This selected number is the sample.

Characteristic of a good sample (Essential qualities of a good sample)

A good sample is that which fulfils the objectives of a research. A sample , in order to be a good sample, must have the following characteristics.

1. Representative : A good sample should be truly representing the universe. A good sample is one which possess by and large , all characteristics of the population. This can be achieved only if the sample is drawn by adopting the proper principle of selecting samples. If the units in the population are not homogeneous, then the sample should be so selected that it contains items of different section of the population so that sample is a true representative of the population.

2. Adequate size of the sample : Only if the sample possesses required number of units selected from the population, it will possess the representative character. A sample should not be very small. If it is very small it cannot represent the population. The size of the sample should be decided properly.
3. Free from bias and prejudice : The selection of the sample should be free from bias and prejudice. Otherwise samples may not be dependable.
4. Conformity to subject matter : The sample selected should fill the needs of the subject matter under study. Sample selected should serve the purpose. All types of sampling will not suit all kinds of enquiry. So the sample selected should be one which suits the situation.
5. Based on past and practical experience : The sample selected should be based on the past experiences and practical experiences of the investigators who have done similar research work.

Size of the sample

The number of sampling units selected from the population is called the size of the sample. If the size of the sample is small it may not represent the universe adequately. If the sample size is very large, it may require more time and money for investigation. Hence the sample size should not be too small or too large. It should be optimum. Optimum size ensures efficiency, representativeness, reliability and flexibility.

Factors influencing sample size.

Sample size is the number of items included in a sample. This is a decision factor in accurately estimating population parameters. Sampling precision depends more on sample size, and not on proportion of population sampled.

In sampling analysis, vital questions are - how large the sample should be? If the sample size is too small, estimation may be inaccurate. If it is too large, heavy cost may be incurred.

Following factors will decide the size of the sample

1. The size of the sample is determined by the nature of the study. If an intensive study is to be made sample of large size shall not be useful, a sample of a small size is more convenient.
2. Nature of the universe also determines the size of the sample. If the universe is homogenous a small size sample will yield required result. If the universe is not homogeneous a small sample will not be of much use.
3. Number of classes to be formed of the population is a criterion for determining the size of the sample. Sample size can be small, if the number of classes is not large in number.
3. Desired precision is another criterion for determining the size of the sample. Sample size is so determined that the results obtained from the sample achieves precision.
5. The size of the sample depends on the size of the population also. If the population is large, sample also is to be large.
6. Cost of selection of items also influences the size of the sample.

Approaches in sample size decision

There are two alternative approaches for determining size of sample. First approach is to specify the precision of estimation desired and then to determine sample size (n) necessary to ensure it. The second approach uses cost of additional information against expected value of additional information. The first approach is capable of solving a mathematical solution, and as such is a frequently used technique of determining sample size. The limitation of this technique is that it does not consider the cost of gathering information. The second approach is theoretically optimal, but is rarely used because of difficulty in measuring the value of information. The second approach employs estimation and guess work. Therefore, we shall concentrate here on the first approach.

Determining sample size –confidence level approach

When a sample study is made, sampling errors are bound to occur, and this can be controlled by selecting sample of adequate size. The precision level must be specified along with confidence level. Sample size can be determined considering such level of confidence, standard deviation and expected error.

Sample size for infinite population

In case of infinite populations, number of elements within the population is indefinite or uncertain. Normally populations belong to this group.

$$\text{Sample size } n = \frac{z^2 \times \sigma^2}{e^2}$$

Where n = sample size

Z = value of significance level = 1.96 or 2.58

σ = standard deviation

e = expected error

Sample size for finite population

In case of finite population, number of elements within population (N) will be certain and given, along with standard deviation and expected error.

$$n = \frac{Z^2 \times N\sigma^2}{(N-1)e^2 + Z^2 \times \sigma^2}$$

Where n = sample size

Z = value of significance level = 1.96 or 2.58

σ = standard deviation

e = expected error

N = Number of items in population

Calculation of sample size when standard deviation is given

Ex 13.1

A Production controller estimates that average production per day is 100 with standard deviation 4.8. He expects an error of ± 3 . How many items he should include in a sample, at 95% level of confidence

$$\sigma = 4.8 \quad e = 3 \quad Z = 1.96 \quad n = ?$$

$$n = \frac{Z^2 \times \sigma^2}{e^2} = \frac{1.96^2 \times 4.8^2}{3^2} = 9.834 \quad \text{Taken as 10}$$

Ex 13.2

Determine sample size for estimating weight of 5000 milk packets with variance of weight 4 gms, and expected error of .8 gms, at 99% level of confidence.

$$N = 5000 \quad Z = 2.58 \quad \sigma = \sqrt{4} = 2 \quad e = .8 \quad n = ?$$

$$n = \frac{Z^2 \times N\sigma^2}{(N-1)e^2 + Z^2 \times \sigma^2} = \frac{2.58^2 \times 5000 \times 2^2}{(4999) \times .8^2 + 2.58^2 \times 2^2} = 40.95 \quad \text{Taken as 41.}$$

Ex 13.3

Determine the number of items to be included in a sample to obtain estimated mean weight of Frooti, at 99% level of confidence, with standard deviation 2 and expected error of .8

$$\sigma = 2 \quad e = .8 \quad Z = 2.58 \quad n = ?$$

$$n = \frac{Z^2 \times \sigma^2}{e^2} = \frac{2.58^2 \times 2^2}{.8^2} = 41.28 \quad \text{Taken as 41}$$

Estimation Approach in Calculation of sample size

How big a sample should I select? , 'What should be my sample size?' and 'How many cases do I need?' These are the most common questions asked. Basically, it depends on what you want to establish. Your purpose in undertaking research is the main determinant of the level of accuracy required in the results, and this level of accuracy is an important determinant of sample size. However, in qualitative research, as the main focus is to explore or describe a situation, issue, process or phenomenon, the question of sample size is less important. You usually collect data till you think you have reached saturation point in terms of discovering new information. Once you think you are not getting much new data from your respondents, you stop collecting further information. Of course, the diversity or heterogeneity in what you are trying to find out about plays an important role in how fast you will reach saturation point. And remember : the greater the

heterogeneity or diversity in what you are trying to find out about, the greater the number of respondents you need to contact to research saturation point. In determining the size of your sample for quantitative studies and in particular for cause-and-effect studies, you need to consider the following:

- At what level of confidence do you want to test your results, findings or hypotheses?
- With what degree of accuracy do you wish to estimate the population parameters?
- What is the estimated level of variation (standard deviation), with respect to the main variable you are studying, in the study population?

Answering these questions is necessary regardless of whether you intend to determine the sample size yourself or have an expert do it for you. The size of the sample is important for testing a hypothesis or establishing an association, but for other studies the general rule is the larger the sample size, the more of your sample. Your skills in selecting a sample, within the constraints of your budget, lie in the way you select your elements so that they effectively and adequately represent your sampling population.

To illustrate this procedure let us take the example of a class. Suppose you want to find out the average age of the students within an accuracy of 0.5 of a year; that is; you can tolerate an error of half a year on either side of the true average age. Let us also assume that you want to find the average age within half a year of accuracy at 95 per cent confidence level; that is, you want to be 95 per cent confident about your findings.

There are a number of formulae for calculating sample size. One among them is

$$\sqrt{n} = \frac{1.96 \times \sigma}{\epsilon}, \text{ where } n = \text{sample size}, 1.96 = \text{value for 5% level of significance},$$

σ = standard deviation, and ϵ = expected error

There is only one unknown quantity in the above equation, that is σ .

Now the main problem is to find the value of σ , without having to collect data. This is the biggest problem in estimating the sample size. Because of this, it is important to know as much as possible about the study population.

The value of σ can be estimated by any one of the following methods:

1 statistical guessing,

2 consulting an expert;

3 obtaining the value of σ from previous comparable studies ; or

4 carrying out a pilot study to calculate the value.

Now in the class room example, we can assume that the σ or standard deviation = 1 year

$$\sqrt{n} = \frac{1.96 \times \sigma}{\epsilon} = \frac{1.96 \times 1}{0.5} = 3.92 \text{ that means } n = 16$$

Hence, to determine the average age of the class at a level of 95 per cent accuracy (assuming $\sigma = 1$ year) with $\frac{1}{2}$ year of error, a sample of at least 16 students is necessary.

Now assume that instead of 95 per cent, if you want to be 99 per cent confident about the estimate age, tolerating an error of $\frac{1}{2}$ year.

$$\sqrt{n} = \frac{2.58 \times 1}{0.5} \quad \sqrt{n} = 5.15 \quad n = 26.54 = 27$$

Hence, if you want to be 99 per cent confident and are willing to tolerate an error of $\frac{1}{2}$ year, you need to select a sample of 27 students. Similarly, you can calculate the sample size with varying values of standard deviation (σ). Remember the golden rule: the greater the sample size, the more accurately your findings will reflect the 'true' picture.

If standard deviation is not available

In the above formulae, standard deviations of population were given or assumed. But in many cases it may not be given or is not available or cannot be assumed. Since we have not yet taken the sample and are in the stage of deciding the size of sample, we cannot calculate standard deviation of population. In such a situation, if we have an idea about the range (difference between highest value and lowest value) we can estimate standard deviation of population as below

Ex . 13.1

Suppose we are estimating wages of workers in a village, and it is learned that there is difference of Rs 40 between the highest wage 640 and lowest wage 600

we know that $\pm 1.96 \sigma$ covers 95% of items in a study. It means 3.92σ covers the given range of population. Thus

$$\text{Given Range} = 3.92 \times \sigma = 40$$

$$\sigma = \frac{\text{Given Range}}{3.92} = \frac{40}{3.92} = 10.20 \text{ ie, about 10 workers}$$

The obtained standard deviation - 10, can be utilized for estimating sample size.

Sample size for estimating proportion

For estimating proportion, the number of elements in a sample is a decision factor which must be specified beforehand. Besides, the expected precision and the confidence

level also should be considered. Then, number of items to be included in a sample for estimating proportion value, will be:

$$n = \frac{z^2 \times p \times q}{e^2}$$

Where n = sample size

Z = value of significance level = 1.96 or 2.58

σ = standard deviation

e = expected error

Ex 13.4

What should be the size of sample, to be drawn from a population, to estimate per cent defective, within 2 % of true value, with 95% level of confidence? For this 100 items were selected, and obtained 2 defectives.

$$\begin{aligned} e &= .02 & Z &= 1.96 & p &= .02 & q &= .98 & n &=? \\ n &= \frac{z^2 \times p \times q}{e^2} = \frac{1.96^2 \times .02 \times .98}{.02^2} \\ &= 188.24 \text{ taken as } 188 \end{aligned}$$

Ex 13.5

In a hotel, 5 out of 100 visitors stay overnight. The management wants to be 95% confident that population percentage to be estimated with $\pm 3\%$ of true value. What should be minimum sample size?

$$\begin{aligned} e &= .03 & z &= 1.96 & p &= .05 & q &= .95 & n &=? \\ n &= \frac{z^2 \times p \times q}{e^2} = \frac{1.96^2 \times .05 \times .95}{.03^2} = 202.75 \text{ or } 203 \text{ visitors} \end{aligned}$$

Review Questions and Exercises

1. Is sampling a critical factor in researches
2. Explain non sampling errors
3. Define sample
4. What is sample size?
5. Explain optimum sample size
6. State the determining factors in sample size
7. Describe steps in determining sample size for mean test
8. What is the procedure in deciding sample size for proportion
9. What are the characteristics of a good sample
10. How is sample size determined for finite population
11. How is sample size determined if standard deviation is not given

12. Determine the size of sample for estimating true mean of the population of size 5000, on the basis of the following:
 Population variance = 4, Level of confidence = 99%, Estimated error = .4
13. what would be the size of sample for estimating mean of the population, if –
 Standard deviation = 15
 Estimated error = 6
 Level of confidence = 99%
14. Determine sample to be drawn from the population of 5000 units to estimate the percentage of defectives on the basis of 3% defectives in the sample within .05 units of its true value. Level of confidence desired is 95%.
15. What should be the size of the sample drawn from a population to estimate the percent defective within 2% of the true value with 95% level of confidence , on the basis of 3% defective in the sample.

UNIT 14 DATA COLLECTION

Introduction

The task of data collection begins after a research problem has been defined and research design/plan chalked out. While deciding about the method of data collection to be used for the study, the researcher should keep in mind two types of data viz., primary and secondary. The primary data are those which are collected afresh and for the first time, and thus happen to be original in character. The secondary data ,on the other hand ,are those which have already been passed through the statistical process. The researcher would have to decide which sort of data he would be using (thus collecting) for his study accordingly he will have to select one or the method of data collection. The methods of collecting primary and secondary data differ since primary data are to be originally collected, while in case of secondary data the nature of data collection work is merely that of compilation. We describe the different methods of data collection, with the pros and cons of each method.

Experiments and surveys

We collected primary data during the course of doing experiments in an experimental research. An experiment refers to an investigation in which factor or variable under test in isolated and its effect(s) measured. In an experiment the investigator measures the effects of an experiment which he conducts intentionally.

However ,in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through direct communication with respondents in one from or another or through personal interviews. Survey refers to the method of securing information concerning a phenomena under study from all or a selected number of respondents of the concerned universe. In a survey , the investigator examines experiment and a survey can be depicted as under.

Tools of Collecting primary data

There are a number of tools or methods of collecting primary data, particularly in surveys and descriptive researches. Some important methods are described below:

Observation Method

The observation method is the most commonly used method specially in studies relating to behavioral sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under the observation method, the information is sought by way of investigator's own direct observation without asking from the respondent. For instance, in a study relating to consumer behavior, the investigator instead of asking the brand of wrist watch used by the respondent, may himself look at the watch.

Merits of observation

1. The main advantage of this method is that subjective bias is eliminated, if observation is done accurately.
2. Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behaviour or future intention or attitudes.
3. Thirdly, this method is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method.
4. This method is particularly suitable in studies which deal with subjects or, respondents, who are not capable of giving verbal reports of their feelings for one reason or the other.

Demerits of observation

However, observation method has various limitations.

1. Firstly, it is an expensive method.
2. Secondly, the information provided by this method is very limited.
3. Thirdly, sometimes unforeseen factors may interfere with the observational task.
4. The fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

Precautions In observation

While using this method, the researcher should keep in mind certain aspects

What should be observed?

How the observation should be recorded?

How the accuracy of observation can be ensured?

Structured and unstructured observation

In case the observation is characterised by a careful definition of the units to be observed, the style of the recording the observed information, standardized conditions of observation and the selection of pertinent data of observation, then the observation is called as structured observation. But when observation is to take place without these characteristics to be thought of in advance, the same is termed as unstructured observation. Structured observation is considered appropriate in descriptive studies, whereas in an exploratory study the observational procedure is most likely to be relatively unstructured.

Participant and non participant observation

We often talk about participant and non-participant types of observation in the context of studies, particularly of social science . This distinction depends upon the observer's sharing or not sharing the life of the group he is observing. If the observer observes by making himself, more or less , a member of the group he is observing so that he can experience what the member of the group experience , the observation is called as the participant observation. But when the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of the type is often termed as non-participant observation. When the observer is observing in such a manner that his presence may be unknown to the people he is observing , such an observation is described as disguised observation.

Controlled and uncontrolled observation

Sometimes we talk of controlled and uncontrolled observation. If the observation takes place in the natural setting, it may be termed as uncontrolled observation, but when observation takes place according to definite pre-arranged plans. involving experimental procedure ,the same is then termed controlled observation. In non-controlled observation ,no attempt is made to use precision instruments. The major aim of this type of observation is to get a spontaneous picture of life and persons. It has a tendency to supply naturalness and completeness of behaviour, allowing sufficient time for observing it. But in controlled observation , we use mechanical (or precision) instruments as aids to accuracy and standardization .Such observation has a tendency to supply formalized data upon which generalization can be built with some degree of assurance. The main pitfall of non-controlled observation is that of subjective interpretation . There is also the danger of having the feeling that we know more about the observed phenomena than we actually do. Generally ,controlled observation takes place in various experiments that are carried out in a laboratory or under controlled conditions, whereas uncontrolled observation is resorted to in case of exploratory researches.

Interview Method

The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral –verbal responses. This method can be used through personal interviews and ,if possible, through telephone interviews.

Personal interview

Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons.(At times the interviewee may also ask certain questions and the interviewer responds to these, but usually the interviewer initiates the interview and collects the information.) This sort of interview may be in the form of direct personal investigation or it may be indirect oral investigation. In the case of direct personal investigation the interviewer has to collect necessary information personally from the sources concerned. He has to be on the spot and has to meet people from whom data have to be collected. This method is particularly suitable for intensive investigations. But in certain cases it may not be possible or worthwhile to contact directly the persons concerned or on account of the extensive scope of enquiry, the direct personal investigation technique may not be used. In such cases an indirect oral examination can be conducted under which the interviewer has not cross-examine other persons who are supposed to have knowledge about the problem under investigation and the information, obtained is recorded . Most of the commissions and committees appointed by government carry on investigations make use of this method.

Types of structured interviews

The method of collecting information through personal interviews is usually carried out in a structured way. As such we call them structured interviews. Such interviews involve the use of a set of predetermined questions and of highly standardized techniques of recording. Thus, the interviewer in a structured interview follows a rigid procedure laid down, asking questions in a form and order prescribed. As against it, the unstructured interviews are characterized by a flexibility of approach to questioning. Unstructured interviews do not follow a system of pre-determined questions and standardized techniques of recording information. In a non-structured interview, the interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so require. He may even change the sequence of questions. He has relatively greater freedom while recording the responses to include some aspects and exclude others. But this sort of flexibility results in lack of comparability of one interview with another and the analysis of unstructured responses becomes much more difficult and time-consuming than that of the structured responses obtained in case of structured interviews. Unstructured interviews also demand deep knowledge and greater skill on the part of the interviewer. Unstructured interview, however, happens to be the central technique of collecting information in case of exploratory or formulative research studies. But in case of descriptive studies, we quite often use the technique of structured interview because of its being more economical, providing a safe basis for generalization and requiring relatively lesser skill on the part of the interviewer.

Focused Interview

We may as well talk about focused interview, clinical interview and the non-directive interview. Focused interview is meant to focus attention on the given experience of the respondent and its effects. Under it the interviewer has the freedom to decide the manner and sequence in which the questions would be asked and has also the freedom to explore reason and motives. The main task of the interviewer in case of a focused interview is to confine the respondent to a discussion of issues with which he seeks conversance. Such interviews as used generally in the development of hypotheses and constitute a major type of unstructured interviews. The clinical interview is concerned with broad underlying feelings or motivations or with the course of individual's discretion. In case of non-directive interview, the interviewers' function is simply to encourage the respondent to talk about the given topic with a bare minimum of direct questioning. The interviewer often acts as a catalyst to a comprehensive expression of the respondents' feelings and beliefs and of the frame of reference within which such feeling and beliefs take on personal significance.

Despite the variations in interview-techniques, the major advantages and weakness of personal interviews can be enumerated in a general way. The chief merits of the interview method are as follows:

Telephone interviews:

This method of collecting information consists in contacting respondents one telephone itself. It is not a very widely used method, but plays important part in industrial surveys, particularly in developed regions. The chief merits of such a system are:

1. It is more flexible in comparison to mailing method.
2. It is faster than other methods i.e., a quick way of obtaining information.
3. It is cheaper than personal interviewing method; here the cost per response is relatively low.
4. Recall is easy; callbacks are simple and economical.

5. There is a higher rate of response than what we have in mailing method; the non-response is generally very low.
6. Replies can be recorded without causing embarrassment to respondents.
7. Interviewer can explain requirements more easily.
8. At times, access can be gained to respondents who otherwise cannot be contacted for one reason or the other.
9. No field staff is required.
10. Representative and wider distribution of sample is possible.

But this system of collecting information is not free from demerits. Some of these may be highlighted.

1. Little time is given to respondents for considered answers: interview period is not likely to exceed five minutes in most cases.
2. Survey are restricted to respondents who have telephone facilities.
3. Extensive geographical coverage may get restricted by cost considerations.
4. It is not suitable for intensive surveys where comprehensive answers are required to various questions.
5. Possibility of the bias of the interviewer is relatively more.
6. Questions have to be short and to the point ; probes are difficult to handle.

Precautions in Interview

Interviewing is an art and one learns it by experience. However , following points may be kept in view by an interviewer for eliciting the desired information:

1. Interviewer must plan in advance and should fully know the problem under consideration. He must choose a suitable time and place so that the interviewee may be at ease during the interview period. For this purpose some knowledge of the daily routine of the interviewee is essential.
2. Interviewer's approach must be friendly and informal . Initially greeting in accordance with the cultural pattern of the interviewee should be exchanged and then the purpose of the interview should be explained.
3. All possible effort should be made to establish proper rapport with the interviewee. People are motivated to communicate when the atmosphere is favorable.
4. Interviewer must know that ability to listen with understanding ,respect and curiosity is the gateway to communication, and hence must act accordingly during the interview. For this, the interviewer must be intelligent and must be a man with self -restraint and self discipline.
5. To the extent possible there should be a free-flowing interview and the questions must be well phrased in order to have full cooperation of the interviewee. But the interviewer must control the course of the interview in accordance with the objective of the study.
6. In case of big enquiries , where the task of collecting information is to be accomplished by several interviewers, there should be an interview guide to be observed by all so as to ensure reasonable uniformity in respect of all salient points in the study.

