Global Shiksha institute in Collaboration with Singhania University

BUSINESS RESEARCH METHODS

COURSE: MBA/EMBA
FIRST SEMESTER

MBA

I SEMESTER

SUBJECT CODE: MB12

Unit – I

Research: Concept, Characteristics and Types; Pure, Applied, Action and Inter Disciplinary Research – Logic and Scientific Investigation.

Unit – II

Research Problem: Identification, Selection and Formulation of a Research Problem - Research design - Literature Search and Review of Literature - Hypothesis – Definition - Types and Characteristics.

Unit – III

Research Methods: Survey – Historical – Case Study – Experimental - Sampling – Definition - Types and Relevance.

Unit – IV

Data Collection: Data Sources – Primary Sources and Secondary Sources - Data Collection Methods – Questionnaire, Interview, Observation

Unit – V

Data Analysis: Analysis and Interpretation – Statistical Tools and Techniques – Measure of Central Tendency - Frequency Distribution - Regression and Correlation – Scales - Statistical Tools – Report Writing - Style and Structure of Presentation of Data..

Books Recommended for Further Readings

1) Busha, Charles H. and Harter, Stephen P. Research Methods in Librarianship: Techniques and Interpretation. New York: Academic Press, 1980.

2) Gopal, M. H. An Introduction to Research Procedure in Social Science. Bombay: Asia, 1964.

3)Goode, William J. and Hatt, Paul K. Methods in Social Research. New York: McGraw Hill, 1952.

4)Krishan Kumar. Research Methods in Library and Information Science, New Delhi Vikas; 1992.

5)Morice, B. L. Library Surveys: An Introduction to the use, Planning Procedure and Presentation of Surveys. 2nd Ed. London: Bingley, 1982.

6)Ravichandara Rao, I. K. Qualitative Methods for Library and Information Science. New Delhi: Wiley Eastern, 1985.

7)Simon, J. L. Basic Research Methods in Social Science: The Art of Empirical Investigation, 1989.

8)Stevens, Rolland E. Research Methods in Librarianship. London: Bingley, 1971.

9)Ward, P. L Introductory Guide to Research in Library and Information Studies in the UKLondon Library Association, 1975.

10)Young, Pauline V. and Schmid, C. F. Scientific Social Surveys and Research. New Delhi: Prentice Hall, 1984.

11)Devarajan, G. Research in Library & Information Science. New Delhi; Ess Ess, 2002

12)Kothari, C. R. Research Methodology – Methods & Techniques. New Delhi; Vishwa Prakasam., 1996.

13)Foster, J. J. Data Analysis Using SPSS for Windows. Versions 8 &10: A Beginer's Guide. New Delhi; SAGE Pub., 2001.

14)Panda, B. D. Research Methodology for Library Science. New Delhi; Anmol, 1997.

15)Santhosh Gupta. Research Methodology and Statistical Techniques. New Delhi; Deep and Deep, 2000.

16)Saravanavel. Research Methodology. Allahabad; Kitab Mahal, 1994.

17)Sehgal, R. L. Applied Statistics for Library Science Research. 2 Vols. New Delhi; Ess Ess, 1998.

UNIT I

CHAPTER I

Introduction to Research Methodology

Research has proved to be an essential and powerful tool in leading man towards progress. Research is a matter of raising question, and then trying to find answer. In other words, Research means a sort of investigation describing the fact that some problem being investigated to shed for generalization. Therefore, Research is the activity of solving problem, which adds new knowledge and developing a theory as well as gathering of evidence to test generalization.

All significant research leads to progress in some field of life or the other. Each year new products, new facts, new concepts and new ways of doing things come into our lives due to ever increasing significant research in physical, biological, social and psychological fields. Research activity is no longer confined to the science laboratory. Even as the manufacturers, the agricultural experts and the archeologists carry on research, in their respective sphere so also the sociologists, economists and educationists. The aim of all research is progress and good life.

Decisions based on systematic research would surely save money, energy, and a lot of failure and frustration and show us the path of progress. Thus, it is not so difficult to show that research is extremely necessary and very worthwhile. It is a careful search for solutions to the problems that plague and puzzle the mankind.

Definition of Research

George G. Mouly of University of Miami defines research as "The systematic and scholarly applications of the scientific method, interpreted in its broader sense to the solutions of educational problems. Conversely, any systematic study designed to promote the development of education as a science can be educational research".

Rusk writes, "Research is a point of view, an attitude of inquiry, or a frame of mind. It asks questions which have hither to not been asked, and it seeks to answer them following a fairly definite procedure. It is not a mere theorizing, but rather an attempt to elicit facts, and to face them once they have been assembled".

Francis G. Cornell feels "to be sure, the best research is that which is reliable, verifiable and exhaustive so that it provides information in which we have confidence. The main point

here is that research is literally speaking, a kind of human behaviour, an activity in which people engage". By this definition, all intelligent human behaviour involves some research.

Maduri Singh defines "Teaching research is essentially a state of mind-friendly, welcoming attitude towards change".

Clifford Woody of the University of Michigan writes in an article in the Journal of Educational Research, "Research is a careful inquiry or examination in seeking facts or principles and diligent investigation to ascertain something".

C. C. Craford, University of Southern California, writes that "Research is simply a systematic and redefined technique of thinking, employing specialized tools, instruments and procedure in order to obtain a more adequate solution of a problem, then would be possible under ordinary means. It starts with a problem, collects data or facts, analyses them critically, and reaches decisions based on the actual evidence. It evolves original work instead of mere exercise of personal opinion. It evolves from a genuine desire to know rather than a desire to prove something. It is quantitative, seeking to know not only what but how much, and measurement, is therefore a central feature of it".

According to J. W. Best, "Research is considered to be the more formal, systematic, intensive process of carrying on the scientific method of analysis. It involves a more systematic structure of investigation usually resulting in some sort of formal record of procedures and a report of results or conclusions".

According to P. M. Cook "Research is an honest, exhaustive, intelligent searching for facts and their meanings or implications with reference to a given problem".

W. S. Monroe defines research as a method of studying problems whose solutions are to be desired partly or wholly from facts. The facts dealt with in research may be statements of opinions, historical facts, those contained in records and reports, the results of tests, answers to questionnaires, experimental data of any sort, and so forth. The final purpose of educational research is to ascertain principles and develop procedures for use in the field of education. Therefore, it should conclude by formulating principles or procedures. The more collection and tabulation of facts is not enough, though it may be preliminary to it or even a part thereof.

Research is thus an original contribution to the existing stock of knowledge of 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. The systematic approach concerning generalization and formulation of a theory is also research. As such the term 'research' refers to the systematic method consisting of enunciating the problem, formulating the hypothesis, collecting the facts or data, analyzing the facts and reacting certain conclusions either in the form of solution(s) towards the concerned problem or in certain generalizations for some theoretical formulation.

According to R. M. Heetehins, "Research in the sense of the development, collaboration and refinement of principles together with the collection and the use of empherical materials to aid in these process, in one of the highest activities of a University and one in which all its professors should be engaged".

In brief, significance of each letter of the word RESEARCH may be seen as below.

R - **R**ational way of thinking

E - **E**xpert and Exhaustive treatment

S - **S**olutions

E - **E**xactness

A - **A**nalysis of Adequate data

R - **R**elationship of facts

C - 1. **C**areful recording
2. **C**ritical observation
3. **C**onstructive attitude
4. **C**ondensed and Compactly stated generalizations

H - **H**onesty and Hard work in all aspects of the treatment of data.

Characteristics of Research

1.Research gathers new knowledge or data from primary or first hand sources

Research endeavors to reach the first hand source (primary source) of data instead of serving its purpose with the data available from second hand sources (secondary source). It is not research when one simply restates or reorganizes what is already known or what has been written.

2.Research in expert, systematic and accurate investigation

The researcher tries to secure an expertise before undertaking any investigation. The data are gathered, recorded and analyzed as accurately as possible. He/she uses standardized and valid data gathering tools or instruments. He/she also employs mechanical means to improve upon the accuracy of human observation, recording and computation of data.

3.Research is logical and objective

The researcher eliminates personal feelings and preferences from his/her research activity. He/she resists the temptation to seek only that data which support his/her hypothesis. There is no attempt to persuade or to prove. The emphasis is on testing, rather than on proving the hypothesis.

4.Research endeavors to organize data in quantitative terms

Research endeavors to organize data in quantitative terms as far as possible. It further tries to express them as numerical measures.

5.Research is patient and unhurried

The researcher is willing to make pains taking efforts. He/she works patiently towards sound conclusions. He/she realizes that significant findings do not come as a result of hurried and careless procedures.

6.Research requires courage

The researcher is not afraid of unpleasant consequences of his/her findings. He/she has the guts to speak and record the truth. He/she is willing to follow his/her procedures to conclusions that may be unpopular and bring social disapproval.

7. Research is highly purposive

It deals with a significant problem, which demands a solution.

8. Research places emphasis upon the discovery of general principles

It places emphasis upon the discovery of general principles and scientific generalizations that can be applied to the solution of a wide range of problems.

9. Research maintains rigorous standards

The researcher is expected to be a scholarly and imaginative person possessing highest integrity. He/she takes each step according to expected specifications and norms. He/she keeps his/her work scrupulously free from loopholes. Research is a job of great responsibility.

10. Research is carefully recorded and reported

Every term is carefully defined, the procedures are described in detail, the limiting factors are recognized, the references are carefully documented and the results are objectively recorded.

11. Research involves hypothesis

Research usually involves, as a step, a hypothesis or a set of hypotheses concerning the tentative explanation of a phenomenon or the solution of a problem.

According to John Best, research is a more systematic activity directed towards discovery and the development of an organized body of knowledge.

Criteria of good research

Whatever may be the types of research works and studies, it is important to note that they all meet on the common ground of scientific method employed by them. The following are the criteria, which we expect the scientific research to satisfy:

1. The purpose of the research should be clearly defined and common concepts should be used.
2. The research procedure used should be described in sufficient details to permit another researcher to repeat the research for further advancement, keeping the continuity of what has already been done.
3. The procedural design of the research should be carefully planned to yield results that are as objective as possible.
4. The researcher should report with complete frankness, flaws in procedural design and estimate their effects upon the findings.
5. The 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 the data should be checked carefully.
6. 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.
7. 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

1. Good Research is systematic

It means that research is structured with specified steps to be taken in a specified sequence in accordance with the 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.

2. Good Research is logical

This implies that research is guided by the rules of logical reasoning and the logical process of induction and deduction are of great value in carrying out research. Induction 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.

3. Good Research is empherical

It implies that research is related basically to one or more aspects of a real situation and deals with concrete data that provides a basis for external validity to research results.

4. Good Research is replicable

This characteristic allows research results to be verified by replicating the study and thereby building a sound basis for decisions.

Hindering factors for research

1. Lack of time, energy and resources has been a great hindrance for most of the researchers.
2. Traditional activities may be a powerful retarding factor.
3. Due to unworthy expenditure of time and money, research activities has failed in one or more instances.
4. Research is out of step with the prevailing educational philosophy.
5. Research procedure may be found difficult – even in comprehensive.
6. Research is considered to be the business of a few arm-chaired academicians.
7. Research has threatened by certain supersensitive and influential teachers, administrators and educators.

Qualities of a good researcher

Like qualities of good research, a researcher must possess some qualities to carryout the investigation successfully. The qualities of a good researcher are as follows.

1. Scientific attitude with reflective thinking
2. Devotion
3. Unbiased attitude
4. Physical characteristics
5. Mental Caliber
6. Ethical standards

Carter V Good gives the following hallmarks to the scientists:

1. Problem awareness
2. Appropriate specialization
3. Acquaintance with the related literature
4. Intellectual curiosity and drive
5. Willingness to pay the price
6. Creativity and ingenuity

M. H. Goal summarizes the qualities of a good researcher as follows:

A true scientist possesses the devotion of a mother, the poise of a judge, the objectivity of a philosopher, the courage of a soldier, the perseverance and patience of bearer, the fervor of a patriot and the vision of a prophet.

In short, a good researcher should

1. work diligently
2. think clearly
3. study carefully and extensively
4. observe minutely
5. collect the data painstakingly
6. analyze data objectively
7. conclude rationally and scientifically.

Objectives of Research

The objective of research is to find out answers to questions through the application of systematic and scientific way.

Though there is a specific purpose behind each research study, however, the objectives can be broadly classified as under;

- To obtain familiarity of a phenomenon.
- To determine the association or independence of an activity.
- To determine the characteristics of an individual or a group of activities and the frequency of its (or their) occurrence.

Nature of Research

The goal of research is to improve the level of living in society. The word research carries an aura of respect. The prestige of much is great and is equally emphasized on the researcher to challenge the problem that needs to be solved even thought the situation is far beyond the scope of research. Some people think the research is a waste of time, effort and money, and further think that more pure research is needed for practical value. Personal opinion guided by prejudice or dogmatism may be conceded in nature of problems but can be solved only on the basis of research evidence. The researcher has to implement more effectively has value judgment for the benefit of much to the society. One can perhaps say that a researcher enjoy the intellectual freedoms, independent thinking and can have rewarding experiences. And further, it is he (researcher) as adviser has to take the society towards the right path of development.

Significance of Research

The significance of research may be analyzed as under:

1. Research has an important role to guiding social plan. Knowledge of the society and the cultural behavior of people require proper planning for their well development because of both knowledge and the cultural behavior of human being are independent. A reliable as well as factual knowledge may be needed to take decision for planning. This is possible only by means of research; Social research is generally worth much more than of the area since the success of the former depends ultimately on people's acceptance and participation. So any plan for success of our economic development needs to be taken research.

2. Knowledge is a kind of power with which one can force the implication of particular phenomena. It also dispels the "Thrust" of odd settings, superstitious, etc., and light is thrown on them for welfare development. Thus social research may have the effect of promoting better understanding and social cohesion.
3. Research is charge with the responsibility for effective functioning of facts. Thus it affords a considerably, sound basis of prediction. Otherwise it leads to failure-bound programme which may have a serious impact on society. For example, Chernobyl Nuclear plant disaster. Another is Bhopal gas disaster. Thus prediction serves better control over the phenomena and helps in successful plan. This leads towards cherished goals. For example the implementation of family planning with prediction is an example.
4. It is the role of the researcher to effect constant improvement in techniques of his trade i.e., Research. He is in spatial temporal contexts; each challenging his attack is faced with the need to improve upon his techniques. In other words the technique of research has to become greater perfection. So big contentment in research and so big revolutionary in technological change. Examples are Electronics, Industrialization.

CHAPTER II

Types of Research

Research can be classified differently depending upon the approach, the purpose and the nature of a research activity. Broadly speaking, research is categorized as (i) Fundamental, pure or Theoretical Research; and (ii) Applied Research and (iii) Action Research.

Fundamental/pure Research

Research motivated by the desire to know or understand for the sake of knowing is called fundamental research. It is mainly concerned with generalization and with the formulation of a theory. On other words, gathering knowledge for the same of knowledge is termed 'pure' or basic research.

Research concerning some natural phenomenon or relating to pure mathematics are examples of fundamental research. Similarly, research studies, concerning human behaviour carried on with a view to make generalizations about human behaviour, are also examples of fundamental research.

The fundamental research always results in the discovery of a new theory. This discovery has nothing to do with n existing theory. Galileo's or Newton's contributions are fundamental in character and they form the basis for different theories.

Fundamental research also leads to the development of the existing theory. It brings out improvement in the existing theory either by relating some of its assumptions or by reinterpreting it. Since theory is always based on assumptions, there exists enormous scope for altering or formulating new set of assumptions and adding new dimension to the existing theory, Relaxing assumptions, altering them or making new ones altogether depends upon how a researcher views the existing theory. In a dynamic society a scholar may ascertain that earlier assumptions have become obsolete or have been inadequately defined. Thus the existing theory may appear to be outdated and implausible with the prevailing conditions.

For example, Malthusian population theory became almost useless in his own country owing to new developments invalidating the assumptions of his theory. Naturally, therefore by dropping out the invalid assumptions, researchers come out with new theories on population behaviour. There have also been attempts to reinterpret the Malthusian doctrine and thus seek to retain its validity character even now. Similarly by questioning some of the assumptions of

Keynesian theory, Friedman came out with new interpretations of the monetary phenomenon. Theories developed in capitalist countries often been challenged by researchers of the socialist block and they have developed new theories or reinterpreted the existing theory.

The researcher engaging in the pure research derives greatest satisfaction from increasing his knowledge in a field of enquiry where many questions remain unanswered. To him the challenge of not knowing is paramount. If he can solve the problem, he is satisfied in the results may or may not have any practical use. The 'pure' scientist would probably argue that knowledge itself is always of practical use in the end. His governing principal is that scientific enquiry is noble in itself; it is its own reward. To keep digging away at the layers of intellectual questions is a challenge enough for a 'pure' scientist. To him "knowledge is the highest good, truth is the supreme value is all the rest is secondary and subordinate".

Applied Research

Applied research is based on the application of known theories and models to the actual operational field or population. The applied research is conducted to test the empirical content or the basic assumptions or the very validity of theory under given conditions. For example, Lewis's growth model for labour surplus economics assumes that real wage rate of labour shall remain constant till the surplus labour is completely wiped out; it may of interest to a researcher to investigate if it so happens in every labour surplus economy. In case of a theory or model not holding good, the researcher's interest may further be stimulated to know why a given model does not apply and what modifications would be required to make the model operational in that situation. It may also form a basis for developing an alternative strategy.

Applied research has practical utility. The central aim of applied research is to discover a solution for some pressing practical problem. The 'applied' scientist always views in research in a practical context from the onset. He would define a problem as one in which some action could be taken to improve matters. The applied scientist is thus, much more likely to be working within a certain set of values and norms preference. Sociologists study a host of problems, viz., the problem of juvenile delinquency, of old people, of gange etc., and find out the measures to remedy the situation.

Applied research often takes the form of a field investigation and aims at collecting the basic data for verifying the applicability of existing theories and models in given situation. Naturally, therefore, the adequacy and accuracy of data will have considerable impact on the

application of the model and reliability of the results. Not only should his data be reliable, he must also be objective, scientific and sharp in identifying the field of application for a given theory. Like "Pure" research, the applied research also contributes to the developments in the following manners.

i) If can contribute new facts

ii) It can put theory to test

iii) It may aid in conceptual classification

iv) It may integrate theory of previous study/studies

In brief, Basic research is derived from fundamental intellectual problems; it tends to be of an original and theoretical nature. On the other hand, applied research is pragmatic; its purposes are more specific and are generally aimed at solving practical problems or at the discovery of new knowledge that can be utilized immediately in actual "real world" situations.