Collecting of Data through Questionnaire

This method of data collection is quite popular , particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organization

and even by governments. In this method a questionnaire is sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire. A questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the question on their own.

The method of collecting data by mailing the questionnaires to respondents is most extensively employed in various economic and business surveys. The merits claimed on behalf of this method are as follows:

1. There is low cost even when the universe is large and is widely spread geographically?
2. It is free from the bias of the interviewer; answers are in respondents' own words.
3. Respondents have adequate time to give well thought out answers.
4. Respondents have adequate time to give well thought out answers.
5. Large samples can be made use of and thus the results can be made more dependable and reliable.

The main demerits of this system can also be listed here:

1. Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
2. It can be used only when respondents are educated and cooperating.
3. The control questionnaire may be lost once it is sent.
4. There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been dispatched.
5. There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
6. It is difficult to know whether willing respondents are truly representative.
7. This method is likely to be the slowest of all.

Before using this method, it is always advisable to conduct 'pilot study'(pilot survey) for testing the questionnaires. In a big enquiry the significance of pilot survey is felt very much. Pilot survey is in fact the replica and rehearsal of the main survey. Such a survey, being conducted by experts, brings to the light the weaknesses(if any) of the questionnaires and also of the survey techniques. From the experience gained in this way, improvement can be effected.

Main aspects of a questionnaire:

Quite often questionnaire is considered as the heart of survey operation. Hence it should be very carefully constructed. If it is not properly set up, then the survey is bound to fail. This fact requires us to study the main aspects of a questionnaire viz., the general form, question sequence and question formulation and wording. Researcher should note the following with regard to these three main aspects of a questionnaire:

General form :

So far as the general form of a questionnaire is concerned, it can either be structured or unstructured questionnaire. Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questions are presented with exactly the same wording and in the same order to all respondents. Resort is taken to this sort of standardization to ensure that all respondents reply to the same set of questions. The form of the question may be either closed(i.e., of the

type 'yes or no') or open (i.e., inviting free response) but should be stated in advance and not constructed during questioning. Structured questionnaires may also have fixed alternatives. Questionnaire is one in which all questions and answers are specified and comments in the respondents' own words are kept to the minimum.

When these characteristics are not present in a questionnaire, it can be termed as unstructured or non-structured questionnaire. More specifically, we can say that in an unstructured questionnaire, the interviewer is provided with a general guide on the type of information to be obtained, but the exact question formulation is largely his own words to the extent possible. In some situations tape recorders may be used to achieve this goal.

Structured questionnaires are simple to administer and relatively inexpensive to analyse. The provision of alternative replies, at times, helps to understand the meaning of the question clearly. But such questionnaires have limitations too. For instance, wide range of data that too in respondent's own words cannot be obtained with structured questionnaires. They are usually considered inappropriate in investigations where the aim happens to be to probe for attitudes and reasons for certain actions or feelings. They are equally not suitable when a problem is being first explored and working hypotheses sought. In such situations, unstructured questionnaires may be used effectively. Then on the basis of the results obtained in pretest (testing before final use) operation from the use of unstructured questionnaires, one can construct a structured questionnaire for use in the main study.

Question sequence:

In order to make the questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chance of individual questions being misunderstood. The question-sequence must be clear and smoothly-moving, meaning thereby that the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because they are likely to influence the attitude of the respondent and in seeking his desired cooperation. The opening questions should be such as to arouse human interest. The following type of questions should generally be avoided as opening question in a questionnaire:

1. Questions that put too great a strain on the memory or intellect of the respondent;
2. Questions of a personal character;
3. Questions related to personal wealth, etc.

Following the opening questions, we should have questions that are really vital to the research problem and a connecting thread should run through successive questions. Ideally, the question-sequence should conform to the respondent's way of thinking. Knowing what information is desired the researcher can rearrange the order of the questions. This is possible in case of unstructured questionnaire, to fit the discussion in each particular case. But in a structured questionnaire the best that can be done is to determine the question-sequence with the help of a pilot Survey which is likely to produce good rapport with most respondents. Relatively difficult questions must be relegated towards the end so that even if the respondent decides not to answer such questions, considerable information would have already been obtained. Thus question-sequence should usually go from the general to the more specific and the researcher must always remember that the answer to a given question is a function not only of the question itself, but of all previous questions as well. For instance, if one question deals with the price usually paid for coffee and the next with

reason for preferring that particular brand, the answer to this latter question may be couched largely in terms of price-differences.

Question formulation and wording:

With regard to this aspect of questionnaire, the researcher should note that each question must be very clear for any sort of misunderstanding can do irreparable harm to survey. Question should also be impartial in order not to give a biased picture of the true state of affairs. Questions should be constructed with a view to their forming a logical part of a well throughout tabulation plan. In general, all questions should meet the following standards-(a) should be easily understood; (b) should be simple i.e., should convey only one thought at a time; (c) should be concrete and should conform as much as possible to the respondent's way of thinking. For instance, instead of asking, "How many razor blades do you use annually?" The more realistic question would be to ask, "How many razor blades did you use last week?"

Concerning the form of questions, we can talk about two principal forms, viz., multiple choice question and the open-end question. In the former the respondent selects one of the alternative possible answers put to him, whereas in the latter he has to supply the answer in his own words. The question with only two possible answers (usually 'Yes or No') can be taken as special case of the multiple choice question, or can be named as a 'closed question'. There are some advantage and disadvantage of each possible form of question. Multiple choice or closed questions have the advantages of easy handling, simple to answer, quick and relatively inexpensive to analyse. They are most amenable to statistical analysis. Some times, the provision of alternative replies helps to make clear the meaning of the question. But the main drawback of fixed alternative questions is that of "putting answers in people's mouths" i.e., they may force a statement of opinion on an issue about which the respondent does not in fact have any opinion on an issue about which the respondent does not in fact have any opinion. They are not appropriate when the issue under consideration happen to be a complex one and also when the interest of the researcher is in the exploration of a process. In such situations, open-ended questions which are designed to permit a free response from the respondent rather than one limited to certain stated alternatives are considered appropriate. Such questions give the respondent considerable latitude in phrasing a reply. Getting the replies in respondent's own words is, thus, the major advantage of open-ended questions. But one should not forget that, from an analytical point of view, open-ended questions are more difficult to handle, leading to problems of interpretation, comparability and interviewer bias.

In practice, one rarely comes across a case when one questionnaire relies on one form of questions alone. The various forms complement each other. As such questions of different forms are included in one single questionnaire. For instance, multiple-choice questions constitute the basis of a structured questionnaire, particularly in a mail survey. But even there, various open-ended questions are generally inserted to provide a more complete picture of the respondent's feelings and attitudes.

Researcher must pay proper attention to the wordings of questions since reliable and meaningful returns depend on it to a large extent. Since words are likely to affect responses, they should be properly chosen. Simple words, which are familiar to all respondents should be employed. Words with ambiguous meaning must be avoided. Caution must also be exercised in the use of phrases which reflect upon the prestige of the respondent. Question wording, in no case, should bias the answer. In fact, question wording and formulation is an art and can only be learned by practice.

Essential of a good questionnaire :

1. To be successful ,questionnaire should be comparatively short and simple i.e., the size of the questionnaire should be kept to the minimum.
2. Questions should proceed in logical sequence moving from easy to more difficult questions. Personal and intimate questions should be left to the end.
3. Technical terms and vague expressions capable of different interpretation should be avoided in an questionnaire .
4. Questions may be dichotomous (yes or no answer) , multiple choice(alternative answers listed) or open-ended. The latter type of the respondent .For instance, a question designed to determine the consumption of particular material may be asked first in terms of financial expenditure and latter in terms of weight.
5. The control questions, thus, introduce a cross-check to see whether the information collected is correct or not.
6. Questions affecting the sentiments of respondents should be avoided.
7. Adequate space for answers should be provided in the questionnaire to help editing and tabulation. There should always be provision for indications of uncertainty, e.g., " do not know." " no preference" and so on.
8. Brief directions with regard to filling up the questionnaire should invariably be given in the questionnaire itself.
9. Finally , the physical appearance of the questionnaire affects the cooperation the researcher receives from the recipients and as such an attractive looking questionnaire, particularly in mail survey, is a plus point for enlisting cooperation. The quality of the paper, along with its colour , must be good so that it may attract the attention of recipients.

Nowadays, the usage of internet is quite popular in collecting data . Questionnaires are sent through e-mails to randomly or purposely selected people. Their e-mail ids can be taken from some other surveys. The trend of paid surveys had also started. There are some website which pays for filling the questionnaires.

Collection of Data through Schedules

This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules , or proforma containing a set of questions are being filled in by the enumerators who are specially appointed for the purpose. These enumeration along with schedules, go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same in the proforma. In certain situations, schedules may be handed over to respondents and enumerators may help them in recording their answers to various questions in the said schedules. Enumerators explain in the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question or the definition or concept of difficult terms.

This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected. The enumeration should be trained to perform their job well and nature and scope of the investigation should be explained to them thoroughly so that they may well understand the implications of different questions put in the schedule. Enumerators should be intelligent and must possess the capacity of cross examination in order to find out the truth. Above all , they should be honest ,sincere, hardworking and should have patience and perseverance.

Difference between Questionnaire and schedule

Both questionnaire and schedule are popularly used methods of collecting data in research surveys. There is much resemblance in the nature of these two methods and this fact has made many people to remark that from a practical point of view, the two methods can be taken to be the same. But from the technical point of view there is difference between the two. The important points of difference are as under:

1. The questionnaire is generally sent through mail to informants to be answered as specified in a covering letter, but otherwise without further assistance from the sender. The schedule is generally filled out by the research worker or the enumerator, who can interpret questions when necessary.
2. To collect data through questionnaire is relatively cheap and economical since we have to spend money only in preparing the questionnaire and in mailing the same to respondents. Here no field staff required. To collect data through schedules is relatively more expensive since considerable amount of money has to be spent in appointing enumerators and in imparting training to them. Money is also spent in schedules.
3. Non-response is usually high in case of questionnaire as many people do not respond many return the questionnaire without answering all questions. Bias due to non-response often remains indeterminate. As against this ,non-response is generally very low in case of schedules. But there remains the danger of interviewer bias and cheating.
4. In case of questionnaire, it is not always clear as to who replies, but in case of schedule the identity of respondent is known .
5. The questionnaire method is likely to be very slow since many respondents do not return the questionnaire in time despite several reminders, but in case of schedules the information is collected well in time as they are filled in by enumerators.
6. Personal contact is generally not possible in case of the questionnaire method as questionnaires are sent to respondents by post who also in turn return the same by post. But in case schedules direct personal contact is established with respondents.
7. Questionnaire method can be used only when respondents are literate and cooperative; but in case of schedules the information can be gathered even when the respondents happen to be illiterate.
8. Wider and more representative distribution of sample is possible under the questionnaire method, but in respect of schedules there usually remains the difficulty in sending enumerators over a relatively wider area.
9. Risk of collecting incomplete and wrong information relatively more under the questionnaire method, particularly when people are unable to understand questions properly . But in case of schedules, the information collected is generally complete and accurate as enumerators can remove the difficulties, if any, faced by respondents in correctly understanding the questions. As a result, the information collected through schedules is relatively more accurate than that obtained through questionnaires.
- 10.The success of questionnaire method lies more on the equality of the questionnaire it self, but in the case of schedules much depends upon the honesty and competence of enumerators.
- 11.In order to attract the attention of respondents , the physical appearance of questionnaire must be quite attractive , but this may not be so in case of schedules as they are to be filled in by enumerators and not by respondents.

12. Along with schedules, observation method can also be used but such a thing is not possible while collecting data through questionnaires.

Guidelines for Constructing Questionnaire/Schedule

The researcher must pay attention to the following points in constructing an appropriate and effective questionnaire or a schedule.

1. The researcher must keep in view the problem he is to study for it provides the starting point for developing the Questionnaire /Schedule. He must be clear about the various aspects of his research problem to be dealt with in the course of his research project.

2. Appropriate form of questions depends on the nature of information sought, the sampled respondents and the kind of analysis intended. The researcher must be constructed with a view to their forming a logical part of a well thought out tabulation plan. The units of enumeration should also be defined precisely so that they can ensure accurate and full information.

3. Rough draft of the Questionnaires/Schedule be prepared, giving due thought to the appropriate sequence of putting questions. Questionnaires or schedule previously drafted(if available) may as well be looked into at this stage.

4. Researcher must invariably re-examine, and in case of need may revise the rough draft for a better one. Technical defects must be minutely scrutinized and removed.

5. Pilot study should be undertaken for pre-testing the questionnaire. The questionnaire may be edited in the light of the result of the pilot study.

6. Questionnaires must contain simple but straight forward directions for the respondents so that they may not feel any difficulty in answering the questions.

Innovative Methods of Data Collection

Let us consider some other innovative or modern methods of data collection, particularly used by big business house in modern times. They are extensively employed in market researches.

Warranty cards:

Warranty cards are usually postal sized cards which are used by dealers of consumer durables to collect information regarding their products. The information sought is printed in the form of questions on the 'warranty cards' which is placed inside the package along with the product with a request to the consumer to fill in the card and post it back to the dealer.

Distributor or store audits:

Distributor or store audits are performed by distributors as well as manufacturers through their salesmen at regular intervals. Distributors get the retail stores audited through salesman and use such information to estimate market size, market share, and seasonal purchasing pattern and so on. The data are obtained in such audits not by questioning but by observation. For instance, in case of a grocery store audit, a sample of stores is visited periodically and data are recorded on inventories on hand either by observation or copying from store records. Store audits are invariably panel operations, for the derivation of sales estimates and compilation of sales trends by stores are their principal 'raison detre'. The principal advantage of this method is that it offers the most efficient way of evaluating the effect on sales of variations of different techniques of in-store promotion.

Pantry audits:

Pantry audit technique is used to estimate consumption of the basket of goods at the consumer level. In this type of audit, the investigator collects an inventory of types, quantities and prices of commodities consumed. Thus in pantry audit data recorded from the examination of consumer's pantry. The usual objective in pantry audit is to find out what

types of consumers buy certain brands., the assumption being that the contents of the pantry accurately portray consumer's preferences. Quite often , pantry audits are supplemented by direct questioning relating to reason and circumstances under which particular products were purchased in an attempt to relate these factors to purchasing habits. A pantry audit may or may not be set up as a panel operation, since a single visit is often considered sufficient to yield an accurate picture of consumers' preference . An important limitation of pantry audit approach is that, at times, it may not be promotion devices produce a marked rise in sales.

Consumer panels:

An extension of the pantry audit approach on a regular basis is known to maintain detailed daily records of their consumption and the same is made available to investigator on demands. In other words , a consumer panel is essentially a sample of consumers who are interviewed repeatedly over a period of time. Mostly consume panels are of two types viz., the transitory consumer panel . A transitory consumer panel is set up to measure the effect of a particular phenomenon. Usually such a panel is conducted on a before-and-after – basis. Initial interviews are conducted before the phenomenon has taken place to find out the consequent changes that might have occurred in the consumer's attitude. It is a favorite tool of advertising and of social research. A continuing consumer panel is often set up for an indefinite period with a view to collect data on particular aspect of consumer behaviour over time, generally at periodic intervals or may be meant to serve as a general purpose panel for researchers on a variety of subjects. Such panels have been used in the area of consumer expenditure, public opinion and radio and TV listenership among others. Most of these panels operate by mail. The representativeness of the panel is relative to the population and the effect of panel for researchers on a variety of subjects. Such panels relative to the population and the effect of panel membership on the information obtained after the two major problems associated with the use of this method of data collection.

Use of mechanical devices:

The use of mechanical devices has been widely made to assess psychological impact and attitude . one of the prominent device was psycho galvanometer. Motion picture camera and Audiometer are the principal devices so far developed and commonly used by modern big business houses , mostly in the developed world for the purpose of collecting the required information

Eye cameras are designed to record the focus of yes of a respondent on a specific portion of a sketch or diagram or written material. Such information is useful in designing advertising material. Pupil metric cameras record dilation of the pupils as a result of a visual stimulus. The extent of dilation shows the degree of interest aroused by the stimulus. psycho galvanometer is used for measuring the extent of body excitement as a result of the visual stimulus. Motion picture cameras can be used to record movement of body of a buyer while deciding to buy consumer goods, from a shop or big store. Influence of packaging or the information given on the label would stimulate a buyer to perform certain physical movements which can easily be recorded by a hidden motion picture camera in the shop's four walls. Audiometers are used by some TV concerns to find out the type of programmes as well as stations preferred by people. A device is fitted in the television instrument itself to record these changes. Such data may be used to find out the market share of competing television stations.

Projective techniques:

Projective techniques (or what are called as indirect interviewing techniques) for the collection of data have been developed by psychologists to use projection of respondents for inferring about underlying motives, urges, or intention which are such that the

respondent either resists to reveal them or is unable to figure out himself. In projective techniques the respondent in supplying information tends unconsciously to project his own attitudes or feeling on the subject under study. Projective techniques play an important role in motivational researches or in attitude surveys.

Use of these techniques requires intensive specialized training. In such techniques, the individual's responses to the stimulus-situation are not taken at their face value. The stimuli may arouse many different kind of reaction. The nature of the stimuli and the way in which they are presented under these techniques do not clearly indicate the way in which the response is to be interpreted. The stimulus may be a photography, a picture, an inkblot and so on. Responses to these stimuli are interpreted as indicating the individual's own view, his personality structure, his needs, tensions, etc. in the context of some-pre established psychological conceptualization of what the individual's responses to the stimulus mean.

We may now briefly deal with the important projective techniques.

Word association tests

These tests are used to extract information regarding such words which have maximum association. In this sort of test the respondents is asked to mention the first words that comes to mind, ostensibly without thinking, as the interviewer reads out each word from a list. If the interviewer says cold, the respondent may say hot and the like ones. The general technique is to use a list of as many as 50 to 100 words. Analysis of the matching words supplied by the respondents indicates whether the given word should be used for the contemplated purpose. The same idea is exploited in marketing research to find out the quality that is mostly associated to a brand of a product. A number of qualities of a product may be listed and informants may be asked to write brand names possessing one or more of these. This technique is quick and easy to use, but yields reliable results when applied to words that are widely known and which possess essentially one type of meaning. This technique is frequently used in advertising research.

Sentence completion tests

These tests happen to be an extension of the technique of word association tests. Under this, informant may be asked to complete a sentence(s) such as : persons who wear Khadi are...) to find association of khadi clothes with certain personality characteristic. Several sentences of this type might be put to the informant on the same subject. Analysis of replies from the same informant reveals his attitude toward that subject, and the combination of these attitudes of all the sample members is then taken to reflect the views of the population. This technique permits the testing not only of words (as in case of word association tests), But of ideas as well and thus, helps in developing hypotheses and in the construction of questionnaires. This technique is also quick and easy to use, but it often leads to analytical problem, particularly when the response happens to be multidimensional.

Story completion tests:

Such tests are a step further wherein the researcher may contrive stories instead of sentences and ask the informants to complete them. The respondent is given just enough of story to focus his attention on a given subject and he is asked to supply a conclusion to the story.

Verbal projection tests :

These are the tests wherein the respondent is asked to comment on or to explain what other people do. For example ,why do people smoke? Answers may reveal the respondent's own motivation.

Pictorial techniques: There are several pictorial techniques. The important ones are as follows:

Thematic Apperception Test (T.A.T.)

TAT consists of a set of pictures. Some of the pictures deal with ordinary day-to-day events while others may be ambiguous pictures of unusual situations that are shown to respondents forms the basis for the investigator to draw inferences about their personality structure ,a attitudes, etc.

Rosenzweig test:

This test uses a cartoon format wherein we have a series of cartoons with words inserted in 'balloons' above. The respondent is asked to put his own words in an empty balloon space provided for the purpose in the picture. From what the respondents write in this fashion, the study of their attitudes can be made.

c) Rorschach test:

This test consists of ten cards having print of inkblots. The design happens to be symmetrical but meaningless. The respondents are asked to describe what they perceive in such symmetrical inkblots and the responses are interpreted on the basis of some pre-determined psychological framework. This test is frequently used but the problem of validity still remains a major problem of this test.

Holtzman inkblot test(HIT) :

This test from W.H.Holtzman is a modification of the Rorschach Test explained above . This test consists of 45 inkblot cards(and not 10 inkblots as we find in case of Rorschach Test) which are based on colour ,movement, shading and other factors involved in inkblot perception. Only one response per card is obtained from the subject(or the respondent) and the responses of a subject are interpreted at three levels of from appropriateness. Form responses are interpreted for knowing the accuracy(F) or inaccuracy(F-) of respondents; shading and colour for ascertaining his affectional and emotional needs; and movement responses for assessing the dynamic aspects of his life.

Special features or advantages.

Firstly, It elicits relatively constant number of responses per respondent. Secondly, it facilitates studying the responses of a respondent to different cards in the light of norms of each card instead of lumping them together. Thirdly, it elicits much more information from the respondent than is possible with merely 10 cards in Roschach test; the 45 cards used in this test provide a variety of stimuli to the respondent and as such the range of responses elicited by the test is comparatively wider.

There are some limitations of this tests as well. One difficulty that remains that remains in using this test is that most of the respondents do not know the determinants of their perceptions, but for the researcher , who as to interpret the protocols of a subject and understand his personality(or attitude) through them, knowing the determinant of each his response is a must. This fact emphasizes that the test must be administered individually and a post-test inquiry must as well be conducted for knowing the nature and sources of responses and this limits the scope of HIT as a group test of personality.

In view of these limitations, some people have made certain changes in applying this test. For instance, Fisher and Cleveland in their approach for obtaining Barrier scope of an individual's personality have developed a series of multiple choice items for 40 of HIT cards. Each of these cards is presented to the subject along with three acceptable choices , such as 'Knight in armor'(Barrier response), 'X-Ray' (Penetrating response) and 'Flower'(Neutral response)]. Subject taking the test is to check the choice the likes most, make a different mark against the one the likes least and leave the third choice blank. The number responses checked by him determines his barrier score on the test.

Tomkins-Horn picture arrangement test :

This test is designed for group administration. It consists of twenty-five plates, each containing three sketches that may be arranged different ways to portray sequence of events. The respondent is asked to arrange them in a sequence which he considers as reasonable. The responses are interpreted as providing evidence confirming certain norms, respondent's attitudes, etc.

Play techniques:

Under play techniques subjects are asked to improvise or act out a situation in which they have been assigned various roles. The researcher may observe such traits as hostility, dominance, sympathy, prejudice or the absence of such traits. These techniques have been used for knowing the attitudes of younger ones through manipulation of dolls. Dolls representing different racial groups are usually given to children who are allowed to play with them freely. The manner in which children organize dolls would indicate their attitude towards the class of persons represented by dolls. This is also known as doll play test, and is used frequently in studies pertaining to sociology. The choice of colour, form, words, the sense of orderliness and other reactions may provide opportunities to infer deep-seated feelings.

Quizzes, tests and examinations: This is also a technique of extracting information regarding specific ability of candidates indirectly. In this procedure both long and short questions are framed to test through them the memorizing and analytical ability of candidates.

Sociometry :

Sociometry is a technique for describing the social relationship among individuals in a group. In an indirect way, sociometry attempts to describe attractions or repulsions between individuals by asking them to indicate whom they would choose or reject in various situations. Thus sociometry is a new technique of studying the underlying motives of respondents. Sociograms are charts that depict the sociometric choice. There are many versions of the sociogram pattern and the reader is suggested to consult specialized references on sociometry for the purpose. This approach has been applied to the diffusion of ideas on drugs amongst medical practitioners.

Depth interviews:

Depth interviews are those interviews that are designed underlying motives and desires and are often used in motivational research. Such interviews are held to explore needs, desire and feelings of respondents. In other words, they aim to elicit unconscious as also other types of material relating especially to personality dynamics and motivations. As such, depth interviews require great skill on the part of the interviewer and at the same time involve considerable time. Unless the researcher has specialized training, depth interviewing should not be attempted.

Depth interview may be projective in nature or it may be a non-projective interview. The difference lies in the nature of the questions asked. Indirect questions on seemingly irrelevant subjects provide information that can be related to the informant's behaviour or attitude towards the subject under study, thus, for instance, the informant may be asked on his frequency of air travel and he might again be asked at a later stage to narrate his opinion concerning the feelings of relatives of some other man who gets killed in an airplane accident. Reluctance to fly can then be related to replies to questions of the latter nature. If the depth interview involves questions of such type, the same may be treated as projective depth interview. But in order to be useful, depth interviews do not necessarily have to be projective in nature; even non-projective depth interviews can reveal important aspects of psycho-social situation for understanding the attitudes of people.

Content-analysis :

Content-analysis consists of analyzing the contents of documentary materials such as books., magazines, newspapers and the contents of all other verbal materials which can be either spoken or printed. Content-analysis prior to 1940's was mostly quantitative analysis of documentary materials concerning certain characteristics that can be identified and counted. But since 1950's content-analysis is mostly qualitative analysis concerning the general import or message of the existing documents.

Content-analysis is a central activity whenever one is concerned with the study of the nature of the verbal materials. A review of research in any area, for instance, involves the analysis of the contents of research articles that have been published. The analysis may be at a relatively simple level or may be a subtle one. It is at a simple level when we pursue it on the basis of certain characteristic of the document or verbal materials that can be identified and counted(such as on the basis of major scientific concepts in a book). It is at a subtle level when researcher makes a study of the attitude, say of the press towards education by feature writers.

Collection of secondary data

Secondary data means data that are already available I.e., they refer to the data which have already been collected and analyzed by someone else. When the researcher utilizes secondary data, then he has to look into various sources from where he can obtain them. In this case he is certainly not confronted with the problems that are usually associated with the collection of original data. Secondary data may either be published data or unpublished data. Usually published data are available in: (a) various publications of the central, state are local governments;(b) various publications of foreign governments or international bodies and their subsidiary organizations;(c) technical and trade journal;(d) books, magazines and newspapers ;(e) reports and publications of various associations connected with business and industry ,banks, stock exchanges, etc.;(f) reports prepared by research scholars, universities ,economists, etc. in different fields; and (g) Public records and statistics, historical documents, and other sources of published information.

Nowadays, data are published on the websites also. For example, the official website of reserve bank of India, National stock exchange , etc. Most of the data freely available and provided in Excel worksheets. The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and also may be available with scholar and research workers, trade associations, labour bureaus and other public/private individuals and organizations.

Researcher must be very careful in using secondary data . He must make a minute scrutiny because it is just possible that the secondary data may be unsuitable or may be inadequate in the context of the problem which the researcher wants to study .In this connection Dr. A. L. Bowley very aptly observes that it is never safe to take published statistic at their face value without knowing their meaning and limitations and it always necessary to criticize arguments that can be based on buy way of caution ,the researcher ,before using secondary data, must see that they possess following characteristics:

1. Reliability of data:

The reliability can be tested by finding out such things about the data: (a) who collected the data?(b) What were the sources of data?(c) Were they collected by using proper methods(d) At what were they collected? (e) Was there any bias of the complier ?(f) What level of accuracy was desired? (g) Was it achieved?

2. Suitability of data : The data that are suitable for one enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher. In this context, the researcher must very carefully scrutinize the definition of various term and units of collection used at the time of collecting the data from the primary source , originally.
3. Similarly, the object, scope and nature of the original enquiry must also be studied. If this researcher finds differences in these, the data will remain unsuitable for the present enquiry and should not be used.
4. Adequacy of data : If the level of accuracy achieved in data is found inadequate for the purpose of the present enquiry, they will be considered as inadequate and should not be used by the researcher. The data will also be considered as inadequate, if they are related to an area which may be either narrower or wider than the area of the present enquiry.

From all this we can say that it is very risky to use the already available data. The already available data should be used by the researcher only when he finds them reliable, suitable, and adequate. But he should not blindly discard the use of such data if they are readily available from authentic sources and are also suitable and adequate for in that case it will not be economical to spend time and energy in field surveys for collecting information. At times , there may be wealth of usable information in the already available data which must be used by an intelligent researcher but with due precaution.

Selection of appropriate method for data collection

Thus, there are various methods of data collection. As such the researcher must judiciously select the method/methods for his own study, keeping in view the following factors:

1. Nature, scope and object of enquiry:

This constitutes the most important factor affecting the choice of particular method selected should be such that it suit the type of enquiry that is to be conducted by the researcher. This factor is also important in deciding whether the data already available(secondary data) are to be used or the data not yet available(primary data) are to be collected.

2. Availability of funds:

Availability of funds for the research project determines to a large extent the method to be used for the collection of data. When funds at the disposal of the researcher are very limited , he will have to select a comparatively cheaper method. Finance ,in fact, is a big constraint in practice and the researcher has to act within this limitation.

3. Time factor:

Availability of time also is to be taken into account in deciding a particular method of data collection. Some methods take relatively more time, whereas with others the data can be collected in a comparatively shorter duration. The time at the disposal of the researcher, thus, affects the selection of the method by which the data are to be collected.

4. Precision required : Precision required is yet another important factor to be considered at the time of selecting the method of collecting of data.

But one must always remember that each method of data collection has its uses and none is superior in all situations. For instance, telephone interview method may be considered appropriate (assuming telephone population) if funds are restricted, time is also restricted and the data is to be collected in respect of few terms with or without a certain degree of precision. In case funds permit and more information is desired, personal interview method may be said to be relatively better. In case time is ample, funds are limited and much information is to be gathered with no precision, then mail-questionnaire method can be regarded more reasonable. When funds are ample, time is also ample and much information with no precision is to be collected, then either personal interview or the mail-questionnaire or the joint use of these two methods may be taken as an appropriate method of collecting data. Where a wide geographic area is to be covered, the use of mail-questionnaires supplemented by personal interview will yield more reliable results per rupees spent than either method alone. The secondary data may be used in case the researcher finds the reliable, adequate and appropriate for his research. While studying motivating influence in market researches or studying people's attitudes in psychological/social surveys, we can resort to the use of one or more of the projective techniques stated earlier. Such techniques are of immense value in case of attitude and the respondent is not aware of it. But when the respondent knows the reason and can tell the same if asked, than a non-projective questionnaire, using direct questions, may yield satisfactory results even in case of attitude surveys. Since projective techniques are as yet in an early stage of development and with the validity of the many of them remaining an open question, it is usually considered better to rely on the straight forward statistical methods with only supplementary use of projective techniques. Nevertheless in pre-testing and in searching for hypotheses they can be highly valuable.

Thus, the most desirable approach with regard to the selection of the method depends on the nature of the particular problem and on the time and resources(money and personnel) available along with the desired degree of accuracy. But, over and above all this, much depends upon the ability and experience of the researcher. Dr. A. L. Bowley's remark in this context is very appropriate when he says that "in collection of statistical data common sense is the chief requisite and experience the chief teacher."

Case study method

The case study method is a very popular form of qualitative analysis and is considered nowadays a common tool of data collection. It involves a careful and complete observation of a social unit, be that unit a person, a family, an institution, a cultural group or even the entire community. It is a method of study in depth rather than breadth. The case study deals with the process that take place and their interrelationship. Thus, case study is essentially an intensive investigation of the particular unit under consideration. The object of the case study method is to locate the factors that account for the behavior-patterns of the given unit as an integrated totality.

Review Questions

1. What is meant by data collection
2. State the role of data collection in research
3. Classify data collection techniques
4. What is primary data
5. What is secondary data
6. Differentiate between primary data and secondary data
7. What are the sources of secondary data

8. What is meant by observation method
9. What is interview method of data collection
10. What is personal interview
11. What are different types of structured interviews
12. Explain telephone interview
13. What are the precautions to be taken before interview
14. What is meant by a questionnaire
15. What are the essentials of a good questionnaire
16. What is meant by a Schedule
17. Differentiate between questionnaire and schedule
18. What are projective techniques
19. Explain some of the projective techniques

UNIT 15 MASUREMENT AND DATA

Introduction

All types of research require data. So, the basic problem of any research is to find facts and figures relating to particular phenomenon under study. The success of research depends on the availability and deployment of data.

In this world, data is continuously generated in large volume – whether for research or recording. When we say that Messi wears Jersey No 10, or Neymar is No 1 among Brazilian strikers with 48 goals, or Kozhikode is 40 kilo meters away from Areacode, or the monthly average expenditure of a student on mobile recharge is Rs 150, really we are generating data. The process of generating data, by assigning numbers or figures led to the idea of measurement. Measurement is a general term referring to the process of giving numbers to concepts and phenomena. After measuring, we have to place the measured object at appropriate scales, so as to enable further manipulation. This process which is more specific is called scaling. Measurement is simple and takes only four forms, whereas, scaling is more complex and takes a number of forms, some of which are even overlapping. This unit deals with the aspect of measurement, and the next unit is devoted to discuss scaling process.

Process of Measurement

In our daily life we are said to measure when we use some yardstick to determine weight, height or some other features of a physical object. We also measure when we judge how well we like song, a painting or personalities of our friends. We, thus, measure physical objects as well as abstract concepts. Measurement is a relatively complex and demanding task, especially so when it concerns qualitative or abstract phenomena. Other examples of qualitative characteristic are taste, honesty, intelligence, consumer's perception and brand loyalty etc. These characteristics are also called as constructs. The feeling and perception of customers and employees are very important for a company.

It is easy to assign number in respect of properties of some objects, but it is relatively difficult in respect of others. For instance, measuring such things as social conformity, intelligence or marital adjustment is much less obvious and requires much closer attention than measuring physical weight, biological age or person's financial assets. In other words, properties like weight, height, etc., can be measured directly with some standard unit of measurement, but it is not that easy to measure properties like motivation to succeed, ability

to stand stress and the like. For the meaningful assessment of the qualitative characteristic it is essential that they are also measured.

Quantitative and qualitative data

Measurement is defined as a process of associating numbers or symbols to observation obtained in a research study. These observations could be qualitative or quantitative.

Most of the analysis can be conducted using quantitative data. For example ,mean, standard deviation, etc. can be computed for quantitative characteristic. Qualitative characteristic can be counted and cannot be computed. Therefore, the researcher must have a clear understanding of the type of characteristic or variable before collecting the data. The observation on qualitative variables may also be assigned numbers. For example ,we can record a person's marital status as 1,2,3 or 4 depending on whether the person is single, married, widowed or devoiced. We can as well record " Yes or No" answers to a question as "0" and "1" (or as 1 and 2 or perhaps as 59 and 60). In this artificial or nominal way, categorical data(qualitative or descriptive) can be made into numerical data and if we thus code the various categories, we refer to the numbers we recorded as nominal data. Nominal data is numerical in name only ,because they do not share any of the properties of the numbers we deal in ordinary arithmetic. For instance if we record marital status as 1,2,3 or 4 as stated above ,we cannot write $4>2$ or $2<$ and we cannot write $3-1=4$, $2+3=4$ or $4/2=2$.

In those situation when we cannot do anything except set up inequalities, we refer to the data as ordinal data. For instance , if one mineral can scratch another, it receives a higher hardness number and on Mohs' scale the numbers from 1 to 10 are assigned respectively to talc, gypsum, calcite, fluorite, apatite, feldspar, quartz, topaz, sapphire and diamond. With these numbers we can write $5>2$ or $6<9$ as apatite is harder than gypsum and feldspar is softer than sapphire, but we cannot write for example $10-9=5-4$, because the difference in hardness between diamond and sapphire is actually much greater than that between apatite and fluorite. It would also be meaningless to say that topaz is twice as hard as fluorite simply because their respective hardness numbers on Mohs' scale are 8 and 4. The greater than symbol (i.e., $>$) in connection with ordinal data may be used to designate "happier than" "preferred to" and so on.

When in addition to setting up inequalities we can also form differences, we refer to the data as Interval data . Suppose we are given the following temperature readings(in degrees Fahrenheit): 58,63,70,95,110,126 and 135. In this case, we can write $100>70$ or $95<135$ which simply means that 110 is warmer than 70 and that 95 is cooler than 135. We can also write for example means that 110 is warmer than 70 and that 95 is cooler than 135. We can also write for example $95-70=135-110$,since equal temperature differences are equal in the sense that the same amount of heat is required to raise the temperature of an object from 70 to 95 or from 110 to 135. On the other hand, it would not mean much if we said that 126 is twice as hot as 63, even though $126/63=2$. To show the reason ,we have only to change to the centigrade scale, where the first temperature become $5/9(126-32)=52$, the second temperature becomes $5/9(63-32)=17$ and the first figure is now more than three times the second. The difficulty arises from the fact that Fahrenheit and Centigrade scales both have artificial origins(zeros) i.e., the number 0 of neither scale is indicative of the absence of whatever quantity we are trying to measure.

When in addition to setting up inequalities and forming differences we can also form quotients(i.e.,) when we can perform all the customary operations of mathematics), we are refer to such data as ratio data . In this sense, ratio data includes all the usual

measurement (or determinations) of length, height, money amounts, weight, volume, area, pressures etc.

The above stated distinction between nominal, ordinal, interval and ratio data is important for the nature of a set of data may suggest the use of particular statistical techniques. A researcher has to be quite alert about this aspect while measuring properties of objects or of abstract concepts.

Concepts, Indicators and Variables

In the process of formulating a research problem, there are two important considerations – use of concepts and measurement of properties. Concepts are highly subjective, as their understanding varies from person to person and therefore, may not be measurable. In a research study, it is important that the concepts used should be operationalised in measurable terms so that the extent of variation in respondents' understanding is reduced, if not eliminated. Techniques to tackle issues like - how to operationalise concepts, how to identify variables and indicators - play an important role in reducing this variability.

Concepts

Concepts are basic elements of scientific method. Concepts emerge from phenomenon. Concepts are more abstract than phenomena. According to Goode and Hatt, a concept is "is like a fact and is an abstraction, not phenomena, and it takes its meanings from the thought framework within which it is placed. Concepts are the foundation of all human communication and thought, besides being basic to scientific method.

Certain concepts are difficult to define, especially when human behavior and attitude are involved. Weight, height, energy are easy to define. They can be pointed out and pictures can be taken of them. However it is difficult to define certain concepts like frustration, happiness, motivation, aggression etc. because they are constructs. They exist in our mental frame work only. They cannot be touched or seen, but can only be felt. They can be called constructs. Constructs represent inferences at a higher level of abstraction from concrete events, and their meanings cannot be easily conveyed, by pointing to specific objects, individuals or events.

Concept and operationalisation

If you are using a concept in your study, you need to consider its operationalisation – that is, how it will be measured. In most cases, to operationalise a concept, you first need to go through the process of identifying indicators – a set of criteria which reflect the concept, which can then be converted into variables. Most of the concepts can easily be converted into indicators and then variables.

For example, we may want to decide if a person is rich or not. Richness is the concept, and one first needs to decide upon the indicators of richness. Monthly Income is one such indicator, and it takes the shape of a variable which can be measured.

Indicators

Each concept has certain indicators. Indicators are characteristics by which concepts can be operationalised and measured. For example, the term asset is a concept. Indicators of this concept are house, car and other investments. Converting the value of each one will give total value of the assets owned by a person. Similarly, the indicators of effectiveness of a health program are patients, change in mortality, change in birth rates etc. There are

appropriate variables to measure these indicators. The choice of indicators for a concept might vary with the researcher but those selected must have a logical link with the concept.

Variables

An image, perception or concept that is capable of measurement is called a variable. A variable is capable of taking on different values. In other words, a concept that can be measured is called a variable. According to Kerlinger, " a variable is a property that takes on different values. Putting it redundantly, a variable is some that varies. A variable is a symbol of which numerals or values are attached. Black and Champion define a variable as "rational units of analysis that can assume any one of a number of designated sets of values. "a concept that can be measured on any one of the four types of measurement scale, which have varying degree of precision in measurement, is called a variable.

There are some who believe that scientific methods are incapable of measuring feelings, preferences, values and sentiments. In our view, most of such attitudes can be measured, though indirectly. These feelings and judgments are based upon observable behavior in real life, though the extent to which the behavior reflect their judgments may vary from person to person.

Types of variables

Variables are numerous types. They can be classified on three criteria – causal relation, design of the study, and unit of measurement.

In studies that attempt to investigate a causal relation, three sets of variables may operate.

Change variable, which are responsible for bringing about change in a phenomenon. They are also called independent variable.

Outcome variable, which are the effects of change variable. They are also called dependent variable.

Extraneous variable, affect cause and effect relationship, but are not considered or measured for the time being.

From the viewpoint of the study design, variables may be active variable and attribute variable. Active variable is that variable that can be manipulated or changed or controlled.

Attribute variable is that variable that cannot be manipulated, changed or controlled, and that reflect the characteristics of the study population. For example – age, gender, education and income.

From the viewpoint of unit of measurement variables may be categorical variables, constant variable, continuous variables etc.

Categorical variables are measured on nominal or ordinal measurement scales. For example – rich/poor, good/bad etc.

Constant variables are those which have only one value or category. For example – taxi, water etc. Continuous variables have continuity in their measurement. For example – age, income attitude etc.

Classifications of measurement scales

From what has been stated above, we can write that scales of measurement can be considered in terms of their mathematical properties. The most widely used classification of measurement scales are : (a) nominal scale; (b) ordinal scale; (c) interval scale; and (d) ratio scale.

Nominal scale:

Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of the assignment of numbers of basketball players in order to identify them. Such numbers cannot be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved. For example, one cannot usefully average the numbers on the back of a group of football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one group with the numbers assigned to another. The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measures of dispersion for nominal scales. Chi-square test is the most common test of statistical significance that can be utilized, and for the measures of correlation, the contingency coefficient can be worked out.

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

Ordinal scale

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

Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real differences between ranks 1 and 2 may be more or less than the differences between ranks 5 and 6. Since the numbers of the scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measure of statistical significance are restricted to the non-parametric methods.

Interval scale

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

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

Ratio scale

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

Ratio scale represents the actual amount of variables. Measures of physical dimension such as weight, height, distance, etc. are examples. Generally, all statistical techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated. Thus, proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly, if the nature of the variables, permits, the researcher in physical science have the advantage to describe variables in ratio scale from but the behavioural sciences are generally limited to describe variables in interval scale from, a less precise type of measurement.