On the whole, completed studies in librarianship have been of an applied nature. Like any developing discipline, the study of library science is characterized by a weak body of theory and research findings that are often irreconcilable with previously acquired knowledge. Most contributions to the study of libraries have been concerned more with the collection of information that allows library phenomena to be identified and described, rather than for the purpose of explaining why and how library circumstances occur. An apparent distrust of theory on the part of many investigators in librarianship has contributed to neglect of basic research in the field.

Action Research

Action Research involves the application of the steps of scientific method to class room problems. Action research is similar to applied research in many ways. Applied research usually involves a larger number of samples as compound with action research. In action research many research projects are carried out in single class room by a single teacher, while others may be carried on by all teachers in a school or even a school district.

Action research is not mainly concerned in obtaining generalized, scientific knowledge about educational problems but in obtaining specific knowledge concerning the subjects involved in the study.

Action research is the study undertakes with the goal of immediate application. In other words, it refers to study of "On the job" problems. In this type of research the researcher

appears as a participant rather than observer, and therefore he is actively and even emotionally involved in the result and their application. It is a special type of research in the sense, testing of applications in accordance with a set of situation and modified according to the local prevailing conditions as well. Another feature is that it adopts itself to the changes that have taken place in the particular community; in case of applied research the evolution of entire community rather individual will be taken with applications of theory. These action research is similar to applied research but differ from the action process carry of action research is depended upon the type of feedback information supplied. Difference phase in action research are to be used for information gatherings. They are

a) A base line survey in order to get possible information relevant to the subject matter should be collected to get an idea of the existing situation.

b) Launching the action research subject.

c) The periodical assessment of the action research project.

d) Making changes and modification in the working of the project.

Depending the availability of information the method must be selected for the action research which usually consisting of the personal interview method, the survey method, the attitude measurement and so on. The action research is not much difference to concurrent evaluation of a project the example for action research are:

As stated earlier. Teach conduct action research to improve his own teaching. The practitioner attempts to study his problem scientifically for guiding, correcting and evaluating actions.

The scope of action research is very vast. The approach of action research is dealing with practical problems seems to be appropriate and promising for all kind of professional workers in education as long as their desire is to improve their own career. An administrator who dissatisfies with his efforts top develop a morale in his staff could approach this problem with action research.

The merits of action research are as follows:

1) Action research helps in decision making and action.

2) It broadens and deepens the fund of knowledge.

3) The helps to acquaint with research sight.

4) The problem solving activity of action research helps in implementing one's own findings in to practice.

5) The experience of action research strengthens the profession as a whole.

6) It brings experimental out look among the workers.

S. No.	Area	Fundamental Research	Action Research
1.	Objectives	Its purpose is to develop and test the educational theories and to obtain principles that will be applicable to a broad population.	Knowledge obtained intended to be applied in local setting and to provide a sort in service training to participating fenches and administrator.
2.	Training	Expert training is needed in measurement research methods and statistics.	Only a limited training is needed. Even if the research skills of scholar are low, action research can be done under the guidance (or) assistance of a consultant.
3.	Locating the research problem	A wide range of methods is used to locate the research problems.	The participating teachers identifying the problems which hinder class room teaching-learning processes.
4.	Involvement	A wide range of methods is used to locate the research problems.	The teacher is invariably involved in research problems
5	Hypothesis	Highly specific hypothesis are developed	Usually a specific statement of the problems serves as hypothesis
6	Review of literature	An exhaustive and through review of the literature is done to have a thorough understanding of the accumulated knowledge in the research processes	There is no need for such a thorough and intensive review. A glance at the review will serve the purpose of developing general understanding of the area.
7	Sampling	Research worker is to obtain a random or otherwise unbiased sample of the population beings studies.	Pupils studying in the class of the teacher are used as subjects
8	Experimental design	Careful attention is given to maintain comparable conditions and reducing error and bias	Bias usually present because participating teachers are ego involved in research situations processing are planned only in general terms.

9	Analysis of Data	Complex analysis is often called for	Simple analysis procedure are usually sufficient
10	Statistical Treatment	Statistical significance is usually stressed	Subjective opinion of participating teachers is often weighed heavily
11	Application of results	Lack of coordination between research workers and teachers generate serious practical problems. The generalization usually remain continued to books and research report	Findings are applied immediately to the classes of participating teachers send lead to for reaching improve in the teaching learning process

The difference between fundamental research and action research should not be magnified. The difference is there but in the emphasis only not in method or spirit. Both type are committed to the high standard of scientific objectively and scholarship.

CHAPTER III

Logic and Scientific Investigation

Meaning of Research

The Webster's International Dictionary proposed a very inclusive definition of research as "a Careful critical inquiry or examination in seeking facts or principles; diligent investigation in order to ascertain something". D.Slesinger and M.Stephenson in the encyclopedia of social sciences define research as 'the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether the knowledge aids in construction of theory or in practice of an art'. Research is any systematic quest for knowledge that is characterized by discipline inquiry. In the process of conducting research, an investigation is concerned with three major activities (1) the definition of terms, (2) the posing of questions or stating of propositions which incorporate the defined terms in an explanatory fashion, and (3) the testing of propositions or the search for answers to exploratory questions within the scope of these activities, the adopt research worker specifies objectives, explains problems with conceptual clarity, imposes structure on the inquiry by relating the research problem to relevant theories, and selects appropriate methods for observing conditions or events and for measuring properties so that interpretations can be made.

Scientific Methods

The overall approach to research of any variety is what is generally termed as the "scientific method". Scientific method comprises of three steps, viz., observation, hypothesis and verification. According to Prof. Northrop, Scientific methods are relative to the stage of enquiry and the type of problem. Different stage of enquiry may have different scientific methods and the method which is scientific for one stage may not be so for another. As there are multiplicities in sciences, there are multiplicities in rational methods. Prof. Caldin holds that there are several types of rational methods. Prof. Hook draw's distinction between scientific technique and scientific method. The former is the instrument available for any research; the latter is the special method of any particular science.

Scientific research requires a logical approach comprised of several independent activities, the essentials of which are the following (1) description of the problem and a critical review of relevant research reports and literature; (2) mustering of previously produced facts by collecting pertinent information about the problem or topic: (3) careful study of available evidence to that research problem can be refined, specific hypotheses or exploratory questions can be posed, and solution can be anticipated: (4) Structuring and conduct of experiments or other careful studies to test the most feasible hypothesis in relation to the most crucial questions: (5) analysis and evaluation of accumulated data and the drawing of relevant conclusions: (6) utilization of research findings to predict effects so that new hypotheses can be generated; and (7) recording of methods, findings, and conclusions in a written research report so that newly acquired insights and knowledge can be communicated to other persons.

The scientific approach has the constituent elements – procedural and personal.

Procedural Components:

Observation, Hypothesis, and Verification are the three procedural approach. Observation is based on the data currently available. Observation depends on knowledge, material, and personal hunches. Observation helps to build up hypothesis. It also helps as a technique of collection of data, there by helping verification. The second step is the formulation of one or more hypotheses. A hypothesis is a tentative conclusion. The main function of hypotheses is to guide the collection and processing of material, and direct investigation. The third step is the verification of hypothesis and arriving at generalization. Verification is helped by analytical tools. The tools of collection and analysis are available

technique used for checking and counter checking. Tools help in the careful definition of the concepts and in the collection of the comprehensive data. Even in social science, the statistical or quantitative tool is coming to greater use.

Personal Components:

The researchers need imagination, analytical ability, resourcefulness, skill, persistence and independence. The capacity to find the heart of the problem is a part of the enquiry. The knowledge of the field of investigation is a fact of the personal component. Researcher's ability and attitude are more important than the method of approach. He should have an objective scientific, and unbiased view. He should have sufficient personal qualifications and professional training, apart from the practical experience, the personal qualifications must enable the researcher sufficiently to assess the adequacy, relevance, and the value of the data. Perusal quality also relates to the integrity, honestly, truthfulness and seniority of purpose. At every step the researcher requires reasoning. There is a great need for balance or poise between mental, physical and moral qualities. Poise gives the ability to see things in their true perspective. Ambition interest and perseverance are very much required to go on successfully with research. In other words a worthy researcher should possess the objectivity of Socrate, the wisdom of Solomon, the courage of David, the strength of Samson, the patience of job, the leadership of Moses, the strategy of Alexander, the tolerance of the carpenter of Nazareth, and above all in an intimate knowledge of every branch of the natural, biological and social sciences.

Scientific Method: Meaning and its essentials

Scientific method is a branch of study which is concerned with observed facts systematically classified, and which includes trustworthy methods for discovery of facts. Every fact is initially nothing but some proposition or problem. At every stage of enquiry hypothesis is necessary. Hypothesis is defined as a suggestion of possible condition between imagined facts and actual facts.

As Wilkinson and Bhandarkar put it, "Science is an intellectual construction – a working thought – model of the world and its aim is to describe and conceptualize the impersonal facts of experience in verifiable terms, as exactly as possible, as simply as possible and as completely and meaningfully as possible".

George Lundberg defines scientific method as one consisting of systematic observation, classification, and interpretation of data. The main difference between day-to-day generalization and the scientific method lies in the degree of formality, rigorousness, verifiability and general validity of the latter.

Scientific method is self corrective in nature. It depends on the methods of developing and testing hypothesis. The method of enquiry itself may be testified and modified. Since scientific method is based on the most probable inference at a particular point of time, a scientific theory is more probable then any other alternative theory. The method of science depends on evidence which is collected from empherical material on the basis of principle. In the process of such dealings, the doubtful matters are detected and clarified.

Assumption of scientific method

The following are some assumptions and objectives of scientific method:

(1) Regularities (2) Verification (3) Techniques (4) Quantification (5) Values(6) Systematization (7) Pure Science and (8) Integration.

(1) Regularities: Scientific method believes that the world is regular and phenomena occur in patterns. Further there are certain discernible uniformities in political and social behaviour, which can be expressed as generalizations which are capable of explaining and predicting political and social phenomena.

(2) Verification: Scientific method prescyoposes that knowledge should consist of propositions that have been subjected to empirical tests and that all evidence must be on observations. Hence scienceis empirical. Experience has come to represent the true test of science. One cannot establish true scientific facts unless one can experience them.

(3) Techniques: Scientific method adopts correct techniques for acquiring and interpreting data. There is need for the use of statistical tools like multi-variate analysis, samples surveys, mathematical models, stimulations, etc. to make the researcher conscious about the methodology. This helps the researcher for planning, executing and assessing his research work. The techniques must be validated for analyzing data.

(4) Quantification: Mathematical formulas and measurements are involved in science. Theoretical concepts must be expressed in mathematical language. Unless they are quantified it would not be possible to obtain precise and accurate knowledge about the complexities of phenomena. Hence it is necessary to quantify all observations as it has advantages in terms of precision and manageability.

(5) Values: Values are facts of two separate things. Social sciences should be studied scientifically through empirical methods and it will have nothing to do with moral and ethical questions.

(6) Systematization: Scientific study demands that research should be systematic. It means it must be theory-oriented and theory-directed. The theory and research should form interrelated parts of a coherent and orderly body of knowledge. It is said that research untutored by theory may prove trivial and theory, unsupported by data, futile. Theory does not consist of speculation and introspection by of analysis, explanation and prediction. It is on the basis of a well organized, logically interrelated structure of concepts and propositions that hypothesis have to be advanced. The hypothesis, in their own turn, has to be capable of undergoing rigorous testing and, then alone, should form the basis of new theories.

(7) Pure science: Scientists insist on pure science approach. They agree that theoretical understandings may lead to an application are parts of problems of life. The theory and its applications are two major parts of the scientific enterprise. Therefore, they attach great importance to pure research and would be content with it even if it cannot be applied to specific social problems.

(8) Integration: Finally, there is integration of each social science with other social sciences. The behaviouralists agree that man is a social animal and while one may try to draw boundary lines between the social, political, economic, cultural and other activities, none of these activities can be understood without placing them in the wider context of his entire life. The study of a political phenomenon, therefore, requires some understanding of how the economic, cultural and other phenomena in society are unfolding themselves. If an effort is made to detach the political man from the economic, social or cultural man, it would not be possible to understand his political

behaviour in its true character. Therefore, a complete and a total perspective comes only through interdisciplinary approach.

Steps in scientific method

The following are the steps involved in applying scientific method:

1. Formulation of hypothesis
2. Defining concepts
3. Establishing working definitions
4. Collection and analysis of data and
5. Relating the findings to the existing knowledge and generalization

Lundberg's four levels of research:

Lundberg has described the following four levels of research.

1. **Random Observation:** At this stage, the investigator has no definite aim or hypothesis. He simply observes or conducts some experiments. Then some day by chance, he may come across some new discovery of relationship. This is a crude method. It involves a lost of fruitless labour. But, it has been found to be very useful in developing a sound basis for intensive research. In physical sciences, some great discoveries like x-ray, penicillin, had taken through this method.
2. **Systematic exploration:** This is an improvement over the first method. The researcher does not leave everything to chance. He proceeds in a systematic manner to explore the variables and their correlation ships.
3. **Testing a well-Defined hypothesis:** Suitable variables are correlated and the nature and degree of their relationships is expressed in the form of a tentative generalization and the research efforts are confined to testing its validity.
4. **Experiment Directed by Systematic Theory:** This is the most refined and advanced method of research. Empirical observations are made and necessary data are collected for interpretation and generalization. Newton's principles and Einesteins Relativity Theory come under these categories. A valid inference is one in which the conclusion follows reasonably from the premises. For ascertaining a valid conclusion, a science should depend on a logical method. Scientific method is therefore the persistent application of logic as the common feature of all systematic and reasoned knowledge.

Formal Logic

Logic is an instrument of reasoning and its character is formal. Logical validity cannot ensure the correctness of the subject matter of science. Logic is more like grammar of which details with the correctness of the form or structure of the language. Similarly logic studies the forms or structure of reasoning and formalities the conditions for the validity of reasoning, but it does not and cannot study the material or correctness or incorrectness of a promise of conclusion, legitimate is formal because it studies the formal structure of reasoning.

Logic and Scientific Methods

Logic involves reasoned language; and all sciences are applied logic. The universal feature of science is in general method which consists in the persistent search for truth. But the search for truth depends on evidence, the determination of which we call logic.

Nature of Scientific Method

The nature of scientific method depends upon the nature and objective of a particular science.There are broadly two methods of sciences:

1. Technical (Technological) and
2. Logical

1. Technical Method: For proper observation and experiment, a science, for its investigation, develops certain technological relations or facts against which the observed facts are considered. Technical methods may simply use certain instruments. The more developed the technical method, the more exact a science becomes in handling data required for experiment. The use of technical methods makes a science progressive.

2. Logical methods: Since science is a systematized knowledge, the importance of the method of reasoning or logic can hardly be over estimated. Logic is the science of reasoning. It formulates conditions through which the validity of reasoning may be tested. Reasoning consists of arriving at a conclusions from certain premises is called informance, which is nothing but divided knowledge. All knowledge consists of aserti or propositions. Inference is a proposition which is derived from some other propositions. A valid inference is one in which the conclusion follows reasonably from the premises. For ascertaining a valid conclusion n, a science should depend on

a logical method. Scientific method is therefore the persistent application of logic as the common feature of all systematic and reasoned knowledge.

Use of Scientific Method and its Value

- Possessing of knowledge becomes a universal human urge. Scientific Method is concerned with the verification of knowledge. It finds out some order in which things are related together. Scientific method is the only way to increase the general body of tested knowledge and to eliminate arbitrary and ambiguous opinion.
- Scientific method is the pursuit of truth as determined by logical considerations. The ideal of science is to achieve this ideal by experimentation, observation, logical arguments and combination of these three in varying proportions.
- In scientific method, logic helps in formulating propositions, explicitly and accurately so that their possible alternatives become clear. Further, logic develops the consequences of such alternatives, and when these are compared with observable phenomena, it becomes possible for the research or scientist to state which alternative is most in harmony with the observed data.
- It minimizes the dangers associated with Novelty, Adventure and uncertainty.
- It relies on empirical evidence.
- It utilizes relevant concepts.
- It is committed to only objective considerations.
- It aims at making only adequate and correct statements about population object.
- It results into probabilistic predictions.
- Its methodology is made known to all concerned.
- It aims at formulating most general axioms.

Difficulties in the Use of Scientific Method in social sciences

Scientific method is more applicable to physical or exact sciences. But it cannot be applied to social science which deals with human behaviour. The people holding this opinion point out the following difficulties in the use of scientific method.

1. As human behaviour is complicated and varied, it is very difficult to categorize human behaviour.

2. When human behaviour is studied and analyzed by other human beings, the personal characteristics of such human beings come into the picture and distort the analytical facts.
3. Human behaviour do not admit measurement because it is psychological in nature.
4. Human behaviour is not uniform and predictable. Therefore, it is uncertain. Different people will have different behaviour in similar circumstances. Similarly, one person may have differently under similar circumstances.
5. The decision making in human behaviour becomes difficult. Thus reliable scientific data cannot always be collected.

Limitations of Scientific Method

Though scientific method has many uses in research, it has got its own limitations.

1. It involves abstractness.
2. Its explanation is never complete.
3. The conclusions arrived by this method is not final.
4. As each science is concerned with a particular area and is based on certain assumptions, it has limited scope.
5. Superstition, cherished beliefs etc., are hostile to the growth of scientific method.
6. Definitions and formal distinctions are not often used properly. Also, statistical information may be irrelevant and inconclusive.
7. Scientific judgment is difficult. Sometimes, it is impossible immediate action is demanded.
8. Development of this method needs necessary time for reflection and material for experiment.
9. Scientific research suits only to the optimists.
10. There is no guarantee of achieving the goal and can prevent human life from being an adventure.

UNIT II

CHAPTER IV

Research Problem

Introduction

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 it 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?

The term problem originates from a Greek word "Proballein" meaning anything thrown forward, a question proposed for solution, a matter stated for examination.

Definition of a Problem

R. S. Woodworth defines problem as "a situation for which we have no ready and successful response by instinct or by previously acquired habit". In other words, it means a situation in which a ready solution is not available. The solution can be found out only after an investigation.

John Dewey writes "the need of clearing up confusion, of straightening out an ambiguity, of overcoming obstacles, of covering the gap between things as they are and as they may be when transformed, is, in germ, a problem".

Selection of a Research Problem

The research problem undertaken for study must be carefully selected. The selection of a research problem is a very important job for a research worker. Before the selection of a research problem, the researcher has to analyze the problem based on the following questions:

1. Is the problem relevant and important?
2. Does the subject area suite to his interest?
3. Does it contain originality and creativeness?
4. Does the problem require extension of knowledge?
5. Is the problem feasible with respect to time and date required in its solution?
6. Does the study reveal some new ideas?
7. Does he study practically fulfill the objectives?
8. Is the problem of investigation too narrow, or too broad?