Goodness of measurement scales

A measurement scale has to have certain desirable qualities to judge their goodness in measuring the characteristic under study. These qualities are described below:

(i) Validity

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

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

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

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

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

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

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

Reliability

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

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

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

Practicality

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

The measuring instrument, in order to be interpretable, must be supplemented by (a) detailed instructions for administering the test; (b) scoring keys; (c) evidence about the reliability and (guides for using the test and interpreting results).

Accuracy

The characteristic of accuracy of a measurement scale means it should be a true representative of the observation of underlying characteristic. For example, measuring with an 'inch' scale will provide accurate value only upto one-eighth of an inch, while measuring with 'cm' scale will provide more accurate value.

Sources of error in measurement

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

Respondent:

At time the respondent may be reluctant to express strong negative feelings or it is just possible that he may very little knowledge but may not admit his ignorance. All this reluctance is likely to result in an interview of 'guesses'. Transient factors like fatigue, boredom, anxiety, etc. may limit the ability of the respondent to respond accurately and fully.

Situation:

Situational factors may also come in the way of correct measurement. Any condition which places a strain on interview can have serious effects on the interviewer-respondent rapport. For instance, if someone else is present, he can distort responses by joining in or merely by being present. If the respondent feels that anonymity is not assured, he may be reluctant to express certain feelings.

Measurer:

The interviewer can distort responses by rewording or reordering questions. His behavior, style and looks may encourage or discourage certain replies from respondents. Careless mechanical processing may distort the findings. Errors may also creep in because of incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data analysis stage.

Instrument:

Error may arise because of the defective measuring instrument. The use of complex words, beyond the comprehension of the respondent, ambiguous meanings, poor printing, inadequate space for replies, response choice omissions, etc. are a few things that make the measuring instrument defective and may result in measurement errors. Another type of instrument deficiency is the poor sampling of the universe of items of concern.

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

Techniques of developing measurement tools

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

Concept development, Specification of concept dimensions, Selection of indicators , and formulation of index.

Concept development

- The first and foremost step is that of concept development which means that the researcher should arrive at an understanding of the major concepts pertaining to his study. This step concept developmental concepts are often already established.

Specification of concept dimensions

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

Selection of indicators

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

Formulation of Index

The last step is that of combining the various indicators into an index, i.e., formation of an index. When we have several dimensions of a concept or different measurements of a dimension, we may need to combine them into a single index. One simple way

For getting an overall index is to provide scale values to the responses and then sum up the corresponding scores. Such an overall index would provide a better measurement tool than a single indicator because of the fact that an individual indicator has only a probability relation to what we really want to know. This way we must obtain an overall index for the various concepts concerning the research study.

Review Questions

1. What is the role of data in research
2. What is meant by measurement
3. How data can be generated
4. State the necessity of measurement
5. Define measurement
6. What are concepts
7. Define indicator
8. How does concept differ from indicator
9. What are variables
10. Explain operationalisation of concept
11. Give a classification of variable
12. What are the different types of measurement scales

13. Explain nominal scale
14. What is a ordinal scale
15. compare interval scale with ratio scale
16. What are the desirable qualities of a scale what are the sources of error in measurement
17. Explain the concept of development measurement tool

UNIT 16 SCALING

Introduction

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

Measurement describes the procedures of assigning numbers to various degrees of opinion, attitude and other concepts. This can be done in two ways viz.,(i)making a judgment about some characteristic of an individual and then placing him directly on a scale that has been defined in terms of that characteristic and (ii) constructing questionnaire in such a way that the score of individual's responses assigns him a place on a measurement scale.

Scaling

Scaling is an extension of the concept of measurement. It can also be said that the result of measurement is scaling . Scaling refers to the process of placing a measured object, characteristic or attitude on the appropriate point on a scale. Scaling is ascertaining the score of characteristics, using measurement, on a continuum, subject to rules. The difference between measurement and scaling is that measurement is the assignment of number s to objects or respondents and scaling is the process of placing them on a continuum with respect to the number scored by them.

Alter L Edwards has defined scaling as referring to measurement. In his words, scaling is a "procedure for assignment of numbers or other symbols to a property of objects in order to impart some of the characteristics of numbers to the properties in question."

Hence the term 'scaling' is applied to the procedures for attempting to determine quantitative measures of subjective and abstract concepts.

Scale

It may be stated here that a scale is a continuum, consisting of the highest point (in terms of some characteristic e.g., preference, favourableness.etc.) and the lowest point along with several intermediate points between these two extreme points. These scale-point positions are so related to each other that when the first point happens to be the highest point the second point indicates a higher degree in terms of a given characteristic as compared to the third points and the third point indicates a higher degree as compared to the fourth and so on. Numbers for meaning the distinction of degree in the attitudes/opinion are thus assigned to individuals corresponding to their scale-positions. Scales may take physical form or hypothetical form.

Determination of Scale

Scales appropriate for measuring a phenomenon or attitude can be determined and classified on the basis of following six factors:

- Objective of study
- Nature of response
- Degree of subjectivity involved
- Number of dimensions against which responses are noted
- Scale properties
- Scale construction technique

Essential of Scaling

- Continuum : If in a social phenomenon , the various factors are logically related and are in the form of a continued measurement they can be called scalable or measurable. If they are stray facts, lying here and there, it would not be possible to measure them. They are logically unrelated and so they cannot be included in the same scale. Investigator should have full knowledge of the universe from which samples have been drawn and also of the subject under study. Then only scaling becomes appropriate appropriate and accurate. Continuum means judging the scalability of the phenomenon under study.
- Reliability : The scale under for measuring the social phenomenon must be reliable. Reliability means that it should consistently produce the same result when applied to the same design. That is , reliability implies uniformity.
- Validity : A scale is called valid if it measures correctly what is expected to measure. Validity implies correct measurement.
- Weighting items : The scale becomes more valid if proper weights are given to various points included in the scale. The attributes involved in a study are not of equal importance and so they are to be provided with proper weights.
- Equality of units : Construction of the scale involves, ordering or ranking of units of a continuum. If the units are unequal, the construction of scale becomes difficult. Equality of units is a desirable characteristic but not essential for sound scientific procedure. In order to make the units equal, sometimes subtraction or addition can be made.

Difficulties in scaling

- Complexity of social phenomena: Social phenomena are a very complex affair. It is not easy define them quantitatively . Because of this complexity, it is not easy to draw scales for measuring them.
- Change in human behavior: Different types of changes are taking place in the circumstance of every one's life. So people have to adjust their life to suit these changes. Therefore they have to change their behaviour pattern. Because of these changes in human behaviour of social phenomena also face problems.
- Heterogeneity in the nature of human beings : people belong to groups of different culture and different social customs. Therefore it is difficult to have a common scale applicable to all people. No universally accepted scale can

be formulated for measuring the human behaviour pattern. Because of these changes in human behaviour of social groups.

- **Qualitative nature of the social phenomena :** Social phenomena are qualitative in nature. Qualitative characters cannot be measured. So they are to be converted to quantitative data, which is not an easy task.

Sources of errors in scaling

The various sources of errors in measurement are :

- **Response error :** Sometimes the respondent may be reluctant to express strong negative feelings. Sometimes the respondent may have very little knowledge about the various aspects but he will not admit this ignorance. The figure, boredom, anxiety etc of the respondent may limit the ability of the respondent to express his response accurately. In these situations the researcher has to make guesses in interviews. These guesses may sometimes be erroneous, thus resulting into wrong measurement.
- **Situational errors :** Situational factors also cause error in measurement. If something happens at the time of interview which place a strain on interview, that will have serious effects on the response from respondent. For example, if somebody is present at the time of interview, the respondent may not reveal facts correctly.
- **Errors due to faulty instrument :** Errors may occur due to defective measuring instrument. Use of complex words, ambiguous meanings, poor printing, inadequate space for replies etc may result in measurement errors.

Scale classification base

The number assigning procedures or the scaling procedures may be broadly classified on one or more of the following bases. (a) subject orientation; (b) response form; (c) degree of subjectivity ;(d) scale properties;(e) number of dimension and (f) scale construction techniques. We take up each of these separately.

(a) Subject orientation:

Under it a scale may be designed to measure characteristics of the respondent who completes it or to judge the stimulus object which is presented to the respondent. In respect of the former, we presume that the stimuli presented are sufficiently homogeneous so that the between-stimuli variation is small as compared to the variation among respondents. In the latter approach , we ask the respondent to judge some specific object in terms of one or more dimension and we presume that the between-respondent variation will be small as compared to the variation among the different stimuli presented to respondents for judging.

(b) Response form:

Under : Under this we may classify the scales as categorical and comparative Categorical scales are also known as rating scales. These scales are used when a respondent scores some object without direct reference to other objects. Under comparative scales, which are also known as ranking scales, the respondent is asked to compare two or more objects. In this sense the respondent may state that one object is superior to the other or that three models of pen rank in order 1,2 and 3 . The essence of ranking is, in fact, a relative comparison of a certain property of two or more objects.

(c) Degree of subjectivity:

With this basis the scale data may be based on whether we measure objective personal preferences or simply make non-preferences judgments. In the former case, the

respondent is asked to choose which person he favours or which solution he would like to see employed, whereas in the latter case he is simply asked to judge which person is more effective in some aspect or which solution will take fewer resources without reflecting any personal preferences.

(d) Scale properties:

Considering scale properties, one may classify the scale as nominal, ordinal, interval and ratio scales. Nominal scales merely classify without indicating order, distance or unique origin. Ordinal scales indicate magnitude relationships of 'more than' or 'less than' but indicate no distance or unique origin. Interval scales have both order and distance values, but no unique origin. Ratio scales possess all these features.

(e) Number of dimensions:

In respect of this basis, scales can be classified as 'unidimensional' and 'multidimensional' scales. Under the former we measure only one attribute of the respondent or object, whereas multidimensional scaling recognizes that an object might be described better by using the concept of an attribute space of 'n' dimensions, rather than a single-dimension continuum.

Scale construction approaches

Following are the five main techniques by which scales can be developed.

(i) Arbitrary approach:

It is an approach where scale is developed on adhoc basis. This is the most widely used approach. It is presumed that such scales measure the concepts for which they have been designed, although there is little evidence to support such an assumption.

(ii) Consensus approach :

here a panel of judges evaluate the items chosen for inclusion in the instrument in terms of whether they are relevant to the topic area and unambiguous in implication.

(iii) Item analysis approach:

Under it a number of individuals items are developed in to a test which is given to a group of respondents. After administering the test, the total scores are calculated for every one. Individual items are then analyzed to determine which items discriminate between persons or objects with high total scores and those with low scores.

(iv) Cumulative scales

Are chosen on the basis of their conforming to some ranking of items with ascending and descending discriminating power. For instance, in such a scale the endorsement of all items indicating a less extreme position is seen.

(v) Factor scales

may be constructed on the basis of inter correlations of items which indicate that a common factor accounts for the relationship between items. This relationship is typically measured through factor analysis method.

Non-comparative scaling Techniques

Most common non-comparative scaling techniques are

Continuous rating or graphic rating

Itemized rating

Simple/Multiple category scale

Verbal/ Frequency scale

These techniques are discussed in brief as follows:

Continuous Rating or Graphic Rating

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

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

Itemized Rating

A scale having numbers or brief descriptions of each category is provided. Categories are ordered in terms of scale positions. The respondents select one of the categories that best describes the stimulus object. Commonly used itemized rating scales are Likert Scale, Thurstone Scale, Guttman Scale , Semantic Differential Scale, Single or Multiple Category Scale etc.

Likert Scale

Likert scales or summated scales are developed by utilizing the item analysis approach wherein a particular item is evaluated on the basis of how well it discriminates between those persons whose total score is high and those whose score is low. Those items or statements that best meet this sort of discrimination test are included in the final instrument.

Thus, summated scales consist of a number of statements which express either a favourable or unfavorable attitude towards the given object to which the respondent is asked to react. The respondent indicates his agreement or disagreement with each statement in the instrument. Each response is given a numerical score, indicating its favorableness' towards an issue.

Most frequently used summated scales in the study of social attitudes follow the pattern devised by Likert. For this reason they are often referred to as Likert-type scales. In a Likert scale, the respondent is asked to respond to each of the statements in terms of several degrees, usually five degrees(but at times 3 or 7 may also be used) of agreement or disagreement. For example, when asked to express opinion whether one considers his job quite pleasant, the respondent may respond in any one of the following ways: (i) strongly agree,(ii) agree,(iii) undecided,(iv) disagree,(v) strongly disagree.

We find that these five points constitute the scale. At one extreme of the scale there is strong agreement with the given statement and at the other, strong disagreement , and between them lie intermediate points. We may illustrate this as under:

Each point on the scale carries a score. Response indicating the least favourable degree of job satisfaction is given the least score(say 1) and the most favourable is given the highest score(say 5). These score-values are normally not printed on the instrument but are shown here just to indicate the scoring pattern. The Likert scaling technique ,thus, assigns a scale value to each of the five responses. The same thing is done in respect of each and every statement in the instrument. This way the instrument yields at total score for each respondent, Which would then measure the respondent's favorableness toward the given

point of view. If the instrument consists of, say 30 statements, the following score values would be revealing.

$30 \times 5 = 150$ Most favourable response possible

$30 \times 3 = 90$ A neutral attitude.

$30 \times 1 = 30$ Most unfavourable attitude.

The scores for any individual would fall between 30 and 150. If the score happens to be above 90, it shows favorable opinion to the given point of view. A score of below 90 would mean unfavourable opinion and a score of exactly 90 would be suggestive of a neutral attitude.

Procedures:

The procedure for developing a Likert-type is as follows:

(i) As a first step, the researcher collects a large number of statements which are relevant to the attitude being studied and each of the statements expresses definite favourableness or unfavourable and statements is approximately equal.

(ii) After the statements have been gathered, a trial test should be administered to a number of subjects. In other words, a small group of people, from those who are going to be studied finally, are asked to indicate their response to each statement by checking one of the categories of agreement or disagreement using a five point scale as stated above.

(iii) The response to various statements are scored in such a way that a response indicative of the most favourable attitude is given the highest score of 5 and that with the most unfavourable attitude is given the lowest score, say of 1.

(iv) Then the total score of each respondent is obtained by adding his scores that he received for separate statements.

(v) The next step is to array these total scores and find out those statements which have a high discriminatory power. For this purpose, the researcher may select some part of the highest and the lowest total scores, say the top 25 per cent and the bottom 25 per cent. These two extreme groups are interpreted to represent the most favourable and the least favourable attitudes and are used as criterion groups by which to evaluate individual statements. This way we determine which statements consistently correlate with low favourability and which with high favourability.

(vi) Only those statements that correlate with the total test should be retained in the final instrument and all others must be discarded from it.

Advantages:

The Likert-type scale has several advantages. Mention may be made of the important ones.

(a) It is relatively easy to construct the Likert-type scale in comparison to Thurstone-type scale can be performed without a panel of judges.

(b) Likert-type scale is considered more reliable because under it respondents answer each statement included in the instrument. As such it also provides more information and data than does the Thurstone-type scale.

(c) Each statement, included in the Likert-type scale, is given an empirical test for discriminating ability and such, unlike Thurstone-type scale, the Likert-type scale permits the use of statements that are not manifestly related (to have a direct relationship) to the attitude being studied.

(d) Liker-type scale can easily be used in respondent-centred and stimulus-centred studies i.e., through it we can study how responses differ between stimuli.

(e) Liker-type scale takes much less time to construct, it is frequently used by the students of opinion research. Moreover ,it has been reported in various research studies that there is high degree of correlation between Liker-type scale and Thurstone-type scale.

Limitations:

There are several limitations of the Liker-type scale as well. One important limitation is that, with this scale, we can simply examine whether respondents are more less favorable to a topic ,but we cannot tell how much more or less they are.

There is no basis for belief that the five position indicated on the scale are equally spaced. The interval between 'strongly agree' and 'agree' , may not be equal to the interval between "agree" and "undecided" . This means that Likert scale does not rise to a stature more than that of an ordinal scale, whereas the designers of Thurstone scale claim the Thurstone scale to be an interval scale.

One further disadvantage is that often the total score of an individual respondent has little clear meaning since a given total score can be secured by a variety of answer patterns. It is unlikely that the respondent can validly react to a short statement on a printed form in the absence of real-life qualifying situations.

Moreover ,there "remains a possibility that people may answer according to what think they should feel rather than how they do feel." This particular weakness of the Liker-type scale is met by using a cumulative scale which we shall take up later in this chapter.

In spite of all the limitations, the Liker-type summated scales are regarded as the most useful in a situation wherein it is possible to compare the respondent's score with a distribution of scores from some well defined group. They are equally useful when we are concerned with a programme of change or improvement in which case we can use the scales to measure attitudes before and after the programme of change or improvement in order to assess whether our efforts have had the desired effects. We can as well correlate scores on the scale to other measures without any concern for the absolute value of what is favourable and what is unfavourable. All this accounts for the popularity of Liker-type scales in social studies relating to measuring attitudes.

Thurstone Scale

To overcome certain problems with the Likert scale, the Thurstone Scale was developed. Thurstone scale considers a 'weight' or 'attitudinal value' for each statement. The weight, which is equivalent to the median value, for each statement is calculated on the basis of rating assigned by a group of judges. Each statement with which respondents express agreement , or to which they respond in the affirmative is given an attitudinal score equivalent to the attitudinal value of the statement.

Procedure in Thurstone Scale

- Construct statements reflecting attitudes towards the issue in question.
- Select a panel of judges who are experts in the field of attitudes being explored.
- Send the statements of respondents to these judges with a request to rate each statement on 11 point scale.
- On the basis of the judges' ratings, calculate the median value of their ratings for each item.

- From the statements, select items that best reflect attitudes towards various aspects of the issue.
- Construct a schedule comprising the selected items.

The Thurstone Scale is also called Equal Appearing Interval Scale. The main advantage of this scale is that, as the importance of each statement is determined by judges, it reflects the absolute rather than relative attitudes of respondents. The scale is thus able to indicate the intensity of people's attitudes and any change in this intensity will be replicated. On the other hand, the scale is difficult to construct, and a major criticism is that judges and respondents may assess the importance of a particular statement differently and, therefore, the respondents' attitudes might not be reflected.

Stapel Scale

This scale was developed by John Stapel. This is a unipolar rating scale with usually 10 categories numbered from -5 to +5. This scale does not have zero or the neutral point. Respondents rate how each term describes the object by selecting the appropriate number. Positive number means the term describes the object accurately, while negative number implies that the term describes the object inaccurately. +5 means the highest degree of accuracy while -5 means the highest degree of inaccuracy.

For example, consider the phrases (i) Tasty food, (ii) Fast Service, and (iii) Good ambience for a restaurant. A respondent is asked to rate accurately these terms or phrases to describe a specified restaurant.

Tasty Food	Fast Service	Good Ambience
+5	+5	+5
+4	+4	+4
+3	+3	+3
+2	+2	+2
+1	+1	+1
-1	-1	-1
-2	-2	-2
-3	-3	-3
-4	-4	-4
-5	-5	-5

The respondent selects +1 for 'tasty food', -2 for 'fast service' and +3 for 'good ambience'. This means the respondent thinks that the phrase 'tasty food' is accurate with a minimum degree of accuracy; phrase 'fast service' is somewhat inaccurate, while 'good ambience' is sufficiently accurate. So according to the respondent the ambience of the restaurant is good, food taste is okay but service is slow.

This method is applicable when the responses are rated on a single dimension. The method is very economic and data can be collected over telephonic interview also. Some researchers think that the method is confusing and of not much use.

Guttman Scale

The Guttman Scale is one of the most difficult scales to construct and therefore is rarely used. This scale does not have much relevance for beginners or students in research and so is not discussed in this study material. Here statements of respondents are arranged

in a perfect cumulative order, which is rather cumbersome. This scaling technique is also called cumulative rating scale.

Semantic Differential Scale

Semantic differential scale or the S.D. Scale developed by Charles E. Osgood, G.J. Suci and P.H Tannenbaum (1975), is an attempt to measure the psychological meaning of an object to an individual. The scale is based on the presumption that an object can have different dimensions of connotative meaning which can be located in multidimensional property space, or what can be called the semantic space in the context of S.D. scale. This scaling consists of a set of bipolar rating scales, usually of 7 points, by which one or more respondents rate one or more concepts on each scale item. For instance, the S.D. scale items for analyzing candidates for leadership position may be shown as under:

Candidates for leadership position (along with the concept-the 'ideal' candidate) may be compared and we may score them from +3 to -3 on the basis of the above stated scales. (The letters, E,P,A showing the relevant factors viz., evaluation, potency and activity, contributed most to meaningful judgments by respondents.

Procedure :

Various steps involved in developing S.D. scale are as follows:

- First of all the concepts to be studied are selected. The concepts are usually chosen by personal judgment, keeping in view the nature of problem.
- The next step is to select the scales bearing in mind the criterion of factor composition and the criterion of scale's relevance to the concept being judged (it is common practice to use at least three scales for each factor with the help of which an average factor score has to be worked out). One more criterion to be kept in view is that scales should be stable across subject and concepts.
- Then a panel judges are used to rate the various stimuli (or objects) on the various selected scales and the responses of all judges would then be combined to determine the composite scaling.