The two important factors that influence the choice of a research problem are the personal value of the researcher and the social condition. The following are the five steps involved in research.

1. Choice of a topic
2. Data Collection
3. Formation of Hypothesis
4. Verification
5. Writing the thesis

Stages in the selection of a Problem

Broadly, the selection of research problem would involve the following three stages.

Stage 1: Selection of Problem Area

Any problem is not of significance to the nation of profession is definitely not worth consideration of the investigator. A research problem can be sponsored by an agency and can also be invented by the investigator himself. It is helpful for the investigator to keep in mind the following aspects while selecting the area of problem for research.

1. The problem to be chosen should be such so as to be meaningfully related to the interest of the investigator himself.
2. The problem having alliance with the chain of thinking of research already in existence can be handled more confidently.
3. The ambitious problems covering a wide range of area of interest should be avoided and the problems of manageable size and limits should be taken up.
4. An important consideration for selecting the problem area relates to its feasibility in terms of the application of scientific techniques, availability of resources in terms of money, personnel and equipments.

Stage 2: Identification of the Problem

The researcher after having carefully understood the pattern of thinking in a particular area of interest seeks to consider the following aspects for the selection of the problem for study. These aspects can be classified as

1. External Factors

While considering the external factors, the following should be thoroughly explored:

a) Novelty

Novelty is one of the fundamental qualities needed for a research problem. The problem shall be of a new area so that the duplication of work is avoided. To achieve this quality, the researcher has to study the early work on this area. The data collected should be recent. Even a problem that has been investigated previously is amenable to further studies using newer and better devices, tools and procedures.

b) Significance

The research problem should be flexible enough to add the existing knowledge or to improve the current practices. It should change the previous findings in any way. The new findings should have some significant role to play in future investigations. If this is not possible, the efforts made will be in vain.

c) Availability of other sources

The facilities in respect of money, time and materials are needed for a research. In case of a good research problem, the following factors should be solved favourably:

1. financial commitment of the problem
2. time required to complete the research work
3. availability of necessary laboratory to carry out the research work

d) Techniques to be employed

e) Sponsorship

f) Working Conditions.

2. Internal Factors

In case of internal factors, the necessary consideration has to be given with respect to the following:

a) Interest

The problem should be interesting to the investigator. During the course of the investigation, the investigator has to face certain obstacles. To face these obstacles boldly, the investigator should have a genuine interest in the problem, otherwise the tendency to discontinue the research may develop. Hence intellectual interest must be attributed as one of the criteria for the successful completion of the research work.

b) Amenability

The problem chosen for research should be amenable to investigation. The researcher should possess the required competence, knowledge, and understanding. He should be skillful enough to develop, administer, and interpret the necessary data gathering devices and procedures. He should possess a reasonable basic knowledge in the necessary statistical techniques. The problem should match his special qualifications, training and experience.

c) Availability of data

After selecting a particular problem for research, the collection of data related to the topic may be sometimes very difficult. In certain cases, the data may be confidential or sensitive. So, for a successful completion of a research work, easy availability of data is a prerequisite. Thus, the topic of the research should ensure the easy collection of data.

d) Availability of Co-operation

The co-operation of various agencies like institution, authorities, and individuals is needed for research work. During the course of investigation, the researcher has to administer several men and materials and to conduct prolonged experiments. He must make sure that the necessary permission and co-operation from authorities will be available when required.

e) Availability of guidance

A researcher has to choose the problem in such a way that necessary guidance is available for the research. Unless there is a competent qualified faculty member who would be willing to supervise the research work, research cannot be carried out successfully.

f) Level of research

It is another criterion to help in the selection of a problem. The nature and scope of a study will be determined on the light of levels like Master's degree, M.Phil., degree and Ph.D. In the higher levels like Ph.D, the problem chosen for research should be more complex and need not be so simple.

g) Intellectual Curiosity

h) Training

i) Temperament and Personal Characteristics

j) Cost involved

k) Risks

l) **Timings**

m) **Motivation**

Stage 3: Interpretation of the Problem (Analysis of the Problem)

The introductory explanation of the problem is usually followed by a detailed definition and development of back ground concerning sub-problems, scope, the review of the related literature, sources of data, explanation of terminology used and assumptions etc. To analyze the problem in its proper perspective he would ask five simple questions to ensure its feasibility.

They are:

1. What do you want to know?
2. Where and how will you get the information?
3. Who will collect the information?
4. How will the information be analyzed?
5. What does it mean?

Scope and limitation

Whatever be the topic chosen for a research work, the scope of the topic has to be limited with reference to the availability of time, money, materials and space. The problem has to be investigated with reference to particular period. The period should be specified. Researcher shall concentrate his study on this particular period only. If the topic is related to a geographical area, that area should be clearly defined, ie., usually the coverage will be that of country, state, district, taluk, village, or in certain cases that of a single institution. Availability of materials can also be pointed out as one of the limitations of the study. But in science and technology field, the investigation always concentrates on a narrow area. And the investigation limits his experiments, on his theoretical pursuit, as the case may be, on a specific problem.

Common Errors

There are some common errors committed in selecting and formulating a research problem. They are:

1. Naming a broad field or area of study instead of a specific problem.
2. Stating it in such a way that the investigation I impossible.
3. Narrowing or localizing a topic.
4. Including in it terms of an unscientific, emotional or biased nature.

5. Lack of precision in the instruments.

Conclusion

A clear statement of the research problem and identification of the limitations of the study are essential for the successful completion of the research.

CHAPTER V

Research Design

Introduction

Having developed research question, identified a reading list, and planned the outline for the literature review, we are now going to move on to look at quantitative research design. What the researcher should keep in my mind is that there are two key, current applications for developing his knowledge about quantitative research design:

(i) to look at how the research question can be examined through a variety of different designs;

(ii) to understand and describe the research designs used in the empirical literature which he is reading for the literature review.

Research design provides the glue that holds the research project together. A design is used to structure the research, to show how all of the major parts of the research project -- the samples or groups, measures, treatments or programs, and methods of assignment -- work together to try to address the central research questions.

Research designs fall into two broad classes: quasi-experimental and experimental. Experimental studies are characterized by the ability to randomize subjects into treatment and control groups. This randomization goes a long way toward controlling for variables which are not included explicitly in the study. Because comparison groups are not true, randomized control groups in quasi-experimental studies, this type of study has to control for confounding variables explicitly through statistical techniques. For this reason, quasi-experimental studies are sometimes labeled *correlational designs*.

Key Concepts and Terms

1. Experimental Designs

A design is experimental if subjects are randomly assigned to treatment groups and to control (comparison) groups. Cook and Campbell (1979) mention ten types of experimental design.

Note that the control group may receive no treatment, or it may be a group receiving a standard treatment (ex., students receiving computer-supported classes versus those receiving conventional instruction). That is, the control group is not necessarily one to be labeled "no treatment."

a) **Classic experimental designs**: randomization of subjects into control and treatment groups is a classic experimental method, amenable to a variety of ANOVA designs. The two broad classes of classic experimental design are:

b) **Between subjects designs:** In this type of design, the researcher is comparing between subjects who experience different treatments. There are different subjects for each level of the independent variable(s) (ex., for each different type of media exposure in a study of the effect of political advertising). Any given subject is exposed to only one level and comparisons are made between subjects' reactions or effects. The researcher relies on randomization of subjects among the treatment groups to control for unmeasured variables, though sometimes stratification of subjects is employed to guarantee proportions on certain key variables (ex., race).

c) **Factorial designs: This design** uses categorical independent variables to establish groups. For instance in a two factor design, the independent variables might be information type (fiction, non-fiction) and media type (television, print, Internet), generating 2 times 3 = 6 categories. An equal number of subjects would be assigned randomly to each of the six possible groups (ex., to the fiction-television group). One might then measure subjects on information retention. A null outcome would be indicated by the average retention score being the same for all six groups of the factorial design. Unequal mean retention scores would indicate a main effect of information type or media type, and/or an interaction effect of both.

d) **Fully-crossed vs. incomplete factorial designs**: A design is fully crossed if there is a study group for every possible combination of factors (independent variables). An incomplete factorial design, leaving out some of the groups, may be preferred if some combinations of values of factors are non-sensical or of no theoretical interest. Also, when one of the factors is treatment vs. control (no treatment) and another factor is

types/levels of treatment, the control subjects by definition will not receive types/levels of treatment so those cells in the factorial design remain empty.

e) **Randomized block designs**: These stratify the subjects and for each strata, a factorial design is run. This is typically done when the researcher is aware of nuisance factors that need to be controlled (example, there might be an air conditioned room stratum and a no air conditioning stratum) or if there were other mitigating structural factors known in advance (ex., strata might be different cities). That is, the blocking variables which stratify the sample are factors which are considered to be control variables, not independent variables as they would be in a simple factorial design. Randomized block designs seek to control for the effects of main factors and their interactions, controlling for the blocking variable(s).

f) **In SPSS:** Consider city to be the blocking variable and information type and media type to be the main factors. In a simple factorial design, city would be an additional factor and in SPSS one would ask for Analyze, General Linear Model, Univariate; the dependent variable would be retention score; city, information type, and media type would be fixed factors; the model would be "full factorial" (the default). In a randomized block design, one would ask for Analyze, General Linear Model, Univariate; the dependent variable would be retention score; information type, and media type would be fixed factors; the blocking variable, city, would be entered as a random factor; click Model and select Custom, then set "Build Term(s)" to "Main Effects" and move all three factors over to the "Model:" box; uncheck "Include Intercept in Model."; Continue; OK. Note that this procedure reflects the fact that in a randomized block design there are no interaction effects, just main effects. Later, for multiple comparisons, repeat this procedure but click the Post Hoc button and enter the main factors in the Post Hoc Tests box; also check the type of test wanted (ex., Tukey's HSD).

g) **Within subjects (repeated measures) designs:** In this type of design, the researcher is comparing measures for the same subjects (hence, "within subjects"). The same subjects are used for each level of the independent variable, as in before-after studies or panel studies. Since the subjects are the same for all levels of the independent variable(s), they are their own controls (that is, subject variables are controlled). However, there is greater danger to validity in the form of carryover effects due to exposure to earlier levels in the

treatment sequence (ex., practice, fatigue, attention) and there is danger of attrition in the sample. *Counterbalancing* is a common strategy to address carryover effects: ex., half the subjects get treatment A first, then B, while the other half get B first, then A, so that the carryover effect washes out in the sense that it is counterbalanced in the overall sample. Keep in mind that counterbalancing does not remove all effects - for instance, if there is a practice effect in a test situation, with higher scores for the second-taken test, on the average both tests will score higher in the overall sample than they would otherwise, since for both tests half the sample had the benefit of a practice effect. Counterbalancing in this situation only seeks that both test scores are biased equally upward, not that bias in absolute scores is eliminated.

h) **Matched pairs designs:** Compared to between-subjects designs, within-subjects designs control for subject variables better but at the expense of greater threat to validity in the form of contamination from influences arising from subjects going from one experimental level (condition) to another. Another type of repeated measures design is matched pairs, where the repeated measurement is not of the same subjects but of very similar subjects matched to have like key attributes. While matched pairs designs avoid some types of invalidity of within subjects designs, such as the threat of subject fatigue across repeated tests, matched pairs designs control only for the matched attributes whereas same-subject within-subjects designs control for both explicit and unmeasured subject variables.

2.Quasi-Experimental Design

a) Nonequivalent Control Group Designs

A design is quasi-experimental if subjects are not randomly assigned to groups but statistical controls are used instead. There may still be a control or comparison group. While subjects are not randomly *assigned*, they are either randomly *selected* (sampled) or are all the relevant cases. For instance, a random sample of cities with council–manager governments may be compared with a random sample of cities with mayor–council governments. Cook and Campbell (1979) outline 11 nonequivalent control group research designs. In each case, due to the non-equivalency of the comparison group, threats to validity are much more possible than in a randomized design and the researcher should consider checklist-style all the types of validity threats.

1. One-Group Posttest-Only Design: Sometimes called the "one-shot case study," this design lacks a pretest baseline or a comparison group, making it impossible to come to valid conclusions about a treatment effect because only posttest information is available. The level of the dependent variable may be due to treatment, or may be due to any number of causes of invalidity such as history (other events coexisting with treatment), maturation (changes in subjects which would have occurred anyway), experimenter expectation (subjects seeking to provide responses known to be desired or simply reacting to the attention of being tested), or other biases discussed in the section on validity.If this design is used, information must be gathered on pretest conditions, if only through respondent recollections, which are often subjective and unreliable.

2. Posttest-Only Design with Nonequivalent Comparison Groups Design: In this common social science design, it is also impossible to come to valid conclusions about treatment effect based solely on posttest information on two nonequivalent groups since effects may be due to treatment or to nonequivalence between the groups. Strategies for improving validity center on trying to create equivalency between groups by random assignment of subjects or matched-pair assignment to groups. When such assignment is impossible, then attempts may be made to control statistically by measuring and using as covariates all variables thought to affect the dependent variable. Nonetheless, many of the same threats to validity exist as in one-group posttest-only designs: history (concurrent events affect the two groups differently), maturation (the two groups would have evolved differently anyway), testing (the two groups have different reactions to testing itself), regression to the mean (the two groups tend to revert to their respective means if starting from extreme levels), etc.

3. Posttest-Only Design with Predicted Higher-Order Interactions: Sometimes the expectation of the treatment effect interacts with a third variable. Instead of the expectation that treatment group subjects will be higher on the dependent, one has the expectation that the subjects will be higher if in the upper half of third variable Y but lower (or not as high) if in the bottom half of Y. For instance, training may lead to greater productivity for high education employees but not for low education employees on the same tasks. The interaction creates two or more expectations compared to the simple one-expectation one-group posttest only design. Because there are more expectations, there is greater verification of the treatment effect.

However, this design is still subject to possible challenges to validity due to such factors as history (subjects high in education had different experiences) -- it is just that the counter-argument has to be more complex to account for the interaction, and therefore may be somewhat less likely to be credible.

4. One-Group Pretest-Posttest Design: This is a common but flawed design in social science. It is subject to such threats to validity as history (events intervening between pretest and posttest), maturation (changes in the subjects that would have occurred anyway), regression toward the mean (the tendency of extremes to revert toward averages), testing (the learning effect on the posttest of having taken the pretest), and most challenges discussed in the separate section on validity. Sometimes the pretest data is collected at the same time as the posttest data, as when the researcher asks for recollection data of the "before" state. This is know as a *proxy pretest-posttest design* and has additional validity problems since the pretest data are usually significantly less reliable.

5. Two-Group Pretest-Posttest Design Using an Untreated Control Group (separate pretest-posttest samples design): If a comparison group which does not receive treatment is added to what otherwise would be a one-group pretest-posttest design, threats to validity are greatly reduced. This is the classic experimental design. Since the groups are not equivalent, there is still the possibility of selection (observed changes are due to selection of subjects, such as working with more motivated volunteers in a treatment group -- see two-stage least squares for a discussion of testing for selection bias). Much depends on the outcome. For instance, if the treatment group starts below the comparison group and ends up above after treatment, a stronger inference of a treatment effect exists than if both groups rise in performance, but the treatment group more so (this might well be due to selection). A strongly recommended modification to this design is to have more than one pre-test. Multiple pretests (at the same interval as between the last pretest and the posttest) help establish the performance trends in both the treatment group and the control group, and treatment should be revealed by a change in the trend line for the treatment group but not the control group.

6. Double pretest designs. One can strengthen pretest-posttest designs by having two (or more) pretest measures. This can establish if there is a trend in the data independent of the

treatment effect measured by the posttest. By seeing if there is a posttest effect over and above the trend, one controls for maturation threats to study validity.

7. Four-group Design with Pretest-Posttest and Posttest-Only Groups. Also known as the "Solomon four-group design," this design has a treatment and control group with both pretests and post-tests and has treatment and control groups with posttests only. This design strengthens the two-group pretest-posttest design because, if the same effect difference is found for treatment vs. control groups in the pretest-posttest set as for the posttest-only set, then the researcher may rule out threats to validity having to do with repeated measurement (ex., learning effects from having taken the test before).

8. Nonequivalent Dependent Variables Pretest-Posttest Design: In this design, the researcher identifies dependent variables related to the treatment-related variable, but where treatment is predicted to have no effect. Then, if the variable thought to be affected by treatment does in fact change in the predicted direction, but there is no change in the other related dependent variables, again as predicted, then the inference is made that the change in question is due to treatment, not some confounding cause such as test experience from the pretest.

9. Removed-Treatment Pretest-Posttest Design: In some situations it is possible not only to introduce a treatment but also to remove it. If the dependent variable goes up after treatment and then goes down when treatment is removed, this is some evidence for the effect of treatment. Of course, if the variable goes up after treatment, it might come down on its own anyway due to a declining return or attrition effect. Cook and Campbell (1979) therefore recommend at least two posttests after treatment and before removal of treatment, in order to establish trend effects after treatment. The researcher also needs to beware of resentment effects due to treatment removal, as these also might cause a decline in the variable measured, depending on the situation.

10. Repeated-Treatment Design: This design is similar to the preceding one but follows a pretest-treatment-posttest-removal of treatment-posttest-restoration of treatment-posttest pattern. The expected treatment effect is for the dependent variable to increase after treatment,

decline after removal of treatment, then increase again with restoration of treatment. Even if this outcome occurs, inference is not foolproof as the decline phase may be due to resentment at removal of treatment rather than direct adverse affects of removal of treatment, and the subsequent rise may be due not to restoration of treatment but removal of the source of resentment. Also, subjects may more easily become aware of experimenter expectations in this design, and may seek to meet (or react against) expectations, thereby contaminating the study.

11. Switching Replications Designs. In this research design, there are two comparison groups and three measures. Both groups are measured under pretest conditions. The treatment is given to one group but not the control group, and a first post-test measure taken. Then the treatment is given to the control group but not the first group, and a second post-test measure is taken.

12. Reversed-Treatment Pretest-Posttest Nonequivalent Comparison Groups Design. This design is one in which the nonequivalent comparison group receives the opposite treatment (ex., the treatment group receives participative leadership while the comparison group receives autocratic leadership). The expectation is that the posttest will show increase for the treatment group and decrease for the comparison group. Cook and Campbell (1979) suggest adding a no-treatment group and even a placebo group where appropriate. Multiple pretests will improve this design by showing preexisting trends in the treatment and nonequivalent comparison group.

13. Cohort Designs with Cyclical Turnover: This design refers to the study of groups as they evolve over time, as in the study of a fourth-grade class in year 1, the corresponding fifth grade class in year two, etc. The expectation is that the class average will increase in the posttest after treatment. This design is liable to the same challenges to validity as simple prettest-posttest designs, but it can be strengthened by *partitioning* the cohort into subgroups according to their exposure to the treatment. In a study of the effects of television violence, for instance, the cohort may be divided into groups of high, medium, and low exposure to violent television shows. The expectation is that the partitions exposed more will show more change on the dependent variable. Where partitioning is not possible, having multiple prettests and posttests can establish trends to rebut "it would have happened anyway" arguments about the validity of conclusions under this design.

14. Regression-Discontinuity Design: One might hypothesize that if there is a treatment effect, then the slope of the regression line relating scores before and after treatment would be the same, but there would be a discontinuous jump in magnitude on the dependent variable immediately after treatment. This test requires verification that the relationship between prettest and posttest scores is linear, as two linear regressions (one before, one after treatment) on a curvilinear underlying relationship could spuriously appear to meet this test. Also, there may be a treatment effect taking the form of a steeper regression slope but no discontinuity at the point of treatment. Such a treatment effect is very difficult to differentiate from a simple curvilinear relationship.

15. Regression Point Displacement Design. In this design there is a treatment group (ex., a county) and a large number of comparison groups (ex., other counties in the state). Consider the case where we wish to estimate the effect of an after-school intervention on juvenile crime. In the pretest condition we regress juvenile crime rates on, say, median income level and we note the position of the test county in the regression scattergram. In the posttest condition, after we have implemented the intervention program, we re-run the regression. If the location of the test county is displaced on the regression scattergram, we conclude that the intervention had an effect.

3. Interrupted Time Series

Above, in the discussion of non-equivalent control group designs, it was suggested that pretest-posttest versions could be improved by having at least two pretests to establish linear tendencies apart from treatment. Cook and Campbell (1979) list six interrupted time series designs which extend this suggestion by having multiple pretests and posttests.

a. **Simple Interrupted Time Series Design**. This is the one-group pretest-posttest design augmented with multiple pretests and posttests. The trend found in multiple pretests can be compared to the trend found in multiple posttests to assess whether apparent post-treatment improvement may simply be an extrapolation of a maturation effect which was leading toward improvement anyway. Since there is no control group, however, the researcher cannot assess other confounding factors such as history-type challenges to validity (the possibility that other factors historically coterminous with the treatment

actually led to the observed effect). There may be other problems such as failure to seasonally adjust data, confounding a seasonal effect with a treatment effect; selection bias, as due to non-random attrition of subjects in the posttest; instrumentation bias (the posttest is not equivalent to the pretest); and testing (there may be a learning effect from the pretest such that the observed effect is one a test artifact rather than a treatment effect).

b. **Interrupted Time Series with a Nonequivalent No-Treatment Comparison Group** : This is the two-group pretest-posttest design using an untreated control group, but with multiple pretests and posttests. By having a comparison group, even if nonequivalent (not randomized), the same threats to validity can occur, but they usually occur in a more complex and hence more easily disproved way. For instance, if this design shows an improvement in the treatment but not comparison group, it may still be true that there is historical bias, but such biasing history factors must be unique to the treatment group for some reason not experienced by the comparison group. There could be seasonal bias, but only if the seasonal factors were thought to be uniquely associated with treatment. The researcher's main validity challenge is to show the two groups were equivalent on all causally important variables prior to treatmen (ex., in a study of a rehabilitation program's effect on recidivism, to show the two groups of prisoners were similar in crime record, age, etc.).. An alternative strategy, which only works for stronger effects, is to select a treatment group which would be expected to measure worse on posttreatment (ex., prisoners with worse criminal records than the comparison group, in a study of recidivism), on the theory that if the effect shows in spite of an adverse starting point for the treatment group, the treatment has an effect.

c. **Interrupted Time Series with Nonequivalent Dependent Variables:** This is the nonequivalent dependent variables pretest-posttest design with multiple pretests and posttests. The object is to find dependent variables related to the dependent being studied, but where the related variables are not thought to be correlated with the treatment variable. Cook and Campbell (1979) give the example of influence on accident rates (the dependent) of breathalyzer tests (the treatment variable) given by police when bars are open weekend nights, but not given at other times. The dependent variable of interest is accident rates on weekend nights. The related dependents are

accident rates on weekday nights when bars are open and accident rates at times when bars are not open. The expectation was that accident rates would be significantly lower on weekend nights because of the presence of the treatment. Counter-explanations for lower accident rates (ex., safer cars, and stricter court treatment of offenders) must explain not only the lower accident rate on weekend nights, but also the lack of effect at other times. Of course, confounding factors may well exist, but they must be unique to the dependent variable of interest.

d. **Interrupted Time Series with Removed Treatment:** This is the removed-treatment pretest-posttest design with multiple pretests and posttests, including ones in between the original treatment and its removal, and hence is a more powerful test. For instance, the threat of history is reduced because any historical forces coincident with treatment would also have increase after treatment and decrease after removal, an unlikely circumstance. Ideally removal of treatment does not occur until enough observations have been taken to rule out any seasonal or other cyclical effects.

e. **Interrupted Time Series with Multiple Replications.** This is simply the interrupted time series with removed treatment design, except that treatment and removal occur multiple times on a schedule. Circumstances rarely permit such a design, but it is stronger yet. By timing the replications randomly, the researcher is able to minimize contamination from cyclical factors. This design assumes one is dealing with a treatment effect which dissipates in a timely manner before the next replication, without carryover effects (otherwise there is "multiple treatment interference," meaning that receiving earlier treatments adds to or multiplies the effect of receiving later treatments).

f. **Interrupted Time Series with Switching Replications.** This is a further refinement in which there are two groups, each serving as either the treatment or comparison group on an alternating basis, through multiple replications of treatment and removal.This requires an even higher level of control over subjects by the researcher but is a particularly strong design in ruling out threats to validity. It does not lend itself to studies where the treatment intervention has been gradual, or where treatment effect does not decay well.

ii. Non-Experimental Designs

A design is non-experimental if the subjects are neither randomly assigned nor randomly selected. There may still be comparison groups.

1. Case study designs are discussed in a separate section.
2. Content analysis is discussed in a separate section. Under some circumstances, content analysis may also be part of a quasi-experimental design.
3. Ethnography is discussed in a separate section.
4. Focus groups are discussed in a separate section.
5. Narrative analysis is discussed in a separate section.
6. Network analysis and sociometry are discussed in a separate section.
7. Participant observation is discussed in a separate section.

Assumptions

i. The researcher is assumed to have a research design!

ii. The researcher is assumed to have considered all threats to validity associated with the design.

iii. In the case of experimental designs, it is assumed that randomization of subjects controls for all unmeasured variables. However, the smaller the sample size, the less likely this is to be true.

Hypothesis

Introduction

Ordinarily, when one talks about hypothesis, one simply means a mere assumption or some other supposition to be proved or disproved. But for a researcher, hypothesis is a formal question that he intends to resolve. Thus, a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of

phenomena either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts.

In a problem-oriented research, it is necessary to formulate a hypothesis or hypotheses. In such researches, hypotheses are generally concerned with the causes of a certain phenomenon or a relationship between two or more variables under investigation.

Definition of Hypothesis

"A hypothesis is shrewd guess or inference that is formulated and provisionally adopted to explain observed facts or conditions and to guide in further investigation".

-Good and Seates.

"A statement capable of being tested and theory by verified or rejected". ***– Rummel.***

"It has been defined as a tentative solution posed on cursory observation of known and available data and adopted provisionally to explain certain events and to guide in the investigation of others. It is in fact, a possible solution to the problems". ***– M.H.Gopal***

Function of Hypothesis

From the above definitions, it is inferred that the following are the three main functions of hypothesis:

1. To Test Theories
2. To suggest Theories
3. To describe Social Phenomena.

1. To Test Theories

The relation between theory and hypothesis is very close indeed. Theories are based on facts, which is if elaborated becomes hypothesis. Theories need to be tested in order to demonstrate its predictive value and its adequacy as a tool of explanation for some

event. The theories lend themselves for empirical test if the facts proposition and assumptions implied in their or split up into specific hypothesis.

2. To Suggest Theories

Another function of hypothesis is to suggest theories. Theory is derived out of hypothesis and is considered as elaborate hypothesis. In other words, hypothesis when refined becomes theory.

There is no sharp line of demarcation between hypothesis and theory. Since the basic difference is one of the complexity and the extent of testing against the evidence. In its early stage of testing a theory usually has been called a hypothesis, but as the hypothesis is checked against the data and their logical implication towards a successful conclusion, it may become known as a theory. Goode and Hart says that "Every worth while theory permits the formulation of additional hypothesis. These when tested are either proved or disproved and in turn constitution further tests of the original theory".

3. To Describe Social Phenomena

When a hypothesis is tested, it explains the phenomenon associated with it. This phenomenon may be new or not known earlier. Neither a theory nor a hypothesis is involved in such a test. Any commonsense knowledge or belief is put to empirical test to determine its validity. Hence new knowledge is gathered. Goode and Hatt observe that "a hypothesis looks forward…. It may be proved to be correct or incorrect on any event, the hypothesis is a question put in such a way that answers some kind can be forthcoming".

Forms of Hypothesis

1. Hypothesis concerning law: This explains how an agent works to produce a particular effect or event.

2. Hypothesis concerning an agent : The law of operation of an agent is known, but the agent which is working to an effect may not be known. This hypothesis is framed to find out agent.
3. Hypothesis concerning collocation: Collocation means an arrangement of circumstances. When a hypothesis is made relating to the circumstances necessary to produce a phenomenon it is known as a hypothesis regarding collocation.
4. Describe hypothesis: It describes the cause and effect relationship of phenomenon.
5. Explanatory hypothesis: It explains the happening of a phenomenon. It reconstructs the situation by extrapolation and arrangement s of facts.

Null Hypothesis

Null means zero. When a hypothesis is stated negatively, it is called Null Hypothesis. The objective of this hypothesis is to avoid the personal bias of the investigator in the matter of collection of data. A null hypothesis is used to collect additional support for the known hypothesis.

Reasoning for possible rejection of proposition is called "Null Hypothesis". It is a very useful tool in testing the significance of difference. H.E. Garret remarks "The Null Hypothesis is akin to the logical principle that a man is incorrect until he proved guilty. It constitutes a challenge and the function of a research is to give facts a chance to refuse this challenge". A null hypothesis in its other forms asserts that the results found to the expected on a probability basis in terms of certain theory. As against null hypothesis, the alternative hypothesis is formulated embracing whole range of values rather than single point. For example, an alternative hypothesis may be stated that:

H_1 = The males visit cinema often than the females.

a null hypothesis may be stated as

H_0 = The males and females do not differ in respect of the frequency of seeingcinema.

The null hypothesis is more useful than other hypothesis because it is exact. It is easier to disprove the category of the hypothesis than to prove it with certainty.

Importance of Hypothesis

In addition to putting the theory to test, a hypothesis performs certain other functions:

1. guiding point to a research
2. prevents a blind search and indiscriminate gathering of data
3. serves as a forerunner to collect data
4. saves a lot of time and keep the researcher away from the considerable amount of confusion
5. solutions to problems
6. has practical values to society
7. helps in understanding the social phenomenon in their proper perspective and refuse certain "Common Sense" Nations about human behavior.

Characteristics of Hypothesis

1. Hypothesis must be defined clearly
2. Hypotheses should have empherical reference
3. The hypothesis must be specific
4. Hypothesis should be related to available techniques
5. The hypothesis should be related to a body of theory.

Testing of Hypotheses

Knowledge or a fact could be accepted only when it has validity. The validity of such knowledge could be accepted only when it is tested with regard to its usefulness or truth. Hence, to accept hypotheses, which are merely hunches or guesses, as facts, they are to be used. Testing hypotheses means subjecting them to some sort of empherical scrutiny to determine if they are supported or refuted by what the researcher observes.

Robert Baes states that the following questions should be asked about hypotheses before they are tested:

1. are the terms empherically specific, so that the concepts or variables can be distinguished in concrete situation?
2. is the relationship between variables such that it could be verified or nullified by means of empherical operations?
3. is there any prior evidence as to the truth or falseness of the relationship?
4. can an appropriate study design be devised?
5. are the variables context bound or could they be equally well applied to other interaction situation?
6. are the generalizations " culture bound" or can they be also applied realistically to other cultures?
7. if other relevant factors are subject to change in the course of the observations, are they adequately specified and enumerated, so that the observer can ascertain whether they have changed during the period of observation?.
8. is the generalization a part of the theoretical system from which it could be deduced as well as being verified by the proposed empherical induction?
9. is the empherical system that is constructed sufficiently praise and articulates to permit predictions in concrete situations?

Problems in Formulating Hypotheses

A researcher has to come across several problems in formulating hypothesis. It is as difficult as finding out a research problem or identifying a suitable theory. The following are some of the difficulties in formulating hypotheses as pointed out by Good and Hart in their book" Methods of social research".

1. Absence of a theoretical frame work
2. Lack of ability to utilize that theoretical frame work logically
3. Failure to acquaint with available research techniques so as to able to phrase the hypothesis properly

The researcher must bear in mind that he has to formulate a good definite and testable hypothesis. At the beginning stage when one is not familiar with various processes of research, the formulation of a sound hypothesis would be some what difficult. He may frame some formulations and can say that they are hypotheses. But they may not have the testability.

Uses of Hypothesis

a) Hypothesis forms the starting point of investigation
b) Hypothesis makes observation and experiment possible
c) Hypothesis is an aid to explanation
d) Hypothesis makes deduction possible
e) Hypothesis act as a guide. In other words, it is investigator's eye-a sort of guiding light in the world of research
f) It prevents blind research. It spells out the difference between precision and hapazards, between fruitful and fruitless research.
g) It provides direction to research identifying which is relevant and prevent irrelevant review of literature and collection of useless or excess data
h) It focuses research, without it is like a random and aimless path.
i) It links up related facts and information in fully understandable.
j) It serves a frame work for drawing meaningful conclusions

UNIT III

CHAPTER VI

Historical Research

Historical research is the critical investigation of events, developments and experiences of past, the careful weighting of evidence of the validity of sources of information on the past, and the interpretation of the weighted evidence. The historical investigator, like others, collects the data, evaluates it for its validity and finally interprets the data. It is also known as historiography and differs from other methods by its nature of subject matter which is usually based on past events. According to Sheik Ali, "Historical Research is digging into the past in order to re-enact the past in its entirely, to reconstruct the past events as fully as they must have

happened and to explain the meaning and significance of those events, to correct the wrong notions so long prevalent if any and to elaborate, analysis, synthesis and philosophise the ideas in the light of the knowledge we possess". In short, the critical examination of the past event or happening in order to know truth and later on to generalization is known as Historical Research.

Aim of Historical Method

The aim of historical method is to discover regularities and events of a given kind at a certain period of time in the past. History is concerned with an accurate account of particular happening, stressing the location and the date.

Thus the aim of historical method is to present an accurate record of how, when and where of the events. The historian explains the events by describing the conditions which led up to it and how it grew. In addition, a historian may indicate some of the important results which have followed as a consequence of the event.

Historical method produces uncritically and unreflectively a final reconstructed historical narrative to learn its bearing upon present phenomena, social institutions, social and or behavioral problems.

Uses Of Historical Methods

The uses of historical methods are enumerated as follows:

i) Historical methods are most frequently used by evaluators of programmes, surveyors and even by the most advanced social or behavioural research workers, both for designing the study and making inferences.

ii) Historical methods provide an interesting material and data in reconstructing the past in order to learn it its influence on present social problem and behavioural complex.

iii) It provides a basis for the formulation, developing awnd initiating projects, programmes and policies.

iv) It extends cludes which might indicate the sources and origin of the problem, the degree of influence it has been exerting on the behavioural aspects and thus shows the way in controlling and harmonizing the sources and influences.

v) It is used as a comparativedata to test the hypothesis.

vi) It is used as "crucial tests" of certain theories. It is an attempt to make a systematic enquiry or investigation into a subject in order to discover facts and revise the known facts and put the facts into theories. It is useful both for theoretical and practical purposes.

vii) It has made an important contribution to various branches of social sciences. Ex-historical methods include study of life history, history of various organizations and institutions.

viii) Historical research enables us to get an accurate knowledge of the past. When we analyse the past and present it is possible to predict the future. Through historical study, it is possible to make comparative studies for a historical period of time factors, situations etc,. In economics, comparative studies may be made in terms of growth, national products. Savings, investments and trends in operation by evaluation. In business, historical studies of individual companies would help us to interpret their contribution to national economy. Thus, historical research enables us to understand the past of plan for the future.

Advantages

The advantages of historical approach are:

a) Some problems are such which can be investigated only by this approach may not offer to other approaches. Therefore, historical approach fills in a big gap of making the research possible and also meaningful on the problems that would otherwise have remained unexplored. Many a times, it is considerable interest to use time series data for assessing the progress or the impact of several policies and initiatives and this can be done by looking into historical records only for such problems, therefore, only historical approach would suit better.

b) Secondly, historical data is not repeatable under any circumstances and, therefore, historical approach serves as a ready hand method to the researchers whose problems depend on historical observations. It is fairly easy to repeat observations in laboratories under controlled conditions but cannot be done in case of historical data. Historical approach, therefore, has an advantage to offer the past data under the then prevailing conditions and afford an opportunity to the researcher to view these observations in the past setting.

c) Thirdly, historical records provide very useful information that goes a long way towards the solution of a research problem. In cases where time series data is unavoidable, it is advantageous to follow historical approach.

How to Conduct a Historical Research

The conduct of historical research entails the following steps:

a) the recognition of a historical problem or the identification of a need for certain historical knowledge.
b) the gathering of as much pertinent information about the problem or topic as possible.
c) if appropriate, the forming of hypotheses that tentatively explain relationships between historical factors (variables).
d) the rigorous collection and organization of evidence, and the verification of the authenticity and varacity of information and its sources.
e) the selection, organization, and analysis of the most pertinent collected evidence, and the drawing of conclusions and
f) the recording of conclusions in a meaningful narrative. Historical research can be conducted most effectively with these procedures.

Library History

Library history is the systematic recounting of past events pertaining to the establishment, maintenance, and utilization of systematically arranged collections of recorded information or knowledge carefully conducted litrary history relates the causes and results of events, it also often recognizes the social, economic, political, intellectual, and cultural environment in which these events occurred. Further more, library history is sometimes considered an exposition of past incidents and developments and their impact on later times.

The term library history is commonly applied to an account of events that affected any library or group of libraries as well as to the social and economic impacts of libraries on their communities.

For example the contents of the Library Trends issue entitled, “American Library History: 1876-1976” illustrate the diversity of historical topics within the profession. Articles are included about the distribution of libraries over the nation, research collection

development, library statistics, library buildings, the library profession, library associations, publishing in the profession, the organization of library resources, and service aspects of librarianship-all are treated from a historical point of view.