The Semantic Differential Scale has a number of specific advantages. It is an efficient and easy way to secure attitudes from a large sample. These attitudes may be measured in both direction and intensity. The total set of responses provides a comprehensive picture of the meaning of an object, as well as a measure of the subject doing the rating. It is a standardized technique that is easily repeated, but escapes many of the problems of response distortion found with more direct methods.

Single/Multiple Category Scales

These scales are also known as dichotomous scales. Here we have two or more mutually exclusive responses. For example, 'Yes' and 'No', 'True' and 'False'. The respondent has to choose only one out of the given categories. It would be clear from the following example questions of a questionnaire:

1. Do you play cricket ?	Yes	No
2. What is your marital status?	Unmarried	Married
	Divorce	Widower
3. What is your employment type?	Salaried	business
	Self employed	Student
	Retired	Homemaker

These scales can also be generalized to have more than one response. For example,

1. What do you own?

Car	House	Computer
-----	-------	----------

2. What are you interested in?

Music	Sports
-------	--------

It should be ensured that the provided choice exhaust almost all possible answers of the asked question. The choice 'others' may be given along with other choices. The collected data is on nominal scale .The method is easy and very popular on internet surveys.

Verbal Frequency Scale

This scale is used when the respondent is unable or unwilling to give the exact numbers in the answer.

For example,

How often do you eat out?

1. Frequently 2. Sometimes 3. Rarely 4. Never

Uni Dimensional scale :

In uni dimensional scale , only one attribute of the object is measured. While, a multidimensional scale considers that an object is described with several dimensions. For example, the popularity of a restaurant can be measured by a single measure food taste. It can be defined on multiple dimensions like food taste, service, cleanliness, ambience, etc.

Balanced and unbalanced scales :

In balanced scale , number of favorable and unfavorable categories are equal, while they are unequal in an unbalanced scale. The scale should be balanced in general to collect an unbiased opinion. However , if the distribution of responses is likely to be skewed on one direction , an unbalanced scale having more categories on the opposite direction is used.

Multidimensional scaling

Multidimensional scaling (MDS) is relatively more complicated scaling device, but with this sort of scaling one can scale objects, individuals or both with a minimum of information. Multidimensional scaling (or MDS) can be characterized as a set of procedures for portraying perceptual or affective dimensions of substantive interest. MDS is used when all the variables (whether metric or non -metric) in a study are to be analyzed simultaneously and all such variables happen to be independent

The underlying assumption in MDS is that the respondents perceive a set of objects as being more or less similar to one another on a number of usually uncorrelated

dimensions. Through MDS technique one can represent geometrically the locations and interrelationships among a set of points. In fact, three techniques attempt to locate the points, given the information about a set of inter point distances, in space of one or more dimensions such as to best summarise the information contained in the inter point distances. The distances in the solution space optimally reflect the distances contained in the input data.

Larger number of scale categories will provide better precision in the collected data. However, it would become inconvenient for the respondent. So, the number of scale categories is a trade-off between precision and convenience. It should be chosen with care.

Review Questions and Exercises

1. What do you mean by Scaling?
2. State the necessity of scaling
3. Define scaling technique
4. What are the essential of scaling?
5. What are the difficulties in Scaling?
6. Define a scale
7. What is nominal scale
8. State characteristics of a scale
9. How is scale for a study determined
10. What are the sources of errors in scaling
11. Explain situational errors
12. Give a classification of scales
13. What is multi dimensional scale
14. What are cumulative scales
15. Why Likert scale is considered superior
16. State the basics of Stapel scale
17. What are Rating Scales
18. What are Ranking Scales

UNIT 17 DATA PREPARATION AND PROCESSING

The data, after collection, has to be prepared and processed, for analysis. The collected data is raw and must be converted to the form that is suitable for the required analysis. The results of the analysis are affected a lot by the form of the data. So, proper data preparation is a must to get reliable results.

Data preparation process

Data preparation or processing is a statistical method by which the collected data is so organized that further analysis and interpretation of the data become easy. Data collected through a statistical investigations are complex and unorganized mass of figures. They are not fit for analysis and interpretation. The limitation of human mind to understand such data make it necessary to use statistical techniques to process the mass of figures. They are to be provided with a form or structure. Processing is a necessary to ensure that we have all relevant data for making contemplated comparison and analysis.

According to Mill, "the first task of the statistician is an organisation of figures in such a form that their significance for the purpose in hand may be appreciated, that comparison with masses of similar data may be facilitated and that further analysis may be possible." Therefore, data are to be boiled down to make them fit for digestion. Processing

implies checking, editing, coding, classifying and tabulating the collected data so that they are amenable for analysis.

The plan of data analysis is decided in advance before collecting the data. Data preparation process is guided by that plan of data analysis. Important steps of data preparation process are as follows: Questionnaire Checking, Editing, Coding, Classification, Tabulation, Graphical representation, Data cleaning etc.

We below describe each of the above processes.

Questionnaire Checking

When the data is collected through questionnaires, the first step of data preparation process is to check the questionnaires if they are acceptable or not. This involves the examination of all questionnaires for their completeness and interviewing quality. Usually this step is undertaken at the time of data collection. If questionnaires checking were not done at the time of collection, it should be done later. A questionnaire may not be acceptable if:

- (a) It is incomplete partially or fully.
- (b) It is answered by a person who has inadequate knowledge or does qualify for the participation.
- (iii) It is answered in such a way which gives the impression that the respondent could not understand the questions.
- (iv) If sufficient number of questionnaires are not accepted the researcher may like to more data.

Editing

Editing of data is a process of examining the collected raw data (specially in surveys) to detect errors and omission and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaires and/ or schedules. Editing is done to assure that the data are accurate, consistent with other facts gathered ,uniformly entered , as completed as possible and have been well arranged to facilitate coding and tabulation.

With regard to points or stages at which editing should be done, one can talk of field editing and central editing. Field editing consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated and /or in illegible form at the time of recording the respondents'. This type of editing is necessary in view of the fact that individuals writing styles often can be difficult for others to decipher. This sort of editing should be done as soon as possible after the interview ,preferably on the very day or not the next day or on the next day. While doing field editing , the investigator must restrain himself and must not correct errors of omission by simply guessing what the information would have said if the question had asked.

Central editing should take place when all forms or schedules have been completed and returned to the office. This type of editing implies that all forms or should get a through editing by single editor in a small study and by a team of editors in case of a large inquiry. Editor(s) may correct the obvious errors such as an entry in the wrong place, entry recorded in months when it should have been recorded in weeks, and the like. In case of inappropriate or missing replies, the editor can sometimes determine the proper answer by reviewing the other information in the schedule . At times, the respondent can be contacted for clarification . The editor must strike out the answer if the same is inappropriate and he has no basis for determining the correct answer or the response. In such a case an editing

entry of 'no answer' is called for. All the wrong replies, which are quite obvious, must be dropped from the final results, especially in the context of mail surveys.

Editors must keep in view several points while performing their work(a) They should be familiar with instructions given to the interviewers and coders as well as with the editing instructions supplied to them for the purpose.(b) While crossing out an original entry for one reason or another, they should just draw a single line on it so that the same may remain legible.(c) They must make entries (if any) on the form in some distinctive colour and that too in a standardized form.(d) They should initial all answers which they change or supply.(e) Editor's initial and the date of editing should be placed on each completed form or schedules.

Coding

Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put in to a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristic of exhaustiveness (i.e., there must be a class for every data item) and also that of mutual exclusivity which means that a specific answer can be placed in one and only one cell in a given category set. Another rule to be observed is that of unidimensionality by which it is meant that every class is defined in terms of only one concept.

Coding is necessary for the efficient analysis and thought it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. This makes it possible to precode the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires. But in case of hand coding some standard method may be used. One such standard method is to code in the margin with a coloured pencil. The other method can be to transcribe the data from the questionnaire to a coding sheet. Whatever method is adopted, one should see that coding errors are altogether eliminated or reduced to the minimum level.

Classification

Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

Classification according to attributes:

As stated above, data are classified on the basis of common characteristic which can either be descriptive (such as literacy, sex, honesty ,etc.) or numerical (such as weight, height, income, etc.) Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively ; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain attributes are known as statistics of attributes and their classification is said to be classification according to attributes.

Such classification can be simple classification or manifold classification. In simple classification we consider only one attribute and divide the universe into two classes-one class consisting of items possessing the given attribute and the other class consisting of items which do not possess the given attribute. But in manifold classification we consider two or more attributes simultaneously , and divided that data into a number of classes (total

number of classes of final order is given by 2^n , where $n = \text{number of attributes considered}$ *. When ever data defined in such a manner that there is least possibility of any doubt/ambiguity concerning the said attributes.

Classification according to class-intervals:

Unlike descriptive characteristic, the numerical characteristics refer to quantitative phenomenon which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and are classified on the basis of class intervals. For instance, persons whose incomes, say, are within ₹ 201 to ₹ 400 can form one group, those whose incomes are within ₹ 401 to ₹ 600 can from another group and so on. In this way the entire data may be divided into a number of groups or classes or what are usually called, 'class-intervals'. Each group of class-interval, thus, has an upper limit as well as a lower limit which are known as class limits. The difference between the two class limits is known as class magnitude. We may have classes with equal class magnitudes or with unequal class magnitudes. The number of items which fall in a given class is known as the frequency of the given class. All the classes or groups, with their respective frequencies taken together and put in the form of a table, are described as group frequency distribution or simply frequency distribution. Classification according to class intervals usually involves the following three main problems:

How many classes should be there? What should be their magnitudes? How to choose class limits? Consistent with this, the class limits should be located at multiples of 2, 5, 10, 20, 100 and such other figures. Class may generally be stated in any of the following forms :

Exclusive type class intervals

They are usually stated as follows:

10-20

20-30

30-40

40-50

The above intervals should be read as under:

10 and under 20

20 and under 30

30 and under 40

40 and under 50

Thus, under the exclusive type class intervals, the items whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly 30 would be put in 30-40 class interval and not in 20-30 class interval. In simple words, we can say that under exclusive type class intervals, the upper limit of a class interval is excluded and items with values less than the upper limit (but not less than the lower limit) are put in the given class interval.

Inclusive type class intervals: They are usually stated as follows:

11-20

21-30

31-41

41-50

How to determine the frequency of each class?

This can be done either by tally sheets or by mechanical aids. Under the technique of tally sheet, the class-groups are written on a sheet of paper (commonly known as the tally sheet) and for each item a stroke (usually a small vertical line) is marked against the class group in which it falls. The general practice is that after every four small vertical lines in a class group, the fifth line for the item falling in the same group, is indicated as horizontal line through the said four lines and the resulting flower (III) represents five items. All this facilitates the counting of items in each one of the class groups. An illustrative tally sheet can be shown as under:

Tabulation

When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise and logical order. The procedure is referred to as tabulation. Thus, tabulation is the process of summarizing raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis. In a broader sense, tabulation is an orderly arrangement of data in column and rows.

Tabulation is essential because of the following reasons.

1. It conserves space and reduces explanatory and descriptive statement to a minimum.
2. It facilitates the process of comparison
3. It facilitates the summation of items and the detection of errors and omissions.
4. It provides a basis for various statistical computations.

Tabulation can be done by hand or by mechanical or electronic devices. The choice depends on the size and type of study, cost considerations, time pressures and the availability of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation if other factors are favourable and necessary facilities are available. Hand tabulation is usually preferred in case of small inquiries where the number of questionnaires is small and they are of relatively short length. Hand tabulation may be done using the directly tally, the list and tally or the card sort and count methods. When there are simple codes, it is feasible to tally directly from the questionnaire. Under this method, the codes are written on a sheet of paper, called tally sheet, and for each response a stroke is marked against the code in which it falls. Usually after every four strokes against a particular code, the fifth response is indicated by drawing a diagonal or horizontal line through the strokes. These groups of five are easy to count and the data are sorted against each code conveniently. In the listing method, the code responses may be transcribed onto a large work-sheet, allowing a line for each questionnaire. This way a large number of questionnaires can be listed on one work sheet. Tallies are then made for each question. The card sorting method is the most flexible hand tabulation. In this method the data are recorded on special cards of convenient size and shape with a series of holes. Each hole stands for a code and when cards are stacked, a needle passes through particular hole representing a particular code. These cards are then separated and counted. In this way frequencies of various codes can be found out by the repetition of this technique. We can as well use the mechanical devices or the computer facility for tabulation purpose in case we want quick results, our budget permits their use and we have a large volume of straight forward tabulation involving a number of cross-breaks.

Generally accepted principles of tabulation

Such principles of tabulation, particularly of constructing statistical tables, can be briefly states as follows:

1. Every table should have a clear ,concise and adequate title so as to make the table intelligible without references to the text and this title should always be placed just above the body of the table.
2. Every table should be given a distinct number to facilitate easy reference.
3. The column headings(captions) and the row heading(stubs) of the table should be clear and brief.
4. The units of measurement under each heading or sub-heading must always be indicated.
5. Explanatory footnotes, if any, concerning the table should be placed directly beneath the table, along with the references symbols used in the table.
6. Source of sources from where the data in the table have been obtained must be indicated just below the table.
7. Usually the columns are separated from one another by lines which make the table more readable and attractive. Lines are always drawn at the top and bottom of the table and below the captions.
8. There should be thick lines to separate the data under one class from the data under another class and the lines separating the sub-divisions of the classes should be comparatively thin lines.
9. The columns may be numbered to facilitate reference.
10. Those columns whose data are to be compared should be kept side by side. Similarly percentages and/or averages must also be kept close to the data.
11. It is generally considered better to approximate figures before tabulation as the same would reduce unnecessary details in the table itself.
12. In order to emphasize the relative significance of certain categories, different kinds of type spacing and indentations may be used.
13. It is important that all column figures be properly aligned. Decimal points and (+) or (-) signs should be perfect alignment.
14. Abbreviations should be avoided to the extent possible and ditto marks should not be used in the table.
15. Miscellaneous and exceptional items, if any ,should be usually placed in the last row of the table.
16. Table should be made as logical ,clear, accurate and simple as possible . If the data happen to be very large, they should not be crowded in a single table for that would make the table unwieldy and inconvenient.
17. Total rows should normally be placed in the extreme right column and that of columns should be placed at the bottom.
18. The arrangement of the categories in a table may be chronological ,geographical, alphabetical or according to magnitude to facilitate comparison. Above all, the must suit the needs and requirements of an investigation.

Graphical Representation

Graphs help to understand the data easily. All statistical packages, MS Excel, and Open Office ,org offer a wide range of graphs. In case of qualitative data (or categorized data) ,most common graphs are bar chart and pie charts.

Bar chart:

A bar chart consists of a series of rectangles (or bars) . The height of each rectangle is determined by the frequency of that category. Suppose that the sales of a popular soft

drink in the year 2010-11, in five geographical regions, denoted as A,B,C,D and E, are 15245,23762,9231,14980, and 12387, respectively, measured in 10,000 USD. A bar chart of this data is as below.

Pie chart :

A pie chart is used to emphasize relative proportion or shares of each category. It's a circular chart divided into sectors, illustrating relative frequencies. The relative frequency in each category or sector is proportional to the arc length of that sector or the area of that sector or the central angle of that sector. Suppose in the previous example, if the soft drink has its market only in the five geographical regions, denoted as A,B,C,D, and E. For the year 2010-11 ,the sales data in these regions are 15245,23762,9231,14980, and 12387, respectively measured in 10,000 USD . A total scale of the soft drink is 75605 times 10,000 USD. A pie chart can be plotted to have the idea of the shares of different markets.

In case of quantitative data, one important chart is histogram which is a generalization of bar chart. The data is first summarized in terms of class interval and each bar represent a class interval. The width of the bar is proportional to the width of corresponding class interval. The area of the bar is proportional to the frequency of corresponding class interval. After making the class intervals in a quantitative data set, a pie chart can also be used to read the share of each class interval.

Data Cleaning

This includes checking the data for consistency and treatment for missing value. Preliminary consistency checks are made in editing . Here we check the consistency in an extensive manner.

Consistency checks look for the data which are unknown or not consistent or outlines. Such data may either be discarded or replaced by the mean value. However , the researcher should be careful while doing this. Extreme values or outlines are not always erroneous.

Missing values are the values which are unknown or not answered by the respondent. In place of such missing values, some neutral value may be used. This neutral value may be the mean of available values. The other opinion could be to use the pattern of responses to other questions to calculate a suitable substitute to the missing values.

Data Adjusting

Date adjusting is not always necessary but it may improve the quality of analysis sometimes. This consists of following methods.

Weight Assigning:

Each respondent or case is assigned a weight to reflect its importance relative to other respondents or cases.

Using this method ,the collected sample can be made a stronger representative of a target population on specific characteristics. For example , the case of educated people could be assigned higher weights and of uneducated people could be assigned lower weights in some survey.

Variable Respecification

This involves creating new variables or modifying existing variables. For example , if the usefulness of a certain product is measured on 10 point scale, it may be reduced on a 4 point scale - 'very useful' , 'useful' , 'neutral' , 'not useful'. Ratio of two variables may also be taken to create a new variable.

Method of dummy variables for respecifying categorical variables is also very popular. Dummy variable is a variable which usually takes numerical values based on the corresponding category in the original variable. For example, a group of people is divided into smokers and non-smokers.

Scale Transformation:

Scale transformation is done to ensure the comparability with other scales or to make the data suitable for analysis. Different type of characteristic are measured on different scales. For example, attitude variables are measured on continuous scale, life style are usually measured on a 5 point Liker scale. So the variables which are measured on different scales, cannot be compared. A common transformation is subtracting all the values of a characteristic by corresponding mean and dividing by corresponding standard deviation.

Problems in preparation process

We can take up the following two problems of processing the data for analytical purposes:

The problem concerning "Don't know" (or DK) responses:

While processing the data, the researcher often comes across some responses that are difficult to handle. One category of such responses may be 'Don't Know Responses' or simply Dk response. When the DK response group is small, it is of little significance. But when it is relatively big, it becomes a matter of major concern in which case the question arises: Is the question which elicited DK response useless? The answer depends on two points viz. the respondent actually may not know the answer or the researcher may fail in obtaining the appropriate information. In the first case the concerned question is said to be alright and DK response is taken as legitimate DK response. But in the second case, DK response is more likely to be a failure of the questioning process.

How DK responses are to be dealt with by researchers? The best way is to design better type of questions. Good rapport of interviewers with respondents will result in minimizing DK responses. But about the DK responses that have already taken place? One way to tackle this issue is to estimate the allocation of DK answers from other data in the questionnaire. The other ways is to keep DK responses as a separate category in tabulation where we can consider it as a separate reply category if DK responses happen to be legitimate, otherwise we would let the reader make his own decision. Yet another way is to assume that DK responses occur more or less randomly and as such we may distribute them among the other answer in the ratio in which the latter have occurred. Similar results will be achieved if all DK replies are excluded from tabulation and that too without inflating the actual number of other responses.

The problem of Use of percentages :

Percentages are often used in data presentation for they simplify numbers, reducing all of them to a 0 to 100 range. Through the use of percentages, the data are reduced in the standard form with base equal to 100 which facilitate relative comparisons.

While using percentages, the following rules should be kept in view by researchers :

1. Two or more percentages must not be averaged unless each is weighted by the group size from which it has been derived.

(2). Use of too large percentages should be avoided, since a large percentage is difficult to understand and tends to confuse, defeating the very purpose for which percentage are used.

- (3). Percentage hide the base from which they have been computed. If this is not kept in view, the real differences may not be correctly read.
- (4). Percentage decrease can never exceed 100 per cent and as such for calculating the percentage of decrease , the higher figure should invariably be taken as the base.
- (5). Percentages should generally be worked out in the direction of the causal -factor in case of two-dimension tables and for this purpose we must select the more significant factor out of the two given factors as the causal factor.

Review Questions

1. why collected data is to be prepared and processed
2. what is meant by data preparation process
3. enlist the activities coming under data preparation
4. state the objective of data processing
5. what is meant by questionnaire checking
6. why data should be edited
7. what do you mean by coding
8. explain classification of data
9. what is meant by attribute
10. give an example of exclusive classification
11. how to determine frequency of each class
12. what is meant by tabulation
13. state principles of tabulation
14. explain graphical representation
15. what is a Pie Chart used for
16. what do you mean by data cleaning
17. what is meant by data adjusting
18. what are the components in data adjustment
19. state problems in data preparation process
20. How the problem of DK responses should be dealt with by a researcher?
21. How will you treat missing data?

UNIT 18 DATA ANALYSIS

Introduction

The purpose of research is to find out something. Analysis of data helps to find out this something. Analysis of data is to be made with references to the problem at hand.

By analysis we mean the computation of certain measures or indices along with searching for patterns of relationship that exist among the data groups. In case of survey and experimental data, analysis involves estimating the values of unknown parameters of the population and testing of hypothesis for drawing inference

As stated earlier, by analysis we mean the computation of certain indices or measures along with searching of patterns of relationship that exist among the data groups. Analysis ,particularly in case of survey or experimental data, involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences.

Approaches to Data Analysis

Coded data can be analysed manually or with the help of a computer. If the number of respondents is reasonably small, there are not many variables to analyze, and you are neither familiar with a relevant computer program, nor wish to learn one, you can manually analyse the data. However, manual analysis is useful only for calculating frequencies and for simple cross tabulations. In addition, if you want to carry out statistical tests, they have to be calculated manually. However, the use of statistics depends upon your expertise and desire or need to communicate the findings in a certain way.