Pierce Butler wrote in "An Introduction to Library science that librarianship can be fully appreciated only through an understanding of its historic origins." The specific value of library history, according to Jesse Shera, is that it allows librarians to synthesize and to make generalizations from reconstructions of the past; this process of synthesis and generalization will not only recreate the past but can serve as as aid in understanding the present. According to Shera, librarians should possess a 'clear historical conscious' because they cannot function effectively (fullfil their social responsibility). When history is regarded merely on esoteric aspect of knowledge.

The other examples for historical approach in libraries are the following:

Spenser's "The Chicago Public Library: origins and Backgrounds".

Shera's "Foundations of the Public Library White fields" Boston Public Library etc.

Librarians who have conducted historical studies have sometimes been interested in the general history of libraries or in library history in specific countries, states or regions. Other librarian historians have concentrated on broad periods of time or "ages" – ancient, medieval or modern. A few history scholars have investigated aspects of large categories of librarianship, such as (a) libraries in the United States during the colonial period: (b) the period in which social libraries existed between 1731 and 1865 or (c) the nineteenth century development period of the tax-supported public library in the United States. A number of historians have conducted enquiries into the histories of particular type of libraries such as academic public or special.

Some library historians are investigatory events and developments in the recent part such as (a) the influence of legislators and of legislation such as Public Law 480, the Library services and construction Act, and the Higher Education Act: (b) the library utilization of a variety of new communication media and technological innovations such as tele-fasimile devices, television, microforms, data-processing equipment, computers, reprographic tools, and audio-visual materials; and (c) the impact of the great Depression and world war II on libraries.

Other subjects of contemporary historical studies have been (a) library services to disadvantaged and culturally deprived citizens; (b) the roles of library associations and their impact on library development (c) overseas library technical assistant, (d) the careers of block librarians; (e) women in positions of leadership in the profession; and (f) political leadership for library development.

Conditions Necessary for Historical Research

The research worker who plans to conduct a history study first determine whether a reasonable amount of evidence about the selected topic is readily available, as well as how and where access to it can be gained.

Good history cannot be written without adequate sources of information. The researcher should avoid selecting topics or problems that are too broad a complex. He should often pose research questions that require very extensive investigations.

The investigator should choose a subject area for investigation in which he or she feels most comfortable so far as personal ability, background, experience, knowledge, and aptitude are concerned. Difficulties are likely to arise when historians have not acquired an in depth knowledge about the targeted subject area.

A thorough and accurate system of bibliographical control and note taking procedures should be devised before a study is initiated. Moreover, literature searchers are essential at an early stage of the investigative process.

Good historians realize that not all knowledge is absolute, nor is all knowledge scientific. Sound knowledge is obtained by posing hypothesis, reasoning out possible answers, and testing the answers. The scientific approach in historical inquiries produces knowledge that is more certain.

The success or failure of historical inquiries will depend greatly upon the ability of the research worker to adequately conceptualize the purpose and problem of the research, to vigorously evaluate and categorize the collected evidence, and to analyse data intelligently in view of research objective.

Operational Research

Operational Research (OR) is the use of scientific methods to study the functions or operation's of an organization so as to develop better type ways of planning and controlling, changes. It can be viewed as a branch of management, engineering, or science or as a special

combination of these three. As part of the field at management, its purpose is to assist decision makers in choosing preferred future courses to action by systematically identifying and examining the alternatives which are available to the manager, and by predicting the possible outcomes from such actions. As a branch of engineering, operational research is concerned with the analysis and design of systems for automating and augmenting management type functions, especially those concerned with information processing and control. Operational research is also viewed as a form of applied science, and is closely related to the fields of statistics, computer science, and optimization methods.

Operational research began with the efforts of small teams of British scientists who were mobilized at the beginning of world war II to help solve pressing military problems. Under the direction of Nobel Laureate P.M.S.Blackett, "Blackett's circus" quickly demonstrated the value of interdisciplinary scientific approaches to operational problems. Their success led to early initation by U.S. and other allied groups, and to the eventual recognition of operations research as an integral part of modern military management. The use of Operational Research Techniques in business, industry and government grew rapidly during the 1950s and 1960s. Today, most large companies have special groups which specialize in the application of the OR approach to corporate problems. Many consulting firms provide such services to business, government and military organizations on a contractual basis, usually to the higher levels of decision making in the organization.

Special courses and academic programmes in operations research were started in the early 1950s at Massa Chusetts Institute of Technology, the JohnsHopkingsUniversity, and CaseWestern ReserveUniversity. By the end of the 1960s most Universities were offering so kind of programme in the subject. Preparation for professional practice as operations research specialist has been strongest in the engineering schools, business school, and military academics. Departments of applied mathematics, industrial and systems engineering, and quantitative methods in business and ecomonic are the kinds of faculties that offer advanced graduate study in operations research. Like statistics and computer science, operations research techniques are now being taught widely as supplementary topics in many other disciplines. As an academic subject in its own right, however, emphasis is placed on the mathematical aspects of the subject and on those methods which have proven to be especially useful in describing and solving many kinds of decision problems. In practice, operations

research remains and electric art and is usually undertaken as a team effort combining the skills from many disciplines. Initially such teams were largely dominated by physical scientists who were interested in applying their scientific methodology to operational problems. More recently, their ranks have tended to give way to engineers, economists, psychologists, mathematicians, and computer scientists. The common unifying theme to their work is the task of formulating reals world problems in such a way hat they can be analysed in an systematic fashion and solutions can be proposed that have credibility through logic and experience. Frequently, it is found that formulations and solution techniques perfected in one problem area have considerable transferability to other problem areas. The two major professional societies in the United States are the Operations Research society of America (ORSA) and The Institute of Management Sciences (TIMS).

The Operations Research Approach

Operations research begins with the study of the people who make decisions in an organization and seeks to make more explicit the alternatives they face and their reasons for making chockes. They are four major steps.

a) formulation of the decision problem

b) construction of a model of the decision problem

c) analysis and design of a system for solving the problem, and

d) incorporation of the system in the organization

The total process is a cyclic one, that is, the research team begins with an elementary formulation of the problem and the construction of as simple a model as is possible to capture the essential features. This leads to a solution system that can be used in an experimental way in the organization to see how it affects the original problem situation. This usually leads to a much better clarification of the problem, enrichment of the model, refinement of the system, and a new round of experimentation in the organization. The process is intended to continue in this way to evolve into a fully satisfactory control of the operations under study, and also to lead to higher levels of management decision making in an organization, it is possible to begin at various problem situations and arrive at the same level of understanding and control. However, it is usually most effective to begin with isolated on situations and then work toward their convergence in a more common structure of decision making, rather than to develop a total management system at the outset.

Steps in the Operations Research Process

Formulation of Decision Problems

Operations research has been a major force in defining the role of the modern manager as primarily that of the decision maker. As Miller and Star point out, the manager has to "do" many things in an organization. "But most organizations make continual attempts to relieve a manager of his more or less routine operations so that he will have more time for the critical decision function. He is rewarded and evaluated in terms of this success at making decision." The following decision making operations will serve as a useful model.

1. Choose the objective: Specify its dimensions and value.
2. Isolate all of the variables that are pertinent to the objective value.
3. Develop the relationships that exist among the variable.
4. Distinguish controllable variables from those which are not controllable, and classify the latter as to whether they are random or due to competition strategies.
5. Develop forecasts and predictions of random variables: determine how stable and reliable the forecasts are.
6. Develop the function that relates the variable to the measure of effectiveness in attaining the objective.
7. Identify the restrictions that limit possible values of controllable variables.
8. Choose those values of the controllable variables that will maximize the degree of attainment of the objective within the limits set by the restrictions.

OR Models and Systems Analysis

When decision problems can be formulated in the theoretical manner described above, it is relatively easy to move on to the next steps building analytic models, and designing systems for applying the results of the models. In practice, ofcourse, it is not very easy to formulate problems so neatly, and it is always helpful to do this with a view toward how the formulation will lead to useful models. For this reason, experience with the modeling process is useful in formulating problems. A considerable amount of interaction between the modeling and the problem-formulation steps is needed in order to capture all the essential elements in the problem situation and to formulate them in manner that will make best use of the analytic tools and experience available.

The most widely used OR model is that of linear programming, which is a technique for finding the best allocation of commonly used resources among the various activities that use those resources. A different kind of modeling that has been especially useful to information and communication systems and other types of service organizations is concerned with the analysis of waiting lines or queues. Professor Morse is a noted expert on queueing theory, and he shows that many aspects of the behaviour of users in a library can be readily modeled as random processes. A more recent study by W.B.Rouse used this technique to determine the optimal staffing of service desks in libraries and to predict the total volume of circulation of a library collection.

An alternative method of studying systems which are subjected to random perturbations is by computer simulation, whereby the computer is made to generate such occurrences by mathematical methods and then keep close track of the consequences of such events over a period of time. This method is often used when the characteristics of the system under study do not confirm to the assumptions and other limitations in the use of queueing theory.

Changing Organizational Behaviour

The final step in the operations research process is that of implementation, or the actual use of models and systems in the operation s of an organization. The systems approach is aimed at a totality of understanding that encompasses all factors including the human ones. R.L.Ackoff has pointed out that the real benefits of systems planning in an organization are not derived by "following a plan" but by the organization engaging directly in the planning process itself as n on going activity. Therefore, the role of the professional planner is not to plan, but to provide everyone who can be affected by planning with an opportunity to participate in it, and to provide those engaged in it with information instructions, questions and answers that enables them to plan better than professional planners can alone.

Operations Research in Libraries

The practical motivation on the part of library managers for operations research studies is the pressure to economize and expand activities by means of computers OR is seen as the best way to conduct exploratory studies before making costly commitments for new computerized systems. As Morse says, the need is for logical models of the library that can

be experimented with in order to see what is most useful and what data are needed to make a new synthesis of libraries and computers work most effectively.

Models of Document Usage in Libraries

Operations research models that have been developed within library settings tend to focus on the library as a document storage and retrieval centre. As early as 1953, Morese and his colleagues in the operations Research centre at the Massachusetts Institute of Technology began a series of studies of the institutes library system to develop methods which could be used in other libraries too. Some of these studies centred on the books and some on user's characteristics. The key model in Morse's study is one in which book usage over time can be predicted to follow a known mathematical form, settling down to a steady-state pattern after an initial period of popularity. The model is used to detect some interesting differences in the usage patterns of books in different subject fields, although the data are admittedly skimpy. The model is then used to make policy recommendations for such activities as book acquisition, circulation, retirement, and storage, in the light of the data obtained in the MIT libraries.

Experimental Method

An experiment is a research process used to establish some truth, principle, or effect. It differs from other investigative methods in that the observed phenomena are controlled to varying degrees by the investigator. Most experiments are conducted under known conditions; attempts are made by experimenters to eliminate as many extraneous factors are possible.

A number of experimental procedures can be used by investigator's no single design could be characterized as the best for all inquiries. In selecting particular experimental techniques, competent investigators remain aware that conceptual requirements of the research hypothesis must be met by controlled experimental conditions.

Basic Concepts

An experiment can be defined as a research situation in which investigators specify exactly, or control, the conditions that will prevail in the investigation. The values of one or more independent variables are then manipulated and the effect of the manipulation on the values of the dependent variables with respect to one or more experimental groups is observed. The effects of other factors that might possibly be relevant to the research

problem (i.e., affect the values of the dependent variable) are minimized through careful experimental design. In this way, conceptual requirements of the research hypothesis are met by controlled experimental conditions.

In librarianship experiments can be used to test new techniques for developing, maintaining, and utilizing library collections, to identify ill-defined or previously unobserved library or informational phenomena and to explore conditions under which certain phenomena in library and information science occur.

Subject of an Experimental Research

A subject of an experiment is the basic unit on which an experiment is performed. In agricultural experimentation, subject might be plants of certain types of perhaps plots of planted ground. In drug research, subjects might be white rates or guines pigs. In library science, information, science, and other social science research, subjects are frequently persons for example, patrons' librarians or students.

Treatment In An Experimental Research

A treatment is the condition that is applied to an experimental group of subjects. In agricultural research, a treatment may be the application of fertilizer to growing plants on the regulation of the amount of moisture which the plants receive. In medical research, the administering of drugs to patents (subjects) or the use of a surgical technique are examples of treatments. In librarianship, a treatment that is being tested might be a system of indexing, a mode of instruction, a type of catalogue organization, or a method of book selection, among others.

Control Group In An Experimental Design

Control is central to experimental research and is the feature that distinguishes experimental research from other research methods, such as those in survey and historical research. The experimental research uses a control group and one or more experimental groups. These groups of subjects are deliberately designed to be equivalent in all important respects in order to control the effects of the many variables that might interfere with the test of hypothesis. These effects can thus be predicted to be similar for the control and experimental groups. The effect on the dependent variable of varying the values of one or more independent variables (the treatment can then be observed in the experimental group).

Complete equivalence between experimental and control groups can seldom be obtained in research in the social sciences, however, because with human subjects there are always numerous environmental and hereditary factors that cannot normally be known. Thus, at best, equivalence can only be an approximate concept.

Consider the following hypothesis that high school students who study a unit of instruction in library skills in conjunction with course work in other subjects such as history or English Literature (i.e. in an integrated approach) learn library skills more effectively than high school students who study an equivalent unit independently of other school subjects. To test this hypothesis, there are certain variables that need to be controlled. The two groups of students should be roughly equivalent in levels of motivation, intelligence, socio-economic status, age, and perhaps other qualities as well. It is easy to see why this is so; any one of these variables might have a differential effect on otherwise equivalent students. For example, if the students of Group A manifest more collective intelligence than Group B, then the factor itself may cause students in Group A to learn library skills more effectively than those in Group B. But by designing and experiment so, the groups are roughly equivalent in intelligence, this variable is controlled, and thus the effect should be the same for both groups. Because of the careful control that is exercised in its use, the experimental method is the most appropriate technique to test hypothesis which involve casual relationships.

Types of Experimental Research Methods

Classical Experiment

In the classical experiment, subjects or objects of a study are randomly assigned to one of two groups an experiments (test) group and a control group. Each subject is given an equal chance of being assigned to the two groups. Both groups are treated similarly with the exception of the application of the key treatment to the experimental group entails the manipulation of one or more variables by the investigator. Factors that are manipulated by the experimenter are called independent variables. When independent variables are varied, the experimenter observes other variables to measure the degree to which the variables are related to changes in the independent variables. These observed factors, measured after the

independent variable has been manipulated are known as dependent variables. The classical experiments have been schematically diagramed in Figure 1.

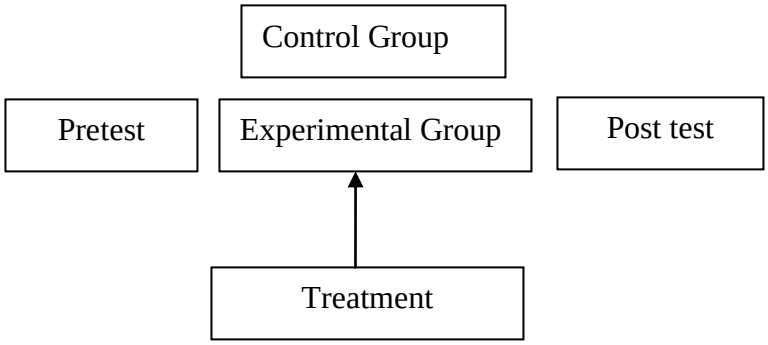

Figure 1. The Classical Experiment

Field Experiment

In a field experiment, the investigator observes a phenomenon in a natural setting, and at the same time, manipulates one or more variables are taken in field experiments to avoid unnecessary disruptions of the natural conditions being observed. When an investigator goes beyond the field experiment and attempts to construct a particular setting or a set of circumstances that typifies a naturally occurring phenomenon, the research design could be described as n experimental simulation. The electronic computer has been found to be a powerful tool for such stimulations.

Laboratory Experiment

A laboratory experiment is a study conducted under highly controlled conditions. The term "laboratory" is used as an indication that an experiment is not conducted in a naturalistic setting but in some other convenient place. When conducting laboratory experiments, the investigator usually attempts to create prototypes of events, situations or processes apart from the naturalistic circumstances in which they occur. Because the settings for most laboratory experiments are artificial or contrived, the investigator must ensure that the research design has a high degree of internal validity.

Significance Of Experimental Method

The experimental method is one of the most useful and powerful techniques for identifying casual relationships between variables. While experiments in library and

information science are conducted to test specific hypotheses, the experimental method is also used advantageously in exploring other kinds of research problems. Experiments can be performed to test new techniques for acquiring, classifying, storing, and relieving information, or to test new library or information services. Further more, experiments allow investigators in library and information science to explore conditions under which a phenomenon occurs. Other experiments might be conducted merely to satisfy curiosity about certain library or information phenomena.

Evaluation Of An Experiment

Once an investigator has conducted an experiment, the results of the experiment the experimental data – must be evaluated Evaluation involves a number of concepts, including hypothesis testing, experimental error, sensitivity, internal validity, and external validity.

The evaluation of an experiment involves a test of a null hypothesis, a statement that no significant difference exists between the control and experimental groups. In other word the null hypothesis asserts that within specified limits of credibility, the control and experimental groups are essentially equivalent. Two groups of subjects will rarely perform identically; some variation will probably occur under circumstances. The investigator, however, must determine. Whether differences between the performances of the control and experimental groups are statistically significant. A test of a null hypothesis is an attempt to disprove the assertion of "no difference"; to show that the experimental groups has in fact been affected by the treatment in such a way as to significantly change the value of the dependent variable. An appropriate procedure to test the significance of a treatment using the four-cell experimental design may be the't'-test, which is applied to the differences between pretest and post test scores for the control and experimental groups. The chi-square test may also be appropriate. Analysis of variance is required for more complex designs.

Sensitivity of an Experiment

The sensitivity of an experiment is its ability to detect relatively small effects. One way to increase the sensitivity of an experiment is to increase the number of subjects; this decreases the chances for random error (or experimental error) to affect the results in a significant way. This option always adds to the expense of conducting the experiment, however. Another way to reduce experimental error is to exert additional control over the experiment, to ensure that the control and experimental groups are exposed of similar subjects. This is sometimes done

by matching subjects on the basis of as many variables as possible. Another measure of an experiment's adequacy is its internal validity. Are variables sufficiently controlled? Has randomization been employed throughout?

A final crucial consideration in the evaluation of an experiment lies with its external validity. Can the results of the experiment be generalized? If so, what is the population to which it can be generalized? External validity is association with questions concerning the extent to which a sample is representative of the target population. Can the results of an experiment be generalized to the real world? It is normally easier to justify such generalization for field experiments than for laboratory experiments because of the artificiality of the latters overall, experimental findings are evaluated in terms of the reliability of the data, the scientific importance of the results, and the extent to which the data can be generalized.