Be aware that manual analysis is extremely time consuming. The easiest way to analyse data manually is to code it directly onto large graph paper in columns in the same way as you would enter it into a computer. On the graph paper you do not need to worry about the column number detailed headings can be used or question numbers can be written on each column to code information about the question.

Types of analysis

Analysis of data can be categorized as Descriptive analysis and Inferential analysis. Descriptive analysis deals with the study of variables. When the study is of one variable, it is called unidimensional analysis and when the study is about two variables, it is called bivariate analysis. If it is about more than two variables, it is called multivariate analysis.

In descriptive analysis, we work out various measures which will show size and shape of a distribution and also measures the degree of the relationship between two or more variables. Descriptive analysis like Correlation analysis, Causal analysis, Regression analysis, etc. deal with relationship between two or more variables for determining the amount of correlation or association, between two or more variables.

Causal analysis is concerned with the study of how one or more variables affect changes in another variable. It assesses functional relationship existing between the variables. This analysis can be termed as correlation analysis. It is important in experimental researchers. But in social and business researchers correlation analysis is relatively more important.

Multivariate analysis deals with simultaneously more than two variables. Multiple Regression Analysis, Multi variant Analysis of variance, Canonical analysis etc. are various multivariate analysis techniques. In multiple regression analysis, the researcher has one dependent variable and two or more independent variables. This analysis facilitates prediction of the value of dependent variable based on its relation with all the concerned independent variables.

When the researcher has a single dependent variable that cannot be measured but can be classified into two or more groups on the basis of some attribute, we have multiple discriminant analysis. With the help of this analysis we can predict an entity's possibility of belonging to a particular group based on several predictor variables. In Multivariate

analysis of variance , we find out the ratio between within – group variance and among-group variance.

Canonical analysis facilitates simultaneous prediction of a set of dependent variables. This analysis can be used in the case of both measurable and non measurable variables.

Correlation analysis and Causal analysis are closely related. Correlation analysis studies the joint variation of two or more variables for determining the amount of correlation between two or more variables. Causal analysis is concerned with the study of how one or more variables affect changes in another variable. It is thus a study of functional relationships existing between two or more variables. This analysis can be termed as regression analysis. Causal analysis is considered relatively more important in experimental researches, whereas in most social and business researches our interest lies in understanding and controlling relationships between variables then with determining causes *per se* and as such we consider correlation analysis as relatively more important.

Multiple regression analysis is adopted when the researcher has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about the dependent variable based on its covariance with all the concerned independent variables.

Multiple discriminant analysis is appropriate when the researcher has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute. The object of this analysis happens to be to predict an entity's possibility of belonging to a particular group based on several predictor variables.

Multivariate analysis variance (or multi- ANOVA): This analysis is an extension of two-way ANOVA , wherein the ratio of among group variance to within group variance is worked out on a set of variables.

Canonical analysis : This analysis can be used in case of both measurable and non-measurable variables for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.

Inferential analysis is concerned with the various tests of significance for testing hypothesis in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

Statistics in research

The role of statistics in research is to function as a tool in designing research, Analyzing its data and drawing conclusion there from. Most research studies result in a large volume of raw data which must be suitably reduced so that the same can be read easily and for further analysis. Clearly the science of statistics cannot be ignored by any

research worker, even though he may not have occasion to use statistical methods in all their details and ramifications. Classification and tabulation ,as stated earlier ,achieve this objective to some extent, but we have to go a step further and develop certain indices or measures to summarize the collected/classified data. Only after this we can adopt the process of generalization from small groups(I.e., samples) to population. If fact ,there are two major areas of statistics viz., descriptive statistics and inferential statistics. Descriptive statistics concern the development of certain indices from the raw data, whereas inferential statistics concern with the process of generalization ,Inferential statistics are also known as sampling statistics and are mainly concerned with two major type of problems: (i) the estimation of population parameters, and (ii) the testing of statistical hypotheses.

Important statistical measures that are used to summarize the survey or research data are:

(1)measures of central tendency or statistical averages;(2) measures of dispersion'(3)measuresof asymmetry (skewness); ,(4)measures of relationship; and(5) other measures.

Among the measures of central tendency ,the three most important ones are the arithmetic average or mean, median and mode. Geometric man and harmonic mean are also sometimes used.

From among the measures of dispersion, variance, and square root-the standard deviation are the most often used measures. Other measures such as mean deviation, range ,etc. are also used. For comparison purpose ,we use mostly the coefficient of standard deviation or the coefficient of variation.

In respect of the measures of skewness and kurtosis, we mostly use the first measure of skewness, based on quartiles or on the methods of moments, are also used sometimes. Kurtosis is also used to measure the peakedness of the curve of the frequency distribution.

Amongst the measures of relationship ,Karl person's coefficient of correlation is the frequently used measures in case of statistics of variables, whereas Yule's coefficient of association is used in case of statistics of attributes. Multiple ecorrelation coefficient ,partial correlation coefficient ,regression analysis, etc., are other important measures often used by a researcher.

Index numbers, analysis of time series, coefficient of contingency, etc., are other measures that may as well be used by a researcher, depending upon the nature of the problem under study.

We present below a brief outline of some important measures , out of the above listed measures often used in the context of research studies.

Role of computers in Data Analysis

Knowledge of computers and the relevant programs is extremely important and ultimately saves time. Data can be analysed with th help of computers even for qualitative and descriptive information. Computers primarily help to save labour in manual data

analysis and in the calculation of descriptive and inferential statistics. Computers are tools that increase your speed, handle complicated statistical and mathematical procedures, print the report, display the analyzed data and present them graphically. They are designed to perform extremely complicated procedures, which otherwise could be very time consuming, in an extremely short time.

While computers are designed to support computing and analyzing research activities, a good understanding of them is required, to be able to benefit from them. The time spent in learning about computers is ultimately worthwhile, whether you classify yourself as a qualitative or quantitative researcher.

Tools for statistical analysis

There are two major areas of statistics, related with research. Descriptive statistics and other is inferential statistics. Descriptive statistics is related with the development of certain indices from the collected data. Inferential statistics or sampling statistic is related with the process of testing of hypothesis and generalization.

The following are the important statistical tools used for summarizing the research data.

Measures of the central tendency or averages

Measures of the dispersion

Measures of the skewness

Measures of the relationship

Other measures.

Measures of the central tendency

Measures of the central value are a typical value of the entire group of data. It describes the characteristics of entire mass of data. It is the average value of the entire data. In other words it is the single value that describes the characteristics of the entire group. There are five types of averages namely mean, median, mode, geometric mean and harmonic mean.

Mean

Mean is known as arithmetic mean. It is the most common measures of central tendency. Its value is obtained by adding all items and by dividing this by the total number of items.

$$\text{Mean } \bar{x} = \frac{\sum X}{N}$$

In short cut method , the concept of assumed mean is used to find out mean.

Mean is the simplest measurement of tendency and it is widely used. It is relatively stable measure of tendency. It is also relatively reliable in the sense that it does vary too much as the repeated samples are from one and the same population. It has limitations also. It is unduly affected by extreme items. That is very large and small items.

Median

Median is the middle value in the distribution when it is arranged in descending or ascending order. It may be defined as the value of that item which divides the series in to equal parts, one half contains values greater than it and the other half contains value less than it,. Therefore the series is to be arranged in ascending or descending order.

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

2

Median is not frequently used in sampling theory.

Mode

It is the most common item of a series. Mode is the positional average and it is not affected by the values of extreme items. Mode is defined as the value of the variable which occurs most frequently in distribution. The chief feature of the mode is that it is the size of that item which has the maximum frequency and is also affected by the frequencies of the neighboring items. It is particularly useful in study of popular sizes.

Geometric Mean

The geometric mean of n positive figures is the n th root of their products. the geometric mean is the obtained by multiplying together all the values of the series and then calculating the root of their product corresponding to the number of items in the group.

$$\text{Geometric Mean} = \sqrt[n]{X_1 \times X_2 \times X_3 \dots X_n}$$

Where n = number of items and X_1, X_2 are various values

It is often used in the preparation of the index numbers or when we deal in ratios. It cannot be computed if any item in the series is negative or zero.

Harmonic Mean

It is special type measure of central tendency used to solve special type of problems . It is the reciprocal of values of various items in the variable.

$$H.M = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \dots}$$

Where X_1, X_2, \dots, X_n refer to the various values in the observations.

It has limited application. It gives largest weight to the smallest item and smallest item to the largest item.

From the above, it is clear that there are large numbers of averages. The selection of the particular statistical average depends upon the nature, type and objectively of the research. All averages are not suitable for all studies.

Measures of dispersion

The measures of central tendency indicate the central tendency of a frequency distribution in the form of an average. These averages tell something about the general level of magnitude of the distribution, but it does not show anything further about the distribution. Especially it fails to give an idea about the dispersion of variables. In order to measure dispersion statistical devise names measures of dispersion is to be calculated.

According to Spiegel, the degree to which numerical data tend to spread about an average value is called the variation or dispersion of data. The variation measures is used to test the representative character of an average. The following are the important measures of dispersion.

Range

It is the simplest measure of dispersion. Its measure depends upon the extreme items and not on all items. It is defined as the differences between the highest value and the lowest values. In a series, it gives an idea about the variability very quickly. The limitation is that the range is affected greatly by fluctuation of sampling. It is not considered as an appropriate measure in research studies.

Mean deviation

The mean deviation is a measure of dispersion based on all items of in a distribution. It is the arithmetic mean of the deviation of series computed from any measure of central tendency. It is the average of the absolute differences of values of items from an average. It is worked out as under

$$\text{Mean deviation} = \frac{\sum D}{N}$$

Where $\sum D$ = sum of deviations and N = number of items.

Mean deviation is calculated by any measure of central tendency. It is an absolute figure. To compare variation among different series which are expressed in the same or different units a relative mean deviation is required. The relative mean deviation is known as coefficient of mean deviation. The coefficient of mean deviation can be obtained by dividing mean by the average used for calculating mean deviation.

Coefficient of mean deviation

$$= \frac{\text{Mean Deviation}}{\text{Mean or Median or Mode}}$$

Mean or Median or Mode

Standard Deviation

Karl Pearson introduced the concept of standard deviation in 1893. It is the most important measure of dispersion and it is widely used in many statistical formulate. It is commonly denoted by the symbol (sigma). It is defined as the positive square root of the arithmetic mean of the squares of the deviation of the given observation from their deviation.

Standard deviation in the case of raw data = $\sqrt{\frac{\sum x^2}{N}}$

Where $\sum x^2$ = sum of squared deviations and N = number of observations

Standard deviation when deviation taken from assumed mean = $\sqrt{\frac{\sum d^2}{N} - (\frac{\sum d}{N})^2}$

Where $\sum d$ = deviation from assumed mean, and N = number of items.

Coefficient of standard deviation

If we divide the standard deviation by the arithmetic average of the series we will get coefficient of standard deviation. It is relative measures which is used to compare it with similar measure of other series.

Coefficient of variation

If we multiply the coefficient of standard deviation with 100, we will get coefficient of variation. The square of the standard deviation is known as variance. It is considered as the best appropriate method of comparing the variability of two or more distribution.

The standard deviation along with several other related measure like variance, coefficient variation etc are used mostly in research studies. It is possible for further algebraic treatment. It is less effected by the fluctuations of sampling. It is the basis for measuring the coefficient of correlation, sampling and statistical inferences. The main limitation of this methods is that, it is an absolute measure of variability , so it cannot be used for the purpose of comparison.

Measures of Skewness

The measures of dispersion give an idea about the spread of the observation about the central tendency of items. The measures of dispersion give an idea about the spread of the observation about the central tendency do not indicate whether the distribution is symmetric or not. We know that frequency distribution differ widely in their nature and composition. Even then they may have central tendencies and the dispersions same.

When the items in series are dispersed in even fashion, the frequency curve representing the distribution will be symmetrical. If the distribution of the items in a series happen to be symmetrical , curve of the distribution will be normal. It is known ass normal curve. In a symmetrical distribution the items show a perfect balance on either side of the central tendency. In this case the mean, median and mode will be the same. There is no skewness. If the distribution is asymmetrical , it shows that it has skewness. So, skewness is a measure of asymmetry. It shows the manner in which the items are clustered around the average. The measure of skewness helps us to determine the nature and extent of concentration of the observations of the observations towards higher or lower values of the variables. It also helps to measure direction. The following formula can be used to find measure of skewness.

Karl Person coefficient of skewness = $\frac{3(\text{Mean} - \text{Median})}{\sigma}$

Where σ = standard deviation.

Correlation

Measures of central tendency, measures of dispersion, skewness etc are used in the context of univariate population. If we have bivariate population it is used for finding out

the relationship and cause and effect relationship between variables. Correlation technique is used to know the correlation between the two or more variables. Regression techniques are used to know the cause and effect relationship between the two variables.

If there is multivariate population coefficient of partial correlation, coefficient of multiple variation, multiple regression etc are used for study.

Scatter diagram method

It is the simplest device for ascertaining the relationship between two variables. Under this method, the values are plotted simply in a graph paper in dots. By looking the scatter of dots, we can form an idea as to whether the variables are related or not. The greater the scatter of the plotted points, lesser is the relationship between the variables. The more closely points come to a straight line, and then it indicates that there is high degree of relationship. If all the points lie on a straight line from the lower left hand corner to the upper right hand corner, the correlation is said to be perfectly positive correlation. ($r = +1$).

Graphic method

Under this method the individual values of the two variables are plotted on a graph paper. The points are then joined together. It will provide two curves. That is one for 'x' variable and other for 'y' variable. If both curves are running parallel, we can conclude that there is perfect positive correlation. If they are running in opposite directions, we can conclude that there is a perfect negative correlation. This method does not give the idea about the degree of correlation between them. Hence this method is treated as rough method for studying correlation.

Coefficient of correlation

Karl Pearson's coefficient correlation is the most popular method used to find out the coefficient of correlation. It is also known as rank correlations. Under short cut method the coefficient of correlation can be found out by using the following formula where the assumed mean is taken.

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}}$$

Where x, y = deviations from actual means.

When deviations are taken from assumed means

$$r = \frac{\sum d_x d_y - \sum d_x \times d_y}{\sqrt{N \sum d_x^2 - (\sum d_x)^2} \cdot \sqrt{N \sum d_y^2 - (\sum d_y)^2}}$$

Where d_x, d_y = deviations from assumed means

The value of 'r' lies between ± 1 . Positive values of 'r' indicate positive correlation between two variables and negative values 'r' indicates negative correlation. When 'r' indicates 0 there is no relationship between the two variables. The method gives a numerical measure of correlation. Great care must be taken while interpreting value of this because it is very often misinterpreted.

Regression Analysis

Correlation analysis helps to determine the relationship between two variables. Regression technique is used to estimate or predict the value of one variable (dependent variable) with the help of another variable (independent variable). Regression is the measure which indicates the average relationship between two or more variables in terms of original units of data. The line describing the average relationship between two variables is

known as line of regression or estimating lie or regression line. The direction of the line indicates the nature of correlation. If the lines are sloping upwards from left to right, it is positive and if they are sloping downwards it is negative. If the lines coincide, there is perfect correlation. This technique is useful for prediction in business.

Regression equations

As there are two regression lines, there are two regression equations. For the two variables X and Y there are two regression equations namely regression equation of X on Y and regression equations Y and X. Where is +_ correlation between variables, the regression lines will coincide. That means it is one line. If regression lines are farther, the lesser is the degree of correlation and nearer the regression lines, correlation is greater.

Regression Equation of Y on X :

$$Y = a + bX \quad Y = \text{dependent variable and } X = \text{independent variable.}$$

To calculate 'a' and 'b' the following normal equations is to be used.

$$\sum Y = Na + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2$$

Regression equation X on Y :

$$X = a + bY \quad X = \text{dependent variable and } Y = \text{independent variable.}$$

$$\sum X = Na + b\sum Y$$

$$\sum XY = a\sum Y + b\sum Y^2$$

Multiple Correlations and Multiple Regressions

When there are more than two independent variables, multiple correlations technique is used to describe relationship. The coefficient of multiple correlations serves as a measure of the degree of correlations between one variable taken as the dependent variable and a group of other variables taken as independent variables. It has many limitations. It is based on the assumption that the relationship between the variables is linear and the effects of independent variable on the dependent variables are separate, additive and distinct.

Partial Correlation

It is the correlation between one dependent variable with other independent variable putting all other variables constant. It helps to measure the relationship between two variables in such a way that the effects of other related variables are eliminated. That means the relationship between a dependent variable and a particular independent variable is studied by putting the other variables as constant.

Other tools of Analysis

Index numbers

When series of data are expressed in different units, we cannot use averages to compare them. In such cases we want to dependent certain other measures through which the figures are reduced to a common base. One such method is to convert series into a series of index numbers. An index number may be described as a specialized average designed to measure the change in a group of related variables over a period of time. So these are devices for measuring the differences in the magnitude of a group of related variables.

Changes in the various economic and business phenomena can be measured economic and compared with the help of index number. It may used to find out the change in price level over a period of time. Price movements can be studied to know its cause or

effect on the economy. Index numbers may measure cost of living index of different classes of people. Index numbers have limitations. It is often difficult to take into account of the changes in the quality of products. Chances of error may remain at one point or the other while constructing an index number.

Time Series Analysis

It is an arrangement of statistical data in accordance with the time of occurrence or in a chronological order. The numerical data which are collected from different points of time is known as time series. This technique has an important role in economics, statistics and commerce. E.g. daily sales in a shop, population growth of ten years etc.

There are four types of variations which are the components of the time series.

Secular trend or long term movement,

Seasonal variation,

Cyclical fluctuation, and

Random fluctuation

Secular trend is the long term trend which requires 15 to 20 years. There are upward trend and downward trend. Seasonal variation is the variation which occurs with some degree of regularity within a specified period of time of one year or shorter. The seasonal variations occur due to change in climate and custom, habits of the people, etc. Cyclical fluctuations are the fluctuations occurs over a long period of time. It is up and down movement. E.g. depression period, then boom period. Random variation is the irregular variations occurs due to unforeseen and unpredictable events like flood, draught, lock outs etc. Generally four methods are used to measure these trends namely graphic method, semi average method, moving average method and method of least squares.

The analysis of time series is done to understand the dynamic conditions to achieve short term and long term goals of business enterprises. The management can predict the future patterns on the basis of the past figures. The knowledge of the seasonal variations is very much useful to take decisions related with production, inventory, purchase etc. by studying cyclical fluctuation; the policy of the business can be formulated. So it is a power tool in the hands of business researchers and analysts.

Growth Rates

Estimates of trend are required for almost research studies. It has a considerable significance relate to academicians and policy makers. For example trend in deposits. Computation of growth rate is the most important method used for this purpose. The entire series of observation is used for the purpose of finding growth rate. There are two common functional forms used for computations of growth rates with the help of least square method.

Linear growth rate

The following equation is used

$$Y_t = A + Bt$$

Y_t = the value of the variables in the period t for which the growth rate is to be estimated.

That $t = 1, 2, \dots, n$

A and B are constant to be estimated.

This formula assumes that the change in Y is a constant linear rate of B units per year and this rate does not depend on the value of Y in the previous year or any other factor.

Compound growth rate

Linear rate of growth is not applicable when the growth in any period is not independent of the value of the variable in the previous period. In this case the compound growth rate is applied.

Inferential analysis

The purpose of sampling is to estimate some characteristic of the population from which the sample is selected. In almost all fields the statistical enquiry deals with drawing of conclusions related with a population from a sample selected from it. Statistical inference is the process by which we draw a conclusion about some measure of population on the basis of a sample value.

The statistical inference or inferential analysis can be classified into two - Test of hypothesis

and Estimation of population parameters.

Parameters and Statistics

Parameters include the statistical constants of population like mean (variance) coefficient correlation (r) etc. the statistical constants for the sample drawn from the population like mean variance, correlation coefficient (r), proportion (p) etc are known as statistic. In other terms parameters are the functions of the population values and statistic are the functions of sample observations. A sampling investigation produces a result which has to be compared with the one expected with on the basis of population parameter.

Testing of Hypothesis

Hypothesis is considered as the most important instrument in research. Hypothesis is a tentative conclusion logically drawn concerning any parameter of the population. In other words a hypothesis is an assumption or some assumption to be proved or disproved. For example-the mean height of students is 65 inches.

Features of Hypothesis

Hypothesis posses the following features.

- It should be capable of being tested.
- It should be capable of being terms.
- It should state the relationship between variables.
- It must be specific.
- It should be consistent with known facts.

Steps in Testing Hypothesis

- The following steps are to be taken to test the hypothesis.
- State the problem.
- Set up a hypothesis.
- Decide a test statistic.

- Select a level of significance.
- Calculate the value of the test statistic.
- Obtain the table value.
- Make a decision to accept or reject hypothesis.

Testing for Means

It frequently happens that samples are from two different sources and it is required to know whether there is any significant difference in their means. We also want to know when the difference is due to chance, whether the samples belong to the same population. Hence the problem is (i) to test the equality of the Population Mean and Sample Mean(ii) to test the significance of the difference between two independent sample Means.

For testing population mean or sample means, compute Z value, if the sample is large, i.e., 30 or more and compare with Z table value. If the sample size is less than 30, compute t value, and compare the computed value of t with the tabulated value of t for the Degree of freedom. The desired level of significance is usually 1% or 5%, we accept null hypothesis when the calculated value is less than table value and vice versa.

Comparison of Sample Mean with Population Mean

Problem 1 A company market bus tyres. Their lives are normally distributed with a mean of 80,000 kms and standard deviation 6000 kms. A change in the production process is believed to result in a better product. A test samples of 64 new tyres has a mean life of 82,400 kms can you conclude that the new product is significantly better than the current one?

Answer : Null hypothesis (H_0) = there is no significant difference between population mean and sample mean.

At 0.05 level of significance, the critical (table) value of Z is 1.96. The computed value of Z is +1.6 which is less than critical value, and falls in the acceptance region. Thus we accept the null hypothesis. The difference is not significant. Hence we conclude that the new tyres are not better than currently one.

Problem 2 Samples of 121 wooden rods are drawn from a large number of wooden rods, whose length are normally distributed with mean 6ft. the standard deviation is 1.2 ft. If the sample mean is 6.4 can the sample be regarded as a truly random sample.

Answer

Sample is large so use Z test

Apply Z test

Number of sample

Null hypothesis (H_0) = there is no significant differences between population mean and sample mean, = 6 ft.

At 0.05 level of significance the critical value as per table value of $Z = 1.96$. The computed value is + 3.67. It falls in the rejection region. Thus we reject the null hypothesis that there is no difference between population means and sample mean. The difference is significant. Hence the sample cannot be regarded as truly random sample.

Problem 3

A talcum powder manufacturing company was distributing a particular brand of talcum powder through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop were 70 dozens. After the campaign, a sample of 10 shops was taken and the mean sales were found to be 74 dozens with standard deviation of 8. Can you consider the advertisement effective?

Answer : Null hypothesis = There is no significant difference between the mean sales. The advertisement is not effective.

No of sample size is less than 30, so apply t test.

For 9 degree of freedom at 5% level of significance the table value of t is 2.26. The calculated value of t 1.49 is less than the table value. So it falls in acceptance region. Hence there is no significant difference between mean samples, and the increase in sales can be attributed to fluctuations of sampling. We conclude that the advertisement is not effective in increasing sales.

Comparison of Two Sample Means

Problem 4 A and B companies manufacture electronic bulbs and gave the following results. Mean life of the two bulbs, in hours are 1300 and 1248 respectively. Number of bulbs used for test purposes are 100 each. Standard deviations for A is 82 and for B is 93 respectively.

Using the standard error of the difference between means state whether there is any significant difference in the mean life of the two makes.

Answer

Let H_0 = No significant difference between the mean lives of the bulbs.

Test statistic for large samples Z test.

Calculated value is 4.19

At 0.05 level of significance the difference is significant for a two tailed test, because Z Table value is 1.96. Thus we reject the null hypothesis, and conclude that the two population means are not same. In other words there is a significant difference in the mean life of the two makers.

Analysis of Variance and Covariance

Z test t test is used to test the hypothesis of equality between two sample means. F test is used to verify the hypothesis of equality between two variances. The same test is extended towards the test of hypothesis of equality amongst several means.

Analysis of variance (ANOVA) is an important statistical tool related with researches in the field of economics, biology, education, psychology, sociology, business and industries, and several other disciplines. If there are only two groups, the researcher can apply 't' test to determine the significance. Even if there are three groups of sample results, they can be compared pair by pair by employing 't' test. But there are several groups; it is difficult to compare the sample statistics pair by pair. In these cases analysis of variance is only the tool that can be employed to test the null hypothesis that the samples are drawn from the same population. This technique is used when multiple sample cases are involved in our studies. It can be used to draw inferences about whether the samples have been drawn from populations having the same mean.

Assumptions in ANOVA

This technique is based on three important assumptions.

- The distribution of individual values in each group is assumed to have normal shape curve.
- The variance of individuals' values within each group is assumed to be the same.
- That the individuals (samples) being observed have been selected randomly from the population.

There are two techniques of analysis of variance - one way classification and two way classification.

One way classification

In one way classification, data are classified on the basis on a single criterion and the data is presented either horizontally or vertically. In one way ANOVA, following are the important steps.

- Calculate mean of each sample.
- Calculate the grand average.
- Take the difference between the means of the various samples and the grand average.
- Square these deviations and find the sum of squares between the samples (SSC)
- Divide SSC by degree of freedom. We will get mean square between samples (MSC)
- Degree of freedom = number of samples - 1.
- Calculate variance within samples
- Calculate mean value of each sample.
- Take the deviations of various item in a sample from the mean values of the respective samples.
- Square these deviations. Then we will get the sum of Squares within the Samples. (SSE)
- Divide the sum of squares with the degree of freedom. That will give Mean Square within Samples. (MSE).
- Degree of freedom = number of observations - number of samples.

Calculate the value of 'F'

It can be found out by using the following formula

$$F = \frac{\text{Mean square between samples}}{\text{Mean square within samples}}$$

Mean square within samples

'F' indicates the variance ratio. F value is compared with the critical value or table value of F distribution. If the computed value is greater than the critical value, we shall

reject the hypothesis that all the groups are drawn from the common population. If the computed value is smaller, accept the null hypothesis.

Analysis of variance – two way classification

In this case the data is classified according to two different criteria or factors. The procedure for calculation of F under two way classifications is summarized in the table given below.

ANOVA TABLE

Source of variations	Sum of squares	df	Mean square
Between columns	SSC	(c-1)	MSC = $SSC/(c-1)$
Between rows	SSR	(r-1)	MSR = $SSR/(r-1)$
Residual	SSE	(c-1)(r-1)	MSE = $SSE/(r-1)(c-1)$
Total	SST	(N-1)	

SSC indicates the sum of squares between columns

MSC indicates mean squares between columns

SSR denotes sum of squares between rows.

MSR indicates mean square between rows

SSE indicates the sum of Squares for the residual

MSE denotes the Mean Square with in residuals

SST is the total sum of squares.

In problems which involve two way classifications, 'residual' is the measuring rod to test significance. It represents the extent of variation due to forces called 'chance'. Variance between columns and Variance between rows are computed as above and compared with Residual Variance and two F values are obtained. These two F values are compared with concerned F table values, and the significance of difference decided.

Residual = total sum of squares – sum of squared between columns and sum of squares between rows.

Problem 6

The Calicut traders wishes to test whether its three salesman named James, Raghu and Basheer tend to make sales of the same size or whether they differ in their selling ability as measured by the average size of their sales. During the last week , there have been 14 sales calls. James made 5 calls, Raghu made 4 calls and the Basheer have made 5 calls. The following is the weekly sales record of the three salesmen.

James (Rs)	Raghul(Rs)	Basheer(Rs)
300	600	700
400	300	300
300	300	400
500	400	600
0	-	500

Perform the analysis of variance and conclude whether the three salesmen differ between themselves in sales efficiency.

Answer

Hypothesis- there is no differences between the selling ability of three salesmen.

This is one way analysis of variance. Coding method is applied to simplify data. '100' is taken as common factor. Calculated F value = 1.83

The table value of F at 5% level of significance for 2,11 degree of freedom is 3.98. The calculated value 1.83, is less than table value. So we accept the null hypotheses. We conclude that there is no significant difference among salesman in their ability measured in terms of mean sales per week.

Non parametric tests

While applying the parametric test like 't', F and Z test, it is assumed that samples are drawn from normally distributed population. It is also assumed that mean; standard deviation or proportion can be derived from sample data and can be used for estimating the corresponding population parameter. In certain situations it is not possible to make any rigid assumption about the distribution of the population from which samples are being drawn. This limitation has led to the development of non parametric tests. It is also known as distribution free methods. So while applying non parametric tests no assumption about population parameter of the populations or populations from which samples are drawn is made. Chi square test is the most widely used test under this category. A large variety of non parametric tests are developed by different statisticians. It includes sign tests, Fisher Irwin tests, Wilcoxon-Mann Whitney Test (or U test), Kruskal-Wallis Test (or H test) and Spearman's Correlation.

Chi square test

Chi square test is one of the important tests developed to test hypothesis. It is a non parametric test. This test is first used by Karl Pearson In the year 1900. We want to ascertain the frequency of events falling in specified categories in a number of business problems. For example in marketing research we count the number of people who prefer a particular brand of talcum powder. Chi square test is applied to such frequencies of occurrence as against the expected ones without referring the population parameters. It is frequently used for testing hypothesis concerning the differences between a set of observed frequencies of a sample and corresponding set of expected or theoretical frequencies. In

other words, it is a test statistic which measures the discrepancy between observed or actual frequencies and their corresponding expected frequencies is called the chi-square test χ^2

$$\chi^2 = \sum (O - E)^2 \text{ where,}$$

O = Observed frequencies

E = Expected frequencies

Degree of freedom = $n - 1$, if observations are given in a series.

Degree of freedom = $(r-1)(c-1)$, if a two way matrix is given.

Steps in Chi square Test

The following steps are required to determine the value of the chi-square test.

- Form null hypothesis that there is no significant difference.
- Consider given frequencies as observed frequencies.
- Calculate expected frequencies
- Take the differences between observed and expected frequencies.
- Obtain the square of the differences
- Divide $(O-E)^2$ with the expected frequency and get chi square values..
- Obtain total χ^2 value, by summatting calculated values.
- Compare the calculated value with Chi square table value.

The greater the discrepancy between observed and expected frequencies, the greater shall be the value of χ^2 . The calculated value of χ^2 is compared with the table value of χ^2 for a given degree of freedom at a certain specified level of significance. At 5% level of significance, the calculated value of χ^2 is more than the table value of χ^2 , the difference between theory and observation is considered as significant. That means it could not have arisen due to fluctuation of simple sampling. The value of χ^2 is always positive.

Problem 5 The table given below depicts the data obtained during an epidemic of cholera.

	Attacked	Not attacked	Total
Inoculated	21	479	500
Not inoculated	1751	1325	1500
Total	196	1804	2000

Test the effectiveness of inoculation in preventing the attack of cholera.

Answer

(H_0) = The inoculation is not effective in preventing the attack of cholera

That is inoculation and attack is independent. On the basis this hypothesis the expected frequency corresponding to the number of persons inoculated and attacked would be -

$$\text{Expectation of } (AB) = \frac{(A) \times (B)}{N}$$

Here A = Total of a column

B = Total of a row

AB = Grand Total.

Therefore A = 196, B = 500, AB = 2000

$AB = \frac{196 \times 500}{2000} = 49$. On the basis of the above value
we can find other values.

	Attacked	Not attacked	Total
Inoculated	49	451	500
Not inoculated	147	1353	1500
Total	196	1804	2000

Now put the observed frequencies and expected frequencies together and find chi square values.

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
21	49	28	784	16
175	147	28	784	5.33
479	451	28	784	1.74
1325	1353	28	784	0.58
				23.65

$$V = (r-1)(c-1) = (2-1)(2-1) = 1$$

When $V=1$, $\chi^2 0.05 = 3.84$

The calculated value of χ^2 is higher than the table value and hence the result of the experiment does not support the hypothesis. We therefore conclude that the inoculation is effective in preventing the attack of cholera.

Mann Whitney U test and Wilcoxon test

It is a most useful test to see whether the values in two samples differ in size. It resembles the Median -Test in scope, but it is much more sensitive. In fact ,for large numbers it is almost as sensitive as the Two Sample Student t-tests. For small numbers with unknown distributions this test is even more sensitive than the Student t-test. As it is only on rare occasions that we do know that values are Normal distributed, this test is to be preferred over the Student t-test. In this test data are ranked. It is used when only populations are involved. All the observations in the two samples are to be ranked. A null hypothesis is fixed that there is no differences between the qualities of two samples. The statistic U is calculated from the sum of the rank in sample one and sample two.

McNemar's Test

McNemar's Test is generally used when the data consist of paired observations of tables. This non-parametric test uses matched-pairs of labels (A,B). It determines whether the proportion of A- and B-labels is equal for both members. It is a very good test when only nominal data are available

The Sign Test

The Sign Test is a nonparametric test that may be of use when it is only necessary (or possible) to know if observed differences between two conditions are significant. That is to say, with appropriate use of the sign test, it would be possible to determine if X is really "more" than Y, or however the conditions are arranged.

The sign test is structured so that plus (+) and (-) "sign" are used to denote change in magnitude, as opposed to any attempt at quantitative measurement.

Factor analysis

Factor analysis is by far the most often used multivariate technique of research studies, specially pertaining to social and behavioural sciences. It is a technique applicable when there is a systematic interdependence among a set of observed or manifest variables and the researcher is interested in finding out something more fundamental or latent which creates this commonality. For instance, we might have data, say, about an individual's income, education, occupation and dwelling area and want to infer from these some factor (such as social class) which summarizes the commonality of all the said four variables. The technique used for such purpose is generally described as factor analysis. Factor analysis, thus, seeks to resolve a large set of measured variables in terms of relatively few categories, known as factors. This technique allows the researcher to group variables into factors (based on correlation between variables) and the factors so derived may be treated as new variables (often termed as latent variables) and the factors so derived by summing the values of the original variables which have been grouped into the factor. The meaning and name of such new variables is subjectively determined by the researcher. Since the factors happen to be linear combinations of data, the coordinates of each observation or variable is measured to obtain what are called factor loadings. Such factor loadings represent the correlation between the particular variable and the factor, and are usually placed in a matrix of correlations between the variable and the factors.

Mathematical basis of factor analysis

The mathematical basis of factor analysis concerns a data matrix (also termed as score matrix), symbolized as S. The matrix contains the scores of N persons of k measures. Thus a_1 is the score of person 1 on measure a, a_2 is the score of person 2 on measure a, and k_n is the score of person N on measure k. The score matrix then takes the form as shown following:

Discriminant Analysis

Discriminant analysis is a technique to discriminate between two or more mutually exclusive and exhaustive groups on the basis of some explanatory variable. These groups are known a-priori. When the criterion variable has two categories, the technique is known as two-group discriminant analysis. When three or more categories are involved, the technique is referred to as multiple discriminant analysis.

Discriminant analysis help the researchers who are interested to understand how consumers differ with respect to demographic and psychographic characteristic. Discriminate analysis is also used to predict the group membership. Banks use discriminate analysis to discriminate between the customers who default and who repay the loan in time, based on their age ,income, assets, number of dependents, and previous outstanding loan, etc.

Consider the following example:

Example 16.1: Suppose a company wants to determine if one of its new product- a new and improved digital camera will be commercially successful or not. The company wants to distinguish the purchasers from non-purchasers. The company makes a panel of 10 potential purchasers and devised the rating scales on three characteristic-durability, performance, and style. The panel is asked to give the digital camera rating from 1 to 10 on these three scales. After the product was evaluated , they are asked their buying intention("would purchase" or " would not purchase"). Five said that they would purchase the digital camera and the other five said that they would not purchase it. Such an analysis may lead to a conclusion based on the discrimination of the purchasers and non purchasers.

Cluster Analysis

The purpose of cluster analysis is to divide large group of subjects or observations, like customers or products, into smaller groups such that the observations within each group are dissimilar or far away. These smaller groups are called as clusters. Thus, resulting clusters exhibit high internal(within cluster) homogeneity and high external (between clusters)heterogeneity.

Cluster analysis is an inter-dependence technique and makes no distinction between dependent (study) and independent (explanatory) variables. Not that, we do not combine the variables here as done in factor analysis. In factor analysis, we reduce number of variables by grouping them into smaller set of factors, while in cluster analysis we reduce number of observations by grouping them into smaller set of clusters. Number of groups(or clusters) is not known in advance ; they are natural clusters as suggested by the data. Main applications of cluster analysis are segmenting the market and determining target markets, product positioning and new product development, and selecting test markets. Some applications of cluster analysis are illustrated below:

- Identifying people with similar patterns of past purchases so that you can tailor your marketing strategies.
- Clubbing television shows into homogeneous categories based on viewer characteristics.(market segmentation)
- Examining patients with a diagnosis of swine flu to determine if distinct subgroups can be identified based on a symptom checklist and results from pathological tests.

In cluster analysis, we group the observations based on a particular set of variables. These variables affect the clusters obtained using cluster analysis. There is no general rule of selecting the variables. We have to be careful while selecting these variables. After selecting the variables, we obtain sample observations on those variables.

Path Analysis

The term 'path analysis' was first introduced by the biologist Sewall Wright in 1934 in connection with decomposing the total correlation between any two variables in a causal system. The technique of path analysis is based on a series of multiple regression analyses with the added assumption of causal relationship between independent and dependent variables. This technique lays relatively heavier emphasis on the heuristic use of visual diagram, technically described as a path diagram. An illustrative path diagram showing interrelationship between Fathers' education, Fathers' occupation, Sons' education, Sons, first and Sons' present occupation can be shown.

Path analysis makes use of standardized partial regression coefficients (known as beta weights) as effect coefficients. In linear additive effects are assumed, then through path analysis a simple set of questions can be built up showing how each variable depends on preceding variables.²⁷ The main principle of path analysis is that any correlation coefficient between two variables, or a gross or overall measure of empirical relationship can be decomposed into a series of part: separate paths of influence leading through chronologically intermediate variable to which both the correlated variables have links.²⁸

The merit of path analysis in comparison to co relational analysis is that it makes possible the assessment of the relative influence of each antecedent or explanatory variable on the consequent or criterion variables by first making explicit the assumptions underlying the causal connections and then by elucidating the indirect effect of the explanatory variables.

The use of the path analysis technique requires the assumption that there are linear additive, a. symmetric relationships among a set of variables which can be measured at least on a quasi-interval scale. Each dependent variable is regarded as determined by the variables preceding it in the path diagram, and a residual variable, defined as uncorrelated with the other variables, is postulated to account for the unexplained portion of the variance in the dependent variable. The determining variables are assumed for the analysis to be given (exogenous in the model).²⁹

In the end, it may again be emphasized that the main virtue of path analysis lies in making explicit the assumption underlying the causal connections and in elucidating the indirect effects due to antecedent variables of the given system.

Canonical Correlation

This technique was first developed by Hotelling, wherein an effort is made to simultaneously predict a set of criterion variables from their joint co-variance with a set of explanatory variables. Both metric and non-metric data can be used in the dependent and

independent variables in such a way that linear composite of the criterion variables has a maximum correlation with the linear composite of the explanatory variables. For example, if we want to relate grade school adjustment to health and physical maturity of the child, we can then use canonical correlation analysis provided, we have for each child a number of adjustment scores(such as tests, teacher's ratings, parent's rating and so on) and also we have for each child a number of health and physical maturity scores(such as heart rate, height, weight, index of intensity of illness and so on). The main objective of canonical correlation analysis is to discover factors separately in the two sets of variables such that the multiple correlation between sets of factors will be the maximum possible. Mathematically, in canonical correlation analysis, the weights of the two sets viz., $a_1, a_2 \text{ and } y_1, y_2, \dots, y_j$ are so determined that the variables $x = a_1x_1 + a_2x_2 + \dots + a_kx_k$ and $Y = y_1 + y_2Y_2 + \dots + y_jY_j$ have a maximum common variance. The process of finding the weights requires factor analyses with two matrices. The resulting canonical correlation solution then gives an overall description of the presence or absence of a relationship between the two sets of variables.

Latent Structure Analysis

This type of analysis shared both of the objectives of factor analysis viz., to extract latent factors and express relationship of observed(manifest) variables with these factors as their indicators and to classify a population of respondents into pure types. This type of analysis is appropriate when the variables involved in a study do not possess dependency relationship and happen to be non-metric.

In addition to the above stated multivariate techniques, we may also describe the salient features of what is known as "Path analysis", a technique useful for decomposing the total correlation between any two variables in a causal system.

Conclusion

Our aim by a brief introduction of some multivariate techniques in this chapter is to motivate the reader to learn these techniques in details. We refer the motivated learners to Johnson and Wichern(2002). From the brief account of multivariate techniques presented above, we many conclude that such techniques are important for they make it possible to encompass all the data from an investigation in one analysis. They in fact result in a clearer and better account of the research effort than do the piecemeal analyses of portions of data. These techniques yield more realistic probability statements in hypothesis testing and interval estimation studies. Multivariate analysis (consequently the use of multivariate techniques) is especially important in behavioural sciences and applied researchers for most of such studies involve problems in which several response variables are observed simultaneously. The common source of each individual observation generally results into dependence or correlation among the dimensions and it is this feature that distinguishes multivariate data and techniques from their univariate prototypes.

In spite of all this, multivariate techniques are expensive and involve laborious computations. As such their applications in the context of research studies have been accelerated only with the advent of high speed electronic computers since 1950's.

Review Questions

1. state the role of data analysis in research
2. state approaches to data analysis
3. classify data analysis methods
4. what role computers play in research
5. classify tools of data analysis
6. what are measures of central tendency
7. how is Geometric Mean calculated
8. what is the significance of standard deviation
9. explain measures of dispersion
10. explain correlation, how is it measured
11. what is rank correlation
12. explain regression analysis
13. why there are two regression equations
14. explain time series analysis
15. what is canonical analysis
16. what is latent structure analysis
17. What do you understand by multivariate techniques?