Expost Facto Study

The Latin expost facto means literally, after the fact or retrospectively. Although an expost facto study is not really on experimental design, such a study sometimes is confused with the experimental method, and is frequently discussed with it. What is lacking in the expost facto study is control. Rather than introducing an independent variable and controlling other variables, the expost facto study seeks to analyse what has already happened in an effort to isolate the cause of the events. Mouly asserts that "This is experimentation in reverse the obvious weakness of such as "experiment" is that we have no control over the situations that have already occurred and we can never be sure of how many other circumstances might have been involved".

Expost facto study is a type of quasi-experimental design. Consider the hypothesis that academic courses in research methods affect graduate library science students in such a way as to cause them to the better librarians than librarians who did not complete such a course.

An expost facto analysis of graduated librarians with respect to an operational definition of success in librarianship (such as salary or job title) might well reveal that a strong relationship does indeed exist between completing a course in research methods and professional success in librarianship. But is the observed relationship casual in nature? Does the course cause the success?

The original hypothesis could be tested, albeit with some difficulty and over a period of years, by the experimental method. Such an approach might involves the random division of a

group of equivalent students into two groups, a control group – students not permitted to take research methods and an experimental group – students who are required to take research methods. The effect of the course on graduate students could then be observed over the student's careers. Relative success or failure in librarianship, however the definition is operationally defined, could then be said to have been caused by the treatment, in this case completing or not taking the course in research methods.

Major Steps in Experimental Method

1. Selecting and determining the problem
2. Reviewing the literature
3. Defining the population
4. Planning the experiment
5. Conducting the experiment
6. Measuring the outcomes
7. Analyzing and interpreting the outcome
8. Drawing up conclusions
9. Reporting the results

1. Selecting And Determining The Problems

i) Investigating the needs in the yield of action and deciding upon a problem.

ii) Conversion of problem into a hypothesis that can be verified or refuted by the experimental data.

2. Reviewing The Literature:

This implies the study of the literature related to similar problem.

3. Defining The Population:

It is necessary to define the population precisely so that there can be no question abut the population to which the conclusions are to apply.

4. Planning The Experiment:

This includes

i) Determining the method of experimentation

ii) Place of the experiment

iii) Duration of the experiment

iv) Determining the materials of the experiment

v) Conducting pilot study

vi) Selecting subjects or groups

5. Conducting The Experiments

This includes the following steps.

a) Control of variables and non-experimental factors

b) Keeping a careful record of steps in the procedure

c) Applying the experiment factor or factors

6. Measuring the Outcomes

This implies giving a careful consideration to the selection of the criterion on the basis of which the results are to be measured

7. Analyzing and Interpreting the Outcomes

This implies the need for competence in statistical data procedures.

8. Drawing Up the Conclusions

Care must be taken to restrict the conclusions to the conditions actually present in the experiment. The conclusions of the study must be restricted to the population actually investigated and care must be exercised not to over generalize the results.

9. Reporting the Results:

The study should be reported in such a way that the reader can make a judgment as to its adequacy.

Case Study

Case Study approach to research is rather a recent development in research. It involves a deeper investigation of a single unit, may be an individual, a family, an institution, a district, a community, or any single event selected for intensive examination case study approach has been developed essentially as a problem solving technique and also as a suggestive device for improvements in various dimensions of the case under examination. Suppose, a firm has been selected for a case study, if not only helps in investigating and finding solutions to some of the problems facing the firm but may also go a long way in suggesting improvements in the present functioning or operations of that particular firm. Thus case study method is more intensive in nature. The field of study is comparatively limited but has more of depth in it. It aims at studying everything about something, rather than something about everything as in the case of statistical method.

According to P.V.Young, "A comprehensive study of a social unit, be that unit a person, a group, a social institution, a district, or a community, is called a case study". According to Good and Hatt, "it is an approach which views and social unit as a whole. It is a way of organizing social data so as to preserve the unitary character of the social object being studied." The case study is thus a form of qualitative analysis involving the very careful and complete observation of a person, a situation or an institution.

This approach to research may not be based on a given hypothesis or on any established conclusions but the study itself may help in developing a well found ed hypothesis for further investigations. This approach to research, is therefore, an open and objective examination of a particular unit with a view to developing a hypothesis for further research, case study approach i.e. also flexible in character as a researcher has complete independence in approaching the problem from any angle that he considers as desirable from his point of view. As compared to historical approach therefore, this approach does not suffer from rigidly of the historical data.

Steps Involved in Case Study

The following are the steps involved in case study:

Step 1:Selection of cases and identification of situation:

The researcher has to decide about the following: (i) Unit has to be taken up for study (ii) Aspect or period of life of unit can be studied (iii) Situations in which unit exists. Therefore, the duty of the researcher is to choose representative and typical units. The selection of such units depends on the ability and skill of the researcher.

Step 2:Collecting and recording of data:

While collecting data, the researcher has to keep in mind that the continuity, breath and level of data to preserve the completeness and validity of data. The researcher has to use different aspects of the unit to collect broad array of facts. For collecting data in case study, study of personal documents, life histories, observations interviews, questionnaires schedules, a variety of tests like intelligent test, achievement test, aptitude test to be adopted. The collected data should be recorded properly. Also the recording of data must be uniform, topic wise accurate and clarity under principal themes of investigation. The data must also be complete and up to date and of easy reference. The careful collection and

recording of data completely depends on the skill and ability of the researcher. He should also try to seek more information and derive special features from them.

Step 3: Interpretation of data

Interpretation is the search for the broader meaning of the research findings. It is the deeper sense of investigation. It places particular event in the larger flow of events. It is the color, the atmosphere, the human element that gives meaning to a fact. It is in short setting, sequence and about all significance. The task of the interpretation falls on the shoulders of the research himself. Mere presentation of the facts is not enough; the facts collected must be clarified, explained and interpreted. The interpretation must be in logical and convenient form. Qualitative classification and building up a proper background theory are also essential. Good interpretation depends on the researcher's knowledge, imagination and wisdom.

Step 4: Reporting of data

The study should be reported in such a way that the reader can make a judgment as to its adequacy.

Advantages

The main advantages of case study approach are

a) It produces new ideas and fresh suggestions
b) It helps in formulating a sound hypothesis; and
c) It may also help in exploring new areas of research.

Since the case study approach makes an in-depth study of a particular unit of investigation and is always approached with an open mind, it bestows upon the researcher a high wealth of new ideas and new suggestions for further exploration of research fields. Investigator of an institution may uncover fresh knowledge about the problems that might not have occurred to the researcher before he undertook the investigation. He may also get new suggestions from the field of operation by intensively carrying out the examination of the case study.

Secondly, case study approach is very useful in helping the researcher to develop and formulate scientifically sound hypothesis for more research on a broader level. As already mentioned, a researcher may not start with a given hypothesis, but may desirably undertake a case study for formulating such hypothesis for further research. Case study approach has

also ad advantage in marking a multi-dimensional exploration of the same unit and this enrich the knowledge pertaining to a particular case for further use in policy formulation.

Thirdly, when a case study is undertaken, some of the areas of research may not have occurred to the researcher's mind and the very case study may open out new avenues of research where fruitful investigations can be undertaken either by the same researcher or other researchers.

Thus, this approach allows a concentrated focus on a single phenomenon and the utilization of a wide array of data gathering methods. The overall purpose a case study is to obtain comprehensive information about the research object. Data gathering methods used in case studies are based primarily upon direct observation; both participant and non-participant observation can be used when necessary, these methods are supplemented by structured techniques such as interviews and questionnaires.

CHAPTER VII

Sampling

Sampling is the act, process, or technique of selecting a suitable sample, or a representative part of a population for the purpose of determining parameters or characteristics of the whole population.

What is a sample?

A sample is a finite part of a statistical population whose properties are studied to gain information about the whole (Webster, 1985). When dealing with people, it can be defined as a set of respondents (people) selected from a larger population for the purpose of a survey.

A population is a group of individuals' persons, objects, or items from which samples are taken for measurement, for example, a population of presidents or professors, books or students.

Purpose of sampling

To draw conclusions about populations from samples, we must use inferential statistics which enables us to determine a population's characteristics by directly observing only a portion (or

sample) of the population. We obtain a sample rather than a complete enumeration (a census) of the population for many reasons. Obviously, it is cheaper to observe a partrather than the whole, but we should prepare ourselves to cope with the dangers of using samples. In this tutorial, we will investigate various kinds of sampling procedures. Some are better than others but all may yield samples that are inaccurate and unreliable. We will learn how to minimize these dangers, but some potential error is the price we must pay for the convenience and savings the samples provide.

There would be no need for statistical theory if a census rather than a sample was always used to obtain information about populations. But a census may not be practical and is almost never economical. There are six main reasons for sampling instead of doing a census. These are; -Economy -Timeliness -The large size of many populations -Inaccessibility of some of the population -Destructiveness of the observation -accuracy

The sampling process comprises several stages:

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

1. Population definition :

Successful statistical practice is based on focused problem definition. Typically, we seek to take action on some population, for example when a batch of material from production must be released to the customer or sentenced for scrap or rework.

Alternatively, we seek knowledge about the cause system of which the population is an outcome, for example when a researcher performs an experiment on rats with the intention of gaining insights into biochemistry that can be applied for the benefit of humans. In the latter

case, the population of concern can be difficult to specify, as it is in the case of measuring some physical characteristic such as the electrical conductivity of copper.

However, in all cases, time spent in making the population of concern precise is often well spent, often because it raises many issues, ambiguities and questions that would otherwise have been overlooked at this stage.

2 . Sampling frame

The sampling frame must be representative of the population and this is a question outside the scope of statistical theory demanding the judgment of experts in the particular subject matter being studied. All the above frames omit some people who will vote at the next election and contain some people who will not. People not in the frame have no prospect of being sampled. Statistical theory tells us about the uncertainties in extrapolating from a sample to the frame. In extrapolating from frame to population, its role is motivational and suggestive.

There is, however, a strong division of views about the acceptability of representative sampling across different domains of study. To the philosopher, the representative sampling procedure has no justification whatsoever because it is not how truth is pursued in philosophy. "To the scientist, however, representative sampling is the only justified procedure for choosing individual objects for use as the basis of generalization, and is therefore usually the only acceptable basis for ascertaining truth." (Andrew A. Marino) It is important to understand this difference to steer clear of confusing prescriptions found in many web pages.

In defining the frame, practical, economic, ethical, and technical issues need to be addressed. The need to obtain timely results may prevent extending the frame far into the future.

The difficulties can be extreme when the population and frame are disjoint. This is a particular problem in forecasting where inferences about the future are made from historical data. In fact, in 1703, when Jacob Bernoulli proposed to Gottfried Leibniz the possibility of using historical mortality data to predict the probability of early death of a living man, Gottfried Leibniz recognized the problem in replying:

"Nature has established patterns originating in the return of events but only for the most part. New illnesses flood the human race, so that no matter how many experiments you have done on corpses, you have not thereby imposed a limit on the nature of events so that in the future they could not vary."

Having established the frame, there are a number of ways for organizing it to improve efficiency and effectiveness.

It is at this stage that the researcher should decide whether the sample is in fact to be the whole population and would therefore be a census.

3. Sampling method

- Quota sampling
- Simple random sampling
- Stratified sampling
- Cluster sampling
- Random sampling
- Matched random sampling
- Systematic sampling
- Mechanical sampling
- Convenience sampling
- Line-intercept sampling

Within any of the types of frame identified above, a variety of sampling methods can be employed, individually or in combination.

Quota sampling

In quota sampling, the population is first segmented into mutually exclusive sub-groups, just as in stratified sampling. Then judgment is used to select the subjects or units from each segment based on a specified proportion. For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.

It is the second step which makes the technique one of non-probability sampling. In quota sampling the selection of the sample is non-random. For example interviewers might be tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for many years.

Simple random sampling

In a simple random sample of a given size, all such subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection: the frame is not subdivided or partitioned. It is possible that the sample will not be completely random.

Stratified sampling

Where the population embraces a number of distinct categories, the frame can be organized by these categories into separate "strata." A sample is then selected from each "stratum" separately, producing a stratified sample. The two main reasons for using a stratified sampling design are

1. to ensure that particular groups within a population are adequately represented in the sample, and
2. to improve efficiency by gaining greater control on the composition of the sample. In the second case, major gains in efficiency (either lower sample sizes or higher precision) can be achieved by varying the sampling fraction from stratum to stratum.

The sample size is usually proportional to the relative size of the strata. However, if variances differ significantly across strata, sample sizes should be made proportional to the stratum standard deviation. Disproportionate stratification can provide better precision than proportionate stratification. Typically, strata should be chosen to:

- have means which differ substantially from one another
- minimize variance within strata and maximize variance between strata.

Cluster sampling

Sometimes it is cheaper to 'cluster' the sample in some way e.g. by selecting respondents from certain areas only, or certain time-periods only. (Nearly all samples are in some sense 'clustered' in time - although this is rarely taken into account in the analysis.)

Cluster sampling is an example of 'two-stage sampling' or 'multistage sampling': in the first stage a sample of areas is chosen; in the second stage a sample of respondent *within* those areas is selected.

This can reduce travel and other administrative costs. It also means that one does not need a sampling frame for the entire population, but only for the selected clusters. Cluster sampling generally increases the variability of sample estimates above that of simple random sampling, depending on how the clusters differ between themselves, as compared with the within-cluster variation.

Random sampling

In random sampling, also known as probability sampling, every combination of items from the frame, or stratum, has a known probability of occurring, but these probabilities are not necessarily equal. With any form of sampling there is a risk that the sample may not adequately represent the population but with random sampling there is a large body of statistical theory which quantifies the risk and thus enables an appropriate sample size to be chosen. Furthermore, once the sample has been taken the sampling error associated with the measured results can be computed. With non-random sampling there is no measure of the associated sampling error. While such methods may be cheaper this is largely meaningless since there is no measure of quality. There are several forms of random sampling. For example, in simple random sampling, each element has an equal probability of being selected. Another form of random sampling is Bernoulli sampling in which each element has an equal probability of being selected, like in simple random sampling. However, Bernoulli sampling leads to a variable sample size, while during simple random sampling the sample size remains constant. Bernoulli sampling is a special case of Poisson sampling in which each element may have a

different probability of being selected. Other examples of probability sampling include stratified sampling and multistage sampling.

Matched random sampling

A method of assigning participants to groups in which pairs of participants are first matched on some characteristic and then individually assigned randomly to groups. (Brown, Cozby, Kee, & Worden, 1999, p.371).

The Procedure for Matched random sampling can be briefed with the following contexts,

a) Two samples in which the members are clearly paired, or are matched explicitly by the researcher. For example, IQ measurements pairs of identical twins.

b) Those samples in which the same attribute, or variable, are measured twice on each subject, under different circumstances commonly called repeated measures. Examples include the times of a group of athletes for 1500m before and after a week of special training; the milk yields of cows before and after being fed a particular diet.

Systematic sampling

Selecting (say) every 10^{th}name from the telephone directory is called an **every 10^{th}**sample, which is an example of systematic sampling. It is a type of probability sampling unless the directory itself is not randomized. It is easy to implement and the stratification induced can make it efficient, but it is especially vulnerable to periodicities in the list. If periodicity is present and the period is a multiple of 10, then bias will result. It is important that the first name chosen is not simply the first in the list, but is chosen to be (say) the 7^{th}, where 7 is a random integer in the range 1,...,10-1. Every 10^{th} sampling is especially useful for efficient sampling from databases.

Mechanical sampling

Mechanical sampling is typically used in sampling solids, liquids and gases, using devices such as grabs, scoops, thief probes, the COLIWASA and riffle splitter. Care is needed in ensuring that the sample is representative of the frame.

Convenience sampling

Sometimes called *grab* or *opportunity* sampling, this is the method of choosing items arbitrarily and in an unstructured manner from the frame. Though almost impossible to treat rigorously, it is the method most commonly employed in many practical situations. In social science research, snowball sampling is a similar technique, where existing study subjects are used to recruit more subjects into the sample.

Line-intercept sampling

Line-intercept sampling is a method of sampling elements in a region whereby an element is sampled if a chosen line segment, called a "transect", intersects the element.

Sampling error

What can make a sample unrepresentative of its population? One of the most frequent causes is sampling error. Sampling error comprises the differences between the sample and the population that are due solely to the particular units that happen to have been selected.

For example, suppose that a sample of 100 American women are measured and are all found to be taller than six feet. It is very clear even without any statistical prove that this would be a highly unrepresentative sample leading to invalid conclusions. This is a very unlikely occurrence because naturally such rare cases are widely distributed among the population. But it can occur. Luckily, this is a very obvious error and can be detected very easily.

The more dangerous error is the less obvious sampling error against which nature offers very little protection. An example would be like a sample in which the average height is overstated by only one inch or two rather than one foot which is more obvious. It is the unobvious error that is of much concern.

There are two basic causes for sampling error. One is chance: That is the error that occurs just because of bad luck. This may result in untypical choices. Unusual units in a population do exist and there is always a possibility that an abnormally large number of them will be chosen. For example, in a recent study in which I was looking at the number of trees, I selected a sample of households randomly but strange enough, the two households in the whole population, which had the highest number of trees (10,018 and 6345) were both selected making the sample average higher than it should be. The average with these two extremes removed was 828 trees. The main protection against this kind of error is to use a large enough sample. The second cause of sampling is sampling bias.

Sampling bias is a tendency to favour the selection of units that have particular characteristics. Sampling bias is usually the result of a poor sampling plan. The most notable is the bias of non response when for some reason some units have no chance of appearing in the sample. For example, take a hypothetical case where a survey was conducted recently by CornellGraduateSchool to find out the level of stress that graduate students were going through. A mail questionnaire was sent to 100 randomly selected graduate students. Only 52 responded and the results were that students were not under stress at that time when the actual case was that it was the highest time of stress for all students except those who were writing their thesis at their own pace. Apparently, this is the group that had the time to respond. The researcher who was conducting the study went back to the questionnaire to find out what the problem was and found that all those who had responded were third and fourth PhD students. Bias can be very costly and has to be guarded against as much as possible. For this case, $2000.00 had been spent and there were no reliable results in addition, it cost the researcher his job since his employer thought if he was qualified, he should have known that before hand and planned on how to avoid it. A means of selecting the units of analysis must be designed to avoid the more obvious forms of bias. Another example would be where you would like to know the average income of some community and you decide to use the telephone numbers to select a sample of the total population in a locality where only the rich and middle class households have telephone lines. You will end up with high average income which will lead to the wrong policy decisions.

Non sampling error (measurement error)

The other main cause of unrepresentative samples is non sampling error. This type of error can occur whether a census or a sample is being used. Like sampling error, non- sampling error may either be produced by participants in the statistical study or be an innocent by product of the sampling plans and procedures.