UNIT 19 INTERPRETATION OF DATA

After collecting and analyzing the data, the researcher has to accomplish the task of drawing inferences followed by report writing. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. It is only through interpretation that the researcher can expose relations and processes that underlie his findings. In case of hypothesis testing studies, if hypothesis is tested and upheld several times, the researcher may arrive at generalizations. But in case the researcher had no hypothesis to start with, he would try to explain his findings on the basis of some theory. This may at times, result in new questions, leading to further researches. All this analytical information and consequential inferences may well be communicated, preferably, through research report, to the consumers of research results, who may be either an individual or a group of individuals or some public or private organization.

Analysis and interpretation are the integral parts of research. The important objective of analysis of data is to provide answers to the questions activated in research. Interpretation refers to drawing inferences from the collected facts after analytical study.

According to C. William Emory interpretation has two major aspects namely establishing continuity in research through linking the results of a given study with those of

another and the establishment of some relationship with the collected data. Analysis is not complete without interpretation. So, both are inter dependent.

In the words of Jahada and Cook scientific interpretation seeks for relationship between the data of study and between the study findings and other scientific knowledge. It helps the researcher to understand the conceptual principles that works underneath his findings. He can connect his findings with those of other studies having relationship with other established concepts.

It is also a critical examination of the results one own analysis of data in the light of all limitations of data collection. It is through the interpretation the researcher is able to understand the real significance of his findings.

Significance of Interpretation

Interpretation refers to the task of drawing inferences from the collected facts after an analytical or experimental study. In fact it is a search for broader meaning of research findings. It gained significance due to following aspects:

- Interpretation is the effort to establish continuity in research through linking the results of a given study with those of another.
- It is the establishment of some explanatory concepts. Interpretation is concerned with relationships within the collected data and the extension of study beyond the collected data as well.
- It is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic components of research process because of the following reasons.
- It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies.
- The interpretation of the findings of exploratory research study often results into hypothesis for experimental research and as such interpretation is involved in the transition from exploratory to experimental research.

Interpretation vs Analysis

There is no clear cut dividing line between analysis and interpretation . Though analysis we attempt to study the collected data and derive results while through interpretation we draw conclusions based on the results obtained. Analysis is not complete without interpretation and interpretation cannot proceed without analysis . So both are interdependent.

Functions of Interpretation

Interpretation refers to the technique of drawing inference from the collected facts and explaining the significance of those inferences after an analytical and experimental study. It is a search for broader and more abstract means of the research findings.

- Interpretation analyses the abstract relations in more concrete terms and tries to unfold the reason for the existing type of relations or findings. It relates the empirical findings with theoretical principles and helps to draw a number of useful inferences from the study. The prime task of interpretation is to bring to the surface the real importance of the findings. An interpretation should explain the findings of the analysis in objective terms.
- It is through interpretation that a researcher can understand real significance of his findings. Through interpretation he can appreciate why the findings are and what they are. Interpretation serves a twofold purpose. First, it gives an understanding of the general factors that seem to explain what has been observed in the course of a study. Second, it provides a theoretical conception which can serve in turn as a guide for further research.
- The work of interpretation is of a technical nature and therefore only experts can do it logically. If the interpretation is not done very carefully, misleading conclusions may be drawn. So any person without adequate knowledge of statistical methods as well as statistical operations and their implications, is likely to draw wrong conclusions.
- In interpretation statistical data, besides technical knowledge, common sense is equally important. A statistician can draw precise inference only if he is fully conversant with the problem under study and its implications. Extra care is to be taken in interpreting results where statistical methods were used. Accuracy in results depends, to a desirable degree, on the personal qualities and abilities of the interpreter. The interpreter must be creative of ideas. He should be free from bias and prejudice.

Fundamental principles of interpretation

Following principles should guide a researcher while interpreting his findings;

- Sound interpretation involves willingness on the part of the interpreter to see what is in the data.
- Sound interpretation requires that the interpreter knows something more than the mere figures.
- Good interpretation demands logical thinking.
- Clear and simple language is necessary for communicating the interpretation.

Need for interpretation

Interpretation is the ultimate gist of a research. Why interpretation is needed. The answer is the following;

- The utility of research findings lies on its proper interpretation. The interpretation of research report is necessary for the following reasons.
- The research involves the effort to establish continuity in research through thinking the results of a given study with those of other studies having the same abstract

principle. It is through interpretation that the interpreter is able to know the abstract principles lying in his conclusions.

- On the basis of the principles underlying his findings, a researcher can make various predictions about the various other events which are unrelated to his area of findings.
- Interpretation leads to the establishment of explaining concepts. The researcher's task is to explain the relations he has observed in the course of his study.
- After a research work is completed and findings are made, there arises an array of new questions constitute the ground work for new research undertakings and formulation of new theories which will either modify or replace old ones. It serves to open new and wider avenues of intellectual and stimulates the quest for more wider avenues of intellectual adventure.

Steps involved in the technique of interpretation

The interpretation is not an easy job. It requires great skill and intelligence. It is an art one learns through practice and experience. The technique of interpretation consists of the following steps.

- Researcher must give reasonable explanations of the relations he has found. He must interpret lines of relationship in terms of the underlying process. He must be able to see uniformity in diversified research findings so that generalization of findings is possible.
- If any extraneous information is collected during the study, it must be considered while interpreting the final results of research study.
- The researcher can consult with those having insight into the study who can point out the omissions and errors in logical arguments. These consultations will help the researcher to do correct interpretation.
- The researcher must consider all relevant factors affecting the problem at the time of interpretation.
- The conclusions appearing correct at the beginning may prove to be inaccurate later. So researcher must not be in a hurry while interpreting.

Precaution in interpretation (Pre-requisites of interpretation)

Interpreter must satisfy himself that the data used for interpretation are adequate and appropriate. If a sample is selected for study, it must be ensured that the sample represents the population truly.

- The data used for interpretation must reflect good homogeneity. If there is no homogeneity, comparison between items becomes difficult.
- Before accepting the data interpretation, it must be seen that there is stability of data. If there is no stability the same results will not repeat in similar experiments.
- The data taken for interpretation must be relevant to the problem.

- It must be ensured that interpretation is done through statistical methods.
- The interpreter must be cautious about the errors that may arise in the process of interpretation. Errors may arise due to false generalization, due to wrong interpretation of statistical measures etc. So the interpreter must remain vigilant about all such things so that false generalization may not take place. He must use correct statistical measures for drawing inferences.
- The interpreter must take all those precautions concerning with the reliability of data, Computational checks, comparison of results etc.

Errors of interpretation

The errors of interpretation can broadly be classified into two groups

Error occur when unwarranted conclusions are drawn from the facts available. For example: On the basis of decrease in the import of cosmetics, drawing conclusions that the women have become less fashionable is not correct. And , Drawing conclusions from an argument running from effect to cause also, will result into errors in interpretation.

- Comparison between two sets of data with unequal base will cause error in interpretation.
- When conclusion are drawn from data irrelevant to the problem, the conclusions are to contain errors.
- Errors are also caused due to false generalization and faulty statistical methods.
- Errors due to misuse of statistical measures
- When conclusions are based on what is true, on an average, such conclusions may some times lead to errors. For example : attributing the average behaviour of a group, to every member of that group results into wrong conclusions.
- When percentages are used for comparisons, it will cause wrong conclusions if the total numbers are different. For example : suppose the number of students appearing for an examination from two colleges A and B are respectively 1 and 50 . If one passes from college A, we say 100% result, while even if 45 students pass from the college B we say only 90% result. The conclusions is college A is better. This is not a fair conclusion. Thus percentage may reveal wrong conclusions.
- When index numbers are used without proper care wrong conclusions may be drawn, since the index number have their own limitations.
- High value of coefficient of correlation may show high degree of dependency. But sometimes , accidentally , also there can be high value of coefficient of correlation even when there is no relation between variables. Such correlation is called causal correlation.
- Conclusions drawn on the basis of coefficient of association may also be sometimes wrong. For example : a positive association between coaching

and passing examination may show that passing the examination is due to coaching. The fact that coaching alone is not the requirement for passing the examination is not revealed by the above result. Thus a wrong conclusion is drawn.

Review Questions and Exercises

- What is meant by interpretation?
- What are the objectives of interpretation
- Explain significance of interpretation
- Compare interpretation with analysis
- State functions of interpretation
- What are the principles of interpretation
- State the steps involved in interpretation
- What are the precautions to be taken in interpretation
- "Interpretation tries to unfold the reasons for the existing type of relations or findings". Substantiate.
- "In interpretation, common sense is important" Elucidate.
- Explain the importance of interpretation.
- What are errors of interpretation? Explain them.

UNIT 20 RESEARCH REPORT WRITING

Introduction

The last step in the research process is writing the research report. Each step of the process is important for a valid study, as negligence at any stage will affect all of the study, not just that part. In a way this last step is the most crucial as it is through the report that the findings of the study and their implications are communicated to your supervisor and readers. Most people will not be aware of the amount and quality of work that has gone into your study, while much hard work and care may have been put into every stage of the research.

The main difference between research writing and other writing is in the degree of control, rigorousness and caution required. Research writing is controlled in the sense that you need to be extremely careful about what you write, the words you choose, the way ideas are expressed and the validity and verifiability of the bases for the conclusions you draw. What most distinguishes research writing from other writings is the high degree of intellectual rigor required. Research writing must be absolutely accurate, clear, free of ambiguity, logical and concise. Assumptions as to the knowledge of readers about the study must be avoided. Bear in mind that you must be able to defend whatever you write should anyone challenge it. Do not use ornamental and superficial language. Even the best researchers make a number of drafts before writing up their final one, so be prepared to undertake this task.

Functions of Research Report

The purpose of a research report is to communicate to the readers regarding the methodology and the results of the study. It is to convince the reader what has been done and what is the outcome.

1. Research Report serves as a means for presenting the problem studied, methods and techniques used , findings conclusions and recommendations in an organized manner.
2. It serves as a reference material for future use in the same or elated area.
3. It serves as a means for judging the quality of the research project.
4. It is a means for evaluating research ability.
5. It provides systematic knowledge on problems and issues analyzed.

Essentials of good report

Following are the principles which should lead to a good research report:

- Clarity – clarity is the first essential requirement of a good research report. The researcher should be clear in his writing. Research report writing is a reasoning process. If there is no clarity of thought, there would not be clarity in writing also.
- Correctness – a research report should be written correctly. Incorrect writing is frequently ambiguous, some reports may become quite misleading and usually difficult to understand. To write correctly, one has to know grammar. Incorrect writing leaves a poor opinion about the researcher and his work. It also affects the image of the research guide.
- Brevity – the researcher should take care of the economy of words and concentration of ideas. Repetition and superfluosity will obstruct the lucidity of the report and will interfere with precision and clarity.
- Objective – the researcher should write an unbiased and objective report fortified by facts. It should not contain sentimental wordings and biased statements.
- Style –A good research report is one that is written in that style which the audience can follow. The style should suit the focus group. It should convince the audience of the truth of the ideas presented.
- Arrangement – proper organization of the report is the important ingredient in the whole process. Simply to know that a report has an introduction, body and conclusion is of little use. The real problem is to decide where to begin and what to say next.
- Consistency – there must be consistency of thought. If there is no consistency in the arguments, then the validity of the report itself will be doubted.
- Plan – planning the study materials and processes form the basis of preparing the research report. A well planned report facilitates proper understanding and correct reading.

Preliminaries to report writing

Before embarking on writing of the report, certain preliminaries have to be planned and implemented.

- The type of readers of the report must be identified. Readers may be from academic community or from general public or some research sponsors. The language, style and presentation must suit the readers.
- The purpose of report matters, and must be determined. Report must be designed to serve this purpose.
- Scope of the report is also to be discussed before preparing the report. Scope depends on the type of readers and the findings.

- Style of the report and format of the report should also be determined before preparing the report.

Significance of Report Writing

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented or written. As a matter of fact, even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others. Following are the aspects leading to the significance of report writing.

- The purpose of research is not well served unless the findings are made known to others, research results must invariably enter the general store of knowledge.
- Most of the universities and other organizations treat the presentation of research results or the writing of report as part and parcel of the research project.
- Writing of research report is the final step in a research study and as such, requires a set of skills somewhat different from those called for in respect of the earlier stages of research.
- This is crucial task which should be accomplished by the researcher with utmost care and he may seek the assistance and guidance of experts for this purpose.

Steps in report writing

Research reports are the product of slow, painstaking, accurate inductive work. The usual steps involved in writing report are the following:

- Logical analysis – it is the first step which is primarily concerned with the development of a subject. There are two ways in which to develop a subject – logically or chronologically. The logical development is made on the basis of mental connections and associations' between the one thing and another by means of analysis. Chronological development is based on a connection or sequence in time or occurrence. Directions for doing or making something usually follow the chronological order.
- Final outline – outlines are the framework upon which long written works are constructed. They are an aid to the logical organization of the material and a reminder of the points to be stressed in the report.
- Rough draft – this follows logical analysis of the subject and the preparation of the final outline. Such a step is of utmost importance for the researcher now sits to write down what he has done in the context of his research study. He will write down the procedure adopted by him in collecting the material for his study also with various limitations faced by him, the technique of analysis adopted by him, the broad findings and generalizations and the various suggestions he wants to offer regarding the problem concerned.
- Rewriting – usually this step required more time than the writing of the rough draft. The careful revision makes the difference between a mediocre and good piece of writing. While reworking and polishing, one should check the report for weaknesses in logical development or presentation. The researcher should also see whether or not the material, as it is presented has unity and cohesion and the report stand upright and firm and exhibit a definite pattern.

- Bibliography – next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended to the research report, is a list of books in some way pertinent to the research which has been done. It should contain all those works which the researcher has consulted.
- Final draft – this constitutes the last step. The final draft should be written in a concise and objective style and in simple language, avoiding vague expressions such as "it seems", "there may be". And so on. While writing the final draft, the researcher must avoid abstract terminology and technical jargon. Illustrations' and examples based on common experiences must be incorporated in the final draft.

Writing Bibliography

There are several well established systems for writing a bibliography and your choice is dependent upon the preference of the discipline and university regulations. In the social sciences, some of the most commonly used ones are the following :

- Harvard System
- American Psychological Association system
- American Medical Association system
- McGraw Hill system
- Modern Languages Association system
- Footnote system.

The entries in bibliography may be made adopting the following order.

For books

Name of author, last name first. Title underlined to indicate italics. Publisher, place and date of publication and number of volumes.

For magazines

Name of the author, last name first, title of article, in quotation marks, name of periodical, underline to indicate italics, volume number and issue number, date of issue, and page number.

Lay out of the Research Report

The research should convey about the study so that he can place it in its general scientific context, judge the adequacy of its methods and thus form an opinion of how seriously the findings are to be taken. For this purpose there is the need of proper lay out of the report. Following should be the broad contents.

1. Preliminary pages – in its preliminary pages the report should carry a title and date, followed by acknowledgement's in the form of Preface or Foreword. Then there should be a table of contents followed by list of tables and illustrations so that the decision maker or anybody interested in reading the report can easily locate the required information in the report.
2. Main Text – the main text proves the complete outline of the research report along with all details. Title of the research study is repeated at the top of the first page of the main text and then follows the other details on pages numbered consecutively, beginning with the second page. The main text of the report should have introduction, statement of the problem, objectives of the study, hypotheses if any, data collection method, analysis, findings and recommendations, the results and a summary.

3. End matter - at the end of the report, appendices should be enlisted in respect of all technical data such as questionnaires, sample information, mathematical derivations and the like on3es. Bibliography of sources consulted should also be given. Index which is an alphabetical listing of names, places and topics along with the numbers of the pages in a book or report , on which they are mentioned or discussed should be invariable given at the end of the report. The value of index lies in the fact that it works as a guide to the reader for the contents in the report.

Types of Reports

Research reports vary greatly in length and type. In each individual case, both the length and form are largely dictated by the problems at hand. For instance, business firms prefer reports in the letter form , just one or two pages in length. Banks, insurance organizations and financial institutions are generally fond of the short balance sheet type of tabulation for their annual reports to their customers and shareholders. Mathematicians prefer to write the results of their investigations in the form of algebraic notations. Chemists report their results in symbols and formulae. Students of universities prefer long reports presenting the critical analysis of some period, based on hypothesis testing and analysis, findings and suggestions.

Thus, research reports can be presented in a number of ways. A preliminary classification of reports throws light on oral report, written report, popular report, technical report, interim report, summary report etc.

Oral Report

What goes on in seminars, conferences, symposia etc are mainly oral presentation. The merit of this approach lies in the fact that it provides an opportunity for give and take decisions which generally lead to a better understanding of the findings and their implications. But the main demerit of this sort of presentation is the lack of any permanent record, concerning the research details. The findings may fade away from people's memory even before an action is taken. Oral presentation is effective when supplemented by various visual devices like slides, charts etc.

Written Report

Presentation of information or results in writing requires many qualities and a great deal of effort by the researchers. Utmost attention should be paid for writing the research report. Most of the universities in India, insist on written report in prescribed formats. Written reports are of different types such as popular report, technical report, interim report, summary report and so on.

Popular report

A popular report is used when the research results have policy implications. This is the report meant to be read by general public or layman. It should present broad facts, findings and recommendations. It must be interesting, simple and lucid. It must avoid all technical jargons and details as to the method of investigation. For example, in a study of Indian economy, the laymen are interested in how economy is performing in broad terms and what is going or happen to their standard of living. They are not generally interested to such details as to the capital -output ratio etc.

Technical report

a technical report is used whenever a full written report of the study is required. technical report is written by an expert to be read by another expert. in this sense, a thesis is a technical report intended to be read by another researcher, or research guide. Not only thesis, but even a monograph, a journal article, class room project etc are technical

writings. The researcher gives full account of the technical aspects both in the sampling methods and in the subject matter. Conceptual and analytical frame work and sample designs are adequately explained. The four distinguishing aspect so technical report are the problem and its nature, method and its scientific accuracy, data and their dependability and the logical relationship of the data and the interpretation.

Interim report

When there is long time gap between data collection and presentation of report in the case of long projects, the study may lose its significance and usefulness. In such cases and interim report can be represented. Interim report may contain either the first results of the analysis or final outcome of the analysis of certain completely analyzed aspect. Interim report helps to take action without waiting for the full report. It prevents misunderstanding about the delay.

Interim report contains narration what has been done so far and what were its outcome it presents a summary of the findings of that part of analysis which has been completed.

Summary Report

A summary report is generally prepared for the consumption of the general public. This type of report is desirable for any study whose findings are of general interest. It is written in simple language avoiding technical jargons, using pictorial charts. It contains brief reference to the objective of the study, major findings and their implications. It is a short report of two or three pages. Its size is limited to be suitable for publication in daily news papers.

Rules in writing Research Report

There are very definite and set of rules which should be followed on the actual preparation of the research report. They are :

- Format – the way in which thoughts and information are to be expressed and presented is called the format of the report. The format should be in the prescribed form, if any or it should be attractive and interest generating. Its overall set up, neat presentations with diagrams, graph and statistical tables, simplicity in sentence construction, clarity in ideas, substantial words phrases and idioms, all contribute to the format of the report. The researcher must show care in using even a word in the report.
- Sections – researcher should divide the text of the report into different parts, sections , sub sections, and chapters. Depending on the number of issues discussed in the chapter, any number of sections may be created. Each section should discuss on a particular issue.
- Headings – proper headings and subheadings are to be given to the divisions, subdivisions, chapters, sections and subsections to make the report more attractive. The kinds of headings employed include central headings, side headings and paragraph heading.
- Paragraphing – for good research writing, appropriate paragraphing is an ornament. Paragraphing is the most vital job in thesis writing. Each part should contain only one central idea. Linking paragraphs and various chapters is an essential technique for maintaining continuity.
- Notes and Footnotes – it is common in the writing of reports to indicate the source of information, wherever the material is borrowed. Such information s are given

separately either at the bottom of the page as footnote or at the end of the chapter as note.

- References – a reference is the material used by the author, when it is not of his own. Referencing should contain certain essential information such as the author, the title, place of publication, publisher and the year of publication.
- Tables and figures – researcher has to make use of tables and figures for presenting matters in his report. A table is a tabular form of presenting statistical data. Matters presented by means of figures can be easily understood and compared. Care should be taken to avoid duplication of the usage of figures , tables, and diagrams simultaneously for a single analysis.

Referencing

The research report should follow an academic style of referencing . there are four referencing systems from which to choose, and you need to adopt the one that is acceptable to your university and academic discipline. The first of these is used in most general books, th second mainly in science and social science books, the third and fourth less frequently used. Those four referencing systems are :

- The short title system,
- The author date system.
- The reference by number system
- The author number system.

Review Questions

1. What do you mean by research reporting
2. State the importance of reporting in research
3. Differentiate research writing and other writing
4. What are the functions of research report
5. State the essentials of good research report
6. What are the preliminaries in research writing
7. Explain the steps to be followed in research reporting
8. What do you mean by bibliography
9. State prominent bibliographic systems
10. What are the contents of a research report
11. Explain types of research reports
12. What is the purpose of an Interim Report
13. State the rules in research report writing
14. What is a summary report
15. What are the rules to be followed in referencing