A non sampling error is an error that results solely from the manner in which the observations are made. The simplest example of non sampling error is inaccurate measurements due to malfunctioning instruments or poor procedures. For example, consider the observation of human weights. If persons are asked to state their own weights themselves, no two answers will be of equal reliability. The people will have weighed themselves on different scales in various states of poor calibration. An individual's weight fluctuates diurnally by several pounds, so that the time of weighing will affect the answer. The scale reading will also vary with the person's state of undress. Responses therefore will not be of comparable validity unless all persons are weighed under the same circumstances.

Biased observations due to inaccurate measurement can be innocent but very devastating. A story is told of a French astronomer who once proposed a new theory based on spectroscopic measurements of light emitted by a particular star. When his collogues discovered that the measuring instrument had been contaminated by cigarette smoke, they rejected his findings.

UNIT IV

CHAPTER VIII

Data Collection

Definition

Data collection is a process of gatheringdatasystematically for a particular purpose from various sources, including questionnaires, interviews, observation, existing records, and electronicdevices. The process is usually preliminary to statistical analysis of the data.

Kinds of Information

Different kinds of information are required in social research. This can be classified into the following two types of information.

1. Primary Data
2. Secondary Data

Primary Data Collection Methods

In primary data collection, the data is collected by the researcher himself using methods such as interviews and questionnaires. The key point here is that the data collected is unique to researcher and his research and, until he publishes, no one else can access to it.There are many methods of collecting primary data and the main methods include:

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

The primary data, which is generated by the above methods, may be qualitative in nature (usually in the form of words) or quantitative (usually in the form of numbers or where we can make counts of words used).

Questionnaires

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

Advantages

- Can be used as a method in its own right or as a basis for interviewing or a telephone survey.
- Can be posted, e-mailed or faxed.
- Can cover a large number of people or organizations.
- Wide geographic coverage.
- Relatively cheap.
- No prior arrangements are needed.
- Avoids embarrassment on the part of the respondent.
- Respondent can consider responses.
- Possible anonymity of respondent.
- No interviewer bias.

Disadvantages

- Design problems.
- Questions have to be relatively simple.
- Historically low response rate (although inducements may help).
- Time delay whilst waiting for responses to be returned.
- Require a return deadline.
- Several reminders may be required.
- Assumes no literacy problems.
- No control over who completes it.
- Not possible to give assistance if required.
- Problems with incomplete questionnaires.
- Replies not spontaneous and independent of each other.

- Respondent can read all questions beforehand and then decide whether to complete or not. For example, perhaps because it is too long, too complex, uninteresting, or too personal.

Design of postal questionnaires

Theme and covering letter

The general theme of the questionnaire should be made explicit in a covering letter. The researcher should state who he is; why the data is required; give, if necessary, an assurance of confidentiality and/or anonymity; and contact number and address or telephone number. This ensures that the respondents know what they are committing themselves to, and also that they understand the context of their replies. If possible, the researcher should offer an estimate of the completion time. Instructions for return should be included with the return date made obvious. For example: 'It would be appreciated if you could return the completed questionnaire by... if at all possible'.

Instructions for completion

The researcher needs to provide clear and unambiguous instructions for completion. Within most questionnaires these are general instructions and specific instructions for particular question structures. It is usually best to separate these, supplying the general instructions as a preamble to the questionnaire, but leaving the specific instructions until the questions to which they apply. The response method should be indicated (circle, tick, cross, etc.). Wherever possible, and certainly if a slightly unfamiliar response system is employed, the researcher should give an example.

Appearance

Appearance is usually the first feature of the questionnaire to which the recipient reacts. A neat and professional look will encourage further consideration of the researcher's request, increasing his response rate. In addition, careful thought to layout should help his analysis. There are a number of simple rules to help improve questionnaire appearance:

- Liberal spacing makes the reading easier.
- Photo-reduction can produce more space without reducing content.
- Consistent positioning of response boxes, usually to the right, speeds up completion and also avoids inadvertent omission of responses.
- Choose the font style to maximize legibility.
- Differentiate between instructions and questions. Either lower case or capitals can be used, or responses can be boxed.

Length

There may be a strong temptation to include any vaguely interesting questions, but the researcher should resist this at all costs. Excessive size can only reduce response rates. If a long questionnaire is necessary, then you must give even more thought to appearance. It is better to leave pages unnumbered; for respondents to flick to the end and see 'page 27' can be very disconcerting!

Order

Probably the most crucial stage in questionnaire response is the beginning. Once the respondents have started to complete the questions they will normally finish the task, unless it is very long or difficult. Consequently, the researcher needs to select the opening questions with care. Usually the best approach is to ask for biographical details first, as the respondents should know all the answers without much thought. Another benefit is that an easy start provides practice in answering questions.

Once the introduction has been achieved the subsequent order will depend on many considerations. The researcher should be aware of the varying importance of different questions. Essential information should appear early, just in case the questionnaire is not completed. For the same reasons, relatively unimportant questions can be placed towards the end. If questions are likely to provoke the respondent and remain unanswered, these too are best left until the end, in the hope of obtaining answers to everything else.

Coding

If analysis of the results is to be carried out using a statistical package or spreadsheet it is advisable to code non-numerical responses when designing the questionnaire, rather than trying to code the responses when they are returned. An example of coding is:

Male [ ]	Female [ ]
1	2

The coded responses (1 or 2) are then used for the analysis.

Thank you

Respondents to questionnaires rarely benefit personally from their efforts and the least the researcher can do is to thank them. Even though the covering letter will express appreciation for the help given, it is also a nice gesture to finish the questionnaire with a further thank you.

Questions

- Keep the questions short, simple and to the point; avoid all unnecessary words.
- Use words and phrases that are unambiguous and familiar to the respondent. For example, 'dinner' has a number of different interpretations; use an alternative expression such as 'evening meal'.
- Only ask questions that the respondent can answer. Hypothetical questions should be avoided. Avoid calculations and questions that require a lot of memory work, for example, 'How many people stayed in your hotel last year?'
- Avoid loaded or leading questions that imply a certain answer. For example, by mentioning one particular item in the question, 'Do you agree that Colgate toothpaste is the best toothpaste?'
- Vacuous words or phrases should be avoided. 'Generally', 'usually', or 'normally' are imprecise terms with various meanings. They should be replaced with quantitative statements, for example, 'at least once a week'.
- Questions should only address a single issue. For example, questions like: 'Do you take annual holidays to Spain?' should be broken down into two discreet stages, firstly find

out if the respondent takes an annual holiday, and then secondly find out if they go to Spain.

- Do not ask two questions in one by using 'and'. For example, 'Did you watch television last night and read a newspaper?'
- Avoid double negatives. For example, 'Is it not true that you did not read a newspaper yesterday?' Respondents may tackle a double negative by switching both negatives and then assuming that the same answer applies. This is not necessarily valid.
- State units required but do not aim for too high a degree of accuracy. For instance, use an interval rather than an exact figure:

'How much did you earn last year?'

Less than £10,000 [ ]

£10,000 but less than £20,000 [ ]

Avoid emotive or embarrassing words – usually connected with race, religion, politics, sex, money.

Types of questions

Closed questions

A question is asked and then a number of possible answers are provided for the respondent. The respondent selects the answer which is appropriate. Closed questions are particularly useful in obtaining factual information:

Sex: Male [ ] Female [ ]

Did you watch television last night? Yes [ ] No [ ]

Some 'Yes/No' questions have a third category 'Do not know'. Experience shows that as long as this alternative is not mentioned people will make a choice. Also the phrase 'Do not know' is ambiguous:

Do you agree with the introduction of the EMU?

Yes [ ] No [ ] Do not know [ ]

What was your main way of traveling to the hotel? Tick one box only.

Car	[ ]
Coach	[ ]
Motor bike	[ ]
Train	[ ]
Other means, please specify _____	

With such lists the researcher should always include an 'other' category, because not all possible responses might have been included in the list of answers. Sometimes the respondent can select more than one from the list. However, this makes analysis difficult.

Why have you visited the historic house? Tick the relevant answer(s). You may tick as many as you like.

I enjoy visiting historic houses	[ ]
The weather was bad and I could not enjoy outdoor activities	[ ]
I have visited the house before and wished to return	[ ]
Other reason, please specify _____	

Attitude questions

Frequently questions are asked to find out the respondents' opinions or attitudes to a given situation. A Likert scale provides a battery of attitude statements. The respondent then says how much they agree or disagree with each one.

Read the following statements and then indicate by a tick whether you strongly agree, agree, disagree or strongly disagree with the statement.

	Strongly agree	Agree	Disagree	Strongly disagree
My visit has been good value for money				

There are many variations on this type of question. One variation is to have a 'middle statement', for example, 'Neither agree nor disagree'. However, many respondents take this as the easy option. Only having four statements, as above, forces the respondent into making a positive or negative choice. Another variation is to rank the various attitude statements; however, this can cause analysis problems:

Which of these characteristics do you like about your job? Indicate the best three in order, with the best being number 1.

Varied work [ ]
Good salary [ ]
Opportunities for promotion [ ]
Good working conditions [ ]
High amount of responsibility [ ]
Friendly colleagues [ ]

A semantic differential scale attempts to see how strongly an attitude is held by the respondent. With these scales double-ended terms are given to the respondents who are asked to indicate where their attitude lies on the scale between the terms. The response can be indicated by putting a cross in a particular position or circling a number.

Work is: (circle the appropriate number)

Difficult	1	2	3	4	5	6	7	Easy
Useless	1	2	3	4	5	6	7	Useful
Interesting	1	2	3	4	5	6	7	Boring

For summary and analysis purposes, a 'score' of 1 to 7 may be allocated to the seven points of the scale, thus quantifying the various degrees of opinion expressed. This procedure has some disadvantages. It is implicitly assumed that two people with the same strength of feeling will mark the same point on the scale. This almost certainly will not be the case. When faced with a semantic differential scale, some people will never, as a matter of principle, use the two end indicators of 1 and 7. Effectively, therefore, they are using a five-point scale. Also scoring the scale 1 to 7 assumes that they represent equidistant points on the continuous spectrum of opinion. This again is probably not true. Nevertheless, within its limitations, the semantic differential can provide a useful way of measuring and summarizing subjective opinions.

Other types of questions to determine peoples' opinions or attitudes are:

Which one/two words best describes...?

Which of the following statements best describes...?

How much do you agree with the following statement...?

Open questions

An open question such as 'What are the essential skills a manager should possess?' should be used as an adjunct to the main theme of the questionnaire and could allow the respondent to elaborate upon an earlier more specific question. Open questions inserted at the end of major sections, or at the end of the questionnaire, can act as safety valves, and possibly offer additional information. However, they should not be used to introduce a section since there is a high risk of influencing later responses. The main problem of open questions is that many different answers have to be summarized and possibly coded.

Testing – pilot survey

Questionnaire design is fraught with difficulties and problems. A number of rewrites will be necessary, together with refinement and rethinks on a regular basis. Do not assume that you will write the questionnaire accurately and perfectly at the first attempt. If poorly designed,

the researcher will collect inappropriate or inaccurate data and good analysis cannot then rectify the situation.

To refine the questionnaire, the researcher needs to conduct a pilot survey. This is a small-scale trial prior to the main survey that tests all his question planning. Amendments to questions can be made. After making some amendments, the new version would be re-tested. If this re-test produces more changes, another pilot would be undertaken and so on. For example, perhaps responses to open-ended questions become closed; questions which are all answered the same way can be omitted; difficult words replaced, etc.

It is usual to pilot the questionnaires personally so that the respondent can be observed and questioned if necessary. By timing each question, the researcher can identify any questions that appear too difficult, and he can also obtain a reliable estimate of the anticipated completion time for inclusion in the covering letter. The result can also be used to test the coding and analytical procedures to be performed later.

Distribution and return

The questionnaire should be checked for completeness to ensure that all pages are present and that none is blank or illegible. It is usual to supply a prepaid addressed envelope for the return of the questionnaire. The researcher needs to explain this in the covering letter and reinforce it at the end of the questionnaire, after the 'Thank you'.

Finally, many organizations are approached continually for information. Many, as a matter of course, will not respond in a positive way.

Interviews

Interviewing is a technique that is primarily used to gain an understanding of the underlying reasons and motivations for people's attitudes, preferences or behaviour. Interviews

can be undertaken on a personal one-to-one basis or in a group. They can be conducted at work, at home, in the street or in a shopping centre, or some other agreed location.

Personal interview

Advantages

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

Disadvantages

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

- If many interviewers, training required.

Types of interview

Structured

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

Semi-structured

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

Unstructured

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

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

Planning an interview:

- List the areas in which you require information.

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

Conducting an interview:

- Personally – arrive on time be smart smile employ good manners find a balance between friendliness and objectivity.
- At the start – introduce yourself re-confirm the purpose assure confidentiality – if relevant specify what will happen to the data.
- The questions – speak slowly in a soft, yet audible tone of voice control your body language know the questions and topic ask all the questions.
- Responses – recorded as you go on questionnaire written verbatim, but slow and time-consuming summarized by you taped – agree beforehand – have alternative method if not acceptable consider effect on respondent's answers proper equipment in good working order sufficient tapes and batteries minimum of background noise.
- At the end – ask if the respondent would like to give further details about anything or any questions about the research thank them.

Telephone interview

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

Advantages

- Relatively cheap.

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

Disadvantages

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

Getting started

- Locate the respondent
    - Repeat calls may be necessary especially if the researcher is trying to contact people in organizations where he may have to go through secretaries.
    - The researcher may not know an individual's name or title – so there is the possibility of interviewing the wrong person.

- The researcher can send an advance letter informing the respondent that you will be telephoning. This can explain the purpose of the research.

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

Ensuring quality

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

Smooth implementation

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

Comparison of postal, telephone and personal interview surveys

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

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

technical skills in conducting interview			
Response rate	Low	Usually high	High

Table: Comparison of the three common methods of surveys

Focus group interviews

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

Observation

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

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

Structured or unstructured

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

In unstructured observation, the researcher monitors all aspects of the phenomenon that seem relevant. It is appropriate when the problem has yet to be formulated precisely and

flexibility is needed in observation to identify key components of the problem and to develop hypotheses. The potential for bias is high. Observation findings should be treated as hypotheses to be tested rather than as conclusive findings.

Disguised or undisguised

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

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

Natural or contrived

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

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

Personal

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

Mechanical

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

Non-participant

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

Participant

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

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

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

Researcher as employee

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

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

Researcher as an explicit role

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

Interrupted involvement

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

Observation alone

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

Choice of roles

The role adopted depends on the following:

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

- The extent to which the researcher would be comfortable in the role: If the researcher intends to keep his identity concealed, will he or she also feel able to develop the type of trusting relationships that are important? What are the ethical issues?

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

Case-studies

The term case-study usually refers to a fairly intensive examination of a single unit such as a person, a small group of people, or a single company. Case-studies involve measuring what is there and how it got there. In this sense, it is historical. It can enable the researcher to explore, unravel and understand problems, issues and relationships. It cannot, however, allow the researcher to generalize, that is, to argue that from one case-study the results, findings or theory developed apply to other similar case-studies. The case looked at may be unique and, therefore not representative of other instances. It is, of course, possible to look at several case-studies to represent certain features of management that we are interested in studying. The case-study approach is often done to make practical improvements. Contributions to general knowledge are incidental.

The case-study method has four steps:

1. Determine the present situation.
2. Gather background information about the past and key variables.
3. Test hypotheses. The background information collected will have been analysed for possible hypotheses. In this step, specific evidence about each hypothesis can be gathered. This step aims to eliminate possibilities which conflict with the evidence collected and to gain confidence for the important hypotheses. The culmination of this step might be the development of an experimental design to test out more rigorously the hypotheses developed, or it might be to take action to remedy the problem.

4. Take remedial action. The aim is to check that the hypotheses tested actually work out in practice. Some action, correction or improvement is made and a re-check carried out on the situation to see what effect the change has brought about.

The case-study enables rich information to be gathered from which potentially useful hypotheses can be generated. It can be a time-consuming process. It is also inefficient in researching situations which are already well structured and where the important variables have been identified. They lack utility when attempting to reach rigorous conclusions or determining precise relationships between variables.

Diaries

A diary is a way of gathering information about the way individuals spend their time on professional activities. They are not about records of engagements or personal journals of thought! Diaries can record either quantitative or qualitative data, and in management research can provide information about work patterns and activities.

Advantages

- Useful for collecting information from employees.
- Different writers compared and contrasted simultaneously.
- Allows the researcher freedom to move from one organization to another.
- Researcher not personally involved.
- Diaries can be used as a preliminary or basis for intensive interviewing.
- Used as an alternative to direct observation or where resources are limited.

Disadvantages

- Subjects need to be clear about what they are being asked to do, why and what you plan to do with the data.
- Diarists need to be of a certain educational level.

- Some structure is necessary to give the diarist focus, for example, a list of headings.
- Encouragement and reassurance are needed as completing a diary is time-consuming and can be irritating after a while.
- Progress needs checking from time-to-time.
- Confidentiality is required as content may be critical.
- Analyses problems, so you need to consider how responses will be coded before the subjects start filling in diaries.

Critical incidents

The critical incident technique is an attempt to identify the more 'noteworthy' aspects of job behaviour and is based on the assumption that jobs are composed of critical and non-critical tasks. For example, a critical task might be defined as one that makes the difference between success and failure in carrying out important parts of the job. The idea is to collect reports about what people do that is particularly effective in contributing to good performance. The incidents are scaled in order of difficulty, frequency and importance to the job as a whole.

The technique scores over the use of diaries as it is centred on specific happenings and on what is judged as effective behaviour. However, it is laborious and does not lend itself to objective quantification.

Portfolios

A measure of a manager's ability may be expressed in terms of the number and duration of 'issues' or problems being tackled at any one time. The compilation of problem portfolios is recording information about how each problem arose, methods used to solve it, difficulties encountered, etc. This analysis also raises questions about the person's use of time. What proportion of time is occupied in checking; in handling problems given by others; on self-generated problems; on 'top-priority' problems; on minor issues, etc? The main problem with this method and the use of diaries is getting people to agree to record everything in sufficient detail for you to analyse. It is very time-consuming!

Sampling

Collecting data is time consuming and expensive, even for relatively small amounts of data. Hence, it is highly unlikely that a complete population will be investigated. Because of the time and cost elements the amount of data you collect will be limited and the number of people or organizations you contact will be small in number. You will, therefore, have to take a sample and usually a small sample.

Sampling theory says a correctly taken sample of an appropriate size will yield results that can be applied to the population as a whole. There is a lot in this statement but the two fundamental questions to ensure generalization are:

1. How is a sample taken correctly?
2. How big should the sample be?

The answer to the second question is 'as large as possible given the circumstances'. It is like answering the question 'How long is a piece of string'? It all depends on the circumstances.

Whilst we do not expect you to normally generalize your results and take a large sample, we do expect that you follow a recognized sampling procedure, such that, if the sample was increased generalization would be possible. You therefore need to know some of the basics of sampling. This will be done by reference to the following example.

The theory of sampling is based on random samples – where all items in the population have the same chance of being selected as sample units. Random samples can be drawn in a number of ways but are usually based on having some information about population members. This information is usually in the form of an alphabetical list – called the sampling frame.

Three types of random sample can be drawn – a simple random sample (SRS), a stratified sample and a systematic sample.

Simple random sampling

Simple random sampling can be carried out in two ways – the lottery method and using random numbers.

The lottery method involves:

- transferring each person's name from the list and putting it on a piece of paper
- the pieces of paper are placed in a container and thoroughly mixed
- the required number are selected by someone without looking
- the names selected are the simple random sample.

This is basically similar to a game of bingo or the national lottery. This procedure is easy to carry out especially if both population and sample are small, but can be tedious and time consuming for large populations or large samples.

Alternatively random numbers can be used. Random numbers are strings of digits that have been generated by the lottery method and can be found in books of statistical tables. An example of these is:

03	47	43	73	86	36	96	47	36	61
97	74	24	67	62	42	81	14	57	20
16	76	62	27	66	56	50	26	71	07
12	56	85	99	26	96	96	68	27	31
55	59	56	35	64	38	04	80	46	22

Random numbers tend to be written in pairs and blocks of 5 by 5 to make reading easy. However, care is needed when reading these tables. The numbers can be read in any direction but they should be read as a singe string of digits i.e. left to right as 0, 3, 4, 7 etc', or top to bottom as 0, 9, 1, 1, 5, 3, 7, … etc. It is usual to read left to right.

The random number method involves:

- Allocating a number to each person on the list (each number must consist of the same number of digits so that the tables can be read consistently).
- Find a starting point at random in the tables (close your eyes and point).

- Read off the digits.
- The names matching the numbers are the sample units.

For the example of selecting nine people at random from 90:

a) The sampling frame is the list of 90 people. Number this list 00, 01, 02, …, 89. Note that each number has two digits and the numbering starts from 00.

b) Suppose a starting point is found at random from the random number tables and let this number be 16. Then the person that has been numbered 16 is the first sample unit.

c) Let the next two digits be 76, then the person numbered 76 is the second sample unit.

This procedure is repeated until the nine people have been identified.

d) Any number occurring for second time is ignored as is any two-digit number over 89.

Simple random number sampling is used as the basis for many other sampling methods, but has two disadvantages:

1. A sampling frame is required. This may not be available, exist or be incomplete.

2. The procedure is unbiased but the sample may be biased. For instance, if the 90 people are a mixture of men and women and all men were selected this would be a biased sample.

To overcome this problem a stratified sample can be taken. In this the population structure is reflected in the sample structure, with respect to some criterion.

For example, suppose the 90 people consist of 30 men and 60 women. If gender is the criterion for stratification then:

$\frac{30}{90}$ of the sample should be men

$$\text{ie.} \quad \frac{30}{90} \times 9 = 3 \text{ men}$$

$\frac{60}{90}$ of the sample should be women

$$\text{ie.} \quad \frac{60}{90} \times 9 = 6 \text{ women}$$

Thus the sample reflects the population structure in terms of gender.

The three men and six women would then be selected by simple random sampling e.g., random numbers.The problem with this approach is the criterion for stratification, (e.g., age, sex, job description), is chosen by you – it is subjective and may not be the best or more appropriate criterion. Also a more detailed sampling frame is required.

Systematic sampling

Whilst not truly random this is a method that is used extensively because it is easy to operate and quick, even when the population and the sample are large.

For example, for the population 90 and sample of nine:

Split the sampling frame in to nine equal groups.

i.e. 1 to 9

10 to 19

etc

80 to 89

Select a number between 1 and 9 using random number tables.

Suppose this number is 6.

Person numbered 6 is chosen.

Then the 16th, 26th, 36th, 46th, 56th, 66th, 76th, and 86th people are the remaining sample units.

If no sampling frame is available access to the population is necessary, such as customers of a business such as a leisure centre, restaurant or museum.

Systematic sampling can be used by selecting a random number say 25.

Then the 25th person to enter is the first sample unit.

The 50th person to enter is the second sample unit.

This process is carried on until the required sample size is met.

This approach usually generates a good cross section of the population. However, you may need a team of people when no sampling frame exists to help with counting, interviewing, etc.

Data recording and analysis

When the researcher is at the planning stage of his research design, it is worth thinking about how he is going to record his data and, even more importantly, how he is going to analyse it. It is pointless collecting data in a form that the researcher cannot understand or analyse!

Secondary Data Collection Methods

All methods of data collection can supply quantitative data (numbers, statistics or financial) or qualitative data (usually words or text). Quantitative data may often be presented in tabular or graphical form. Secondary data is data that has already been collected by someone else for a different purpose to yours. For example, this could mean using:

- data collected by a hotel on its customers through its guest history system
- data supplied by a marketing organisation

- annual company reports
- government statistics.

Secondary data can be used in different ways:

- The researcher can simply report the data in its original format. If so, then it is most likely that the place for this data will be in his main introduction or literature review as support or evidence for his argument.

- The researcher can do something with the data. If he uses it (analyse it or re-interpret it) for a different purpose to the original then the most likely place would be in the 'Analysis of findings' section of his dissertation. A good example of this usage was the work on suicide carried out by Durkheim. He took the official suicide statistics of different countries (recorded by coroners or their equivalent) and analysed them to see if he could identify variables that would mean that some people are more likely to commit suicide than others. He found, for example, that Catholics were less likely to commit suicide than Protestants. In this way, he took data that had been collected for quite a different purpose and used it in his own study – but he had to do a lot of comparisons and statistical correlations himself in order to analyse the data. (See Haralambos, 1995, for details of Durkheim's work).

Most research requires the collection of primary data (data that you collect at first hand), and this is what students concentrate on. Unfortunately, many dissertations do not include secondary data in their findings section although it is perfectly acceptable to do so, providing you have analyzed it. It is always a good idea to use data collected by someone else if it exists – it may be on a much larger scale than you could hope to collect and could contribute to your findings considerably.

As secondary data has been collected for a different purpose to yours, you should treat it with care. The basic questions you should ask are:

- Where has the data come from?
- Does it cover the correct geographical location?

- Is it current (not too out of date)?
- If you are going to combine with other data are the data the same (for example, units, time, etc.)?
- If you are going to compare with other data are you comparing like with like?

Thus the researcher should make a detailed examination of the following:

- Title (for example, the time period that the data refers to and the geographical coverage).
- Units of the data.
- Source (some secondary data is already secondary data).
- Column and row headings, if presented in tabular form.
- Definitions and abbreviations, for example, what does SIC stand for? For example, how is 'small' defined in the phrase 'small hotel'? Is 'small' based on the number of rooms, value of sales, number of employees, profit, turnover, square metres of space, etc., and do different sources use the word 'small' in different ways? Even if the same unit of measurement is used, there still could be problems. For example, in Norway, firms with 200-499 employees are defined as 'medium', whereas in the USA firms with less than 500 employees are defined as 'small'.

There are many sources of data and most people tend to underestimate the number of sources and the amount of data within each of these sources.

Sources can be classified as:

- paper-based sources – books, journals, periodicals, abstracts, indexes, directories, research reports, conference papers, market reports, annual reports, internal records of organisations, newspapers and magazines
- electronic sources– CD-ROMs, on-line databases, Internet, videos and broadcasts.

The main sources of qualitative and quantitative secondary data include the follwing:

- Official or government sources.
- Unofficial or general business sources.

The output of all publishers of non-official sources is included in the most comprehensive directory available:

Mort D. (1997) Sources of Unofficial UK Statistics 3rd Edition Aldershot: Gower

The guide lists 1,059 statistical titles and series published by 635 different organisations. It excludes one-off surveys or market reports.

The arrangement is alphabetical by organisation with details of titles produced and contacts for further information. It lists references to the following types of sources:

  - trade associations
  - trade and other journals
  - private research publishers
  - stockbroking firms
  - large company market reports
  - local authorities
  - professional bodies
  - academic institutions.
- European Union (Community) sources.
- International sources.
  - Organisation for Economic Co-operation and Development (OECD)
  - United Nations and related organisations.

Sources for the last two categories are many and varied. If your dissertation requires these sources you need to conduct a more thorough search of your library and perhaps seek the assistance of the librarian.

Conclusion

In this unit, we have covered the various methods used to collect data and the various methods of sampling and assessing the data collection process. There is a lot to absorb and the researcher needs to make decisions about the best methods to use to collect the data that the researcher needs to answer his research questions. Each piece of research is very individual and when he comes to writing up his methodology section he needs to be able to justify and evaluate the methods he has used. So it is a good idea to have planned this aspect very carefully!

References

http://brent.tvu.ac.uk/dissguide/hm1u2/hm1u2fra.htm

UNIT V

Data analysis

Introduction

Data analysis is the process of looking at and summarizing **data** with the intent to extract useful information and develop conclusions.Developments in the field of statistical data analysis often parallel or follow advancements in other fields to which statistical methods are fruitfully applied. Because practitioners of the statistical analysis often address particular applied decision problems, methods developments is consequently motivated by the search to a better decision making under uncertainties.

Statistical Data Analysis

Data are not Information!

To determine what statistical data analysis is, one must first define statistics. Statistics is a set of methods that are used to collect, analyze, present, and interpret data. Statistical methods are used in a wide variety of occupations and help people identify, study, and solve

many complex problems. In the business and economic world, these methods enable decision makers and managers to make informed and better decisions about uncertain situations.

Vast amounts of statistical information are available in today's global and economic environment because of continual improvements in computer technology. To compete successfully globally, researchers and decision makers must be able to understand the information and use it effectively. Statistical data analysis provides hands on experience to promote the use of statistical thinking and techniques to apply in order to make educated decisions in the research.

Computers play a very important role in statistical data analysis. The statistical software package, SPSS, which is used in this course, offers extensive data-handling capabilities and numerous statistical analysis routines that can analyze small to very large data statistics. The computer will assist in the summarization of data, but statistical data analysis focuses on the interpretation of the output to make inferences and predictions.

Studying a problem through the use of statistical data analysis usually involves four basic steps.

1. Defining the problem
2. Collecting the data
3. Analyzing the data
4. Reporting the results

Defining the Problem

An exact definition of the problem is imperative in order to obtain accurate data about it. It is extremely difficult to gather data without a clear definition of the problem.

Collecting the Data

We live and work at a time when data collection and statistical computations have become easy almost to the point of triviality. Paradoxically, the design of data collection, never sufficiently emphasized in the statistical data analysis textbook, have been weakened by an apparent belief that extensive computation can make up for any deficiencies in the design of data collection. One must start with an emphasis on the importance of defining the population about which we are seeking to make inferences; all the requirements of sampling and experimental design must be met.

Designing ways to collect data is an important job in statistical data analysis. Two important aspects of a statistical study are: Population - a set of all the elements of interest in a study Sample - a subset of the population Statistical inference is refer to extending your knowledge obtain from a random sample from a population to the whole population. This is known in mathematics as an Inductive Reasoning. That is, knowledge of whole from a particular. Its main application is in hypotheses testing about a given population. The purpose of statistical inference is to obtain information about population form information contained in a sample. It is just not feasible to test the entire population, so a sample is the only realistic way to obtain data because of the time and cost constraints. Data can be either quantitative or qualitative. Qualitative data are labels or names used to identify an attribute of each element. Quantitative data are always numeric and indicate either how much or how many.

For the purpose of statistical data analysis, distinguishing between cross-sectional and time series data is important. Cross-sectional data re data collected at the same or approximately the same point in time. Time series data are data collected over several time periods.

Data can be collected from existing sources or obtained through observation and experimental studies designed to obtain new data. In an experimental study, the variable of interest is identified. Then one or more factors in the study are controlled so that data can be obtained about how the factors influence the variables. In observational studies, no attempt is made to control or influence the variables of interest. A survey is perhaps the most common type of observational study.

Analyzing the Data

Statistical data analysis divides the methods for analyzing data into two categories: exploratory methods and confirmatory methods. Exploratory methods are used to discover what the data seems to be saying by using simple arithmetic and easy-to-draw pictures to summarize data. Confirmatory methods use ideas from probability theory in the attempt to answer specific questions. Probability is important in decision making because it provides a mechanism for measuring, expressing, and analyzing the uncertainties associated with future events. The majority of the topics addressed in this course fall under this heading.

Reporting the Results

Through inferences, an estimate or test claims about the characteristics of a population can be obtained from a sample. The results may be reported in the form of a table, a graph or a set of percentages. The reported results must reflect the uncertainty through the use of probability statements and intervals of values, because only a small collection (sample) has been examined and not an entire population.

Measures of Central Tendency

Condensation of data is necessary in statistical analysis. This is because a large number of big figures are not only confusing but also difficult to analyze. Therefore, in order to reduce the complexity of data and to make them comparable it is necessary that various phenomena, which are being compared, are reduced to a single figure. The first of such measure is averages or measures of central tendency.

Measures of central tendency are measures of the location of the middle or the center of a distribution. The definition of "middle" or "center" is purposely left somewhat vague so that the term "central tendency" can refer to a wide variety of measures.

Definition

Some of the important definitions are given below:

1. “Statistics is a value which is typical or representative of a set of data”
2. “Average is an attempt to find one single value to describe the whole of figures”
3. “The average is sometimes described as number which is typical of the whole group”
4. “An average is only a short way of expressing an arithmetic result”
5. “An average is a single number describing some features of a set of data”.

Characteristics of a typical average

A measure of central tendency is a typical value around which other figures congregate. According to Yule and Kendall, average must possess the following characteristics:

It should be

1. rigidly defined so that there is no confusion with regard to its meaning connotation.
2. easy to understand.
3. simple to compute.
4. based on all the items in the data.
5. capable of further algebraic treatment.
6. capable of being used in further statistical computation.
7. Its definition should be in the form of a mathematical formula.
8. It should not be unduly influenced by any single item or group of items.
9. It should have sampling stability.

Uses of an Average

1. To get one single value that describes the salient features of a mass of complex data.
2. To help comparison.
3. To trace the mathematical relationship.
4. To help in decision making.
5. To know about the Universe.

Types of Averages

The three most significant types of averages are

1. Arithmetic Mean
2. Median
3. Mode

1. Arithmetic Mean

Arithmetic average is also called as Mean. It is the most common type and widely used measure of central tendency. Arithmetic average of a series is the figure obtained by dividing the total value of the various items by the number of numbers.

a) Individual Observation

The simple arithmetic mean (AM) of an individual series is equal to the sum of the values divided by the number of observations. Let the values of the variable X be X_1, X_2, K, X_n and let $\overline{X}$ denote the AM of X.

Step1: Add up all the values of the variable X and find out $\sum X$.

Step2: Divide $\sum X$ by the number of observations (N).

That is. $\overline{X} = \frac{X_1 + X_2 + \text{K} + X_n}{N} = \frac{\sum X}{N}$.

Example

Calculate mean from the following data:

Roll No.	1	2	3	4	5	6	7	8	9	10
Marks	40	50	55	78	58	60	73	35	43	48

Solution:

Here, Marks is the variable X. The calculation is tabulated as follows:

Roll No.	Marks
1	40

2	50
3	55
4	78
5	58
6	60
7	73
8	35
9	43
10	48
$N = 10$	$\sum X = 540$

$$\overline{X} = \frac{\sum X}{N} = \frac{540}{10} = 54 \text{ marks.}$$

Interpretation:

The arithmetic mean mark of the given 10 students is 54 marks.

b) Discrete Series

Step1: Multiply each size of item (x) by its frequency (f) -fx

Step2: Add all the fx-$\sum fx$

Step3: Divide $\sum fx$ by the total frequency (N = $\sum f$)

Arithmetic Mean of a discrete series is given by $\overline{X} = \frac{\sum fx}{N}$.

Example

Calculate mean from the following data:

Value	1	2	3	4	5	6	7	8	9	10
Frequency	21	30	28	40	26	34	40	9	15	57

Solution:

x	f	fx
1	21	21
2	30	60
3	28	84
4	40	160
5	26	130
6	34	204
7	40	280
8	9	72
9	15	135
10	57	570
	$N = \sum f = 300$	$\sum fx = 1716$

$$\overline{X} = \frac{\sum fx}{N} = \frac{1716}{300} = 5.72$$

Continuous Series

In continuous frequency distribution, the value of each individual frequency distribution is unknown. Therefore, an assumption is made to make them precise or on the assumption that the frequency of the class intervals is concentrated at the centre that the mid point of each class interval has to be found out.

The following procedure is to be adopted for calculating arithmetic mean in a continuous series.

Step1: Find out the mid value of each group or class. The mid value, denoted by '*m*', is obtained by adding the lower limit and upper limit of the class and dividing the total by two. For example, in a class interval 10 – 20, the mid value is 15 ($\frac{10+20}{2}=\frac{30}{2}=15$)

Step 2: Multiply the mid value of each class by the frequency of the class. That is, find *fm*.

Step 3: Add up all the products - $\sum fm$

Step 4: $\sum fm$ is divided by *N*.

Apply the formula: $\overline{X}=\frac{\sum fm}{N}$.

Example

Find out the mean profit from the following data.

Profits per shop (Rs.)	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Number of shops	10	18	20	26	30	28	18

Solution

Profits (Rs.)	Mid point (*m*)	No.of shops (*f*)	*fm*
100-200	150	10	1500
200-300	250	18	4500
300-400	350	20	7000
400-500	450	26	11700
500-600	550	30	16500
600-700	650	28	18200
700-800	750	18	13500
		$N=\sum f$ = 150	$\sum fm$ = 72900

$$\overline{X} = \frac{\sum fm}{N} = \frac{72900}{150} = 486.$$

Interpretation

The average profit is Rs. 486.

Merits of Mean

Arithmetic mean is the simplest measurement of central tendency of a series. It is widely used because

1. It is easy to understand.
2. It is easy to calculate.
3. It is used in further calculation.
4. It is rigidly defined.
5. It is based on the value of every item in the series.
6. It provides a good basis for comparison.
7. Arithmetic mean can be calculated if we know the number of items and aggregate. If the mean and the number of items are known we can find the aggregate.
8. Its formula is rigidly defined. The mean is the same for the series who ever calculates it.
9. It can be used for further analysis and algebraic treatment.
10. The mean is more stable measure of central tendency (Ideal average)

Demerits of arithmetic mean

1. The mean is unduly affected by the extreme items.
2. It is unrealistic
3. It may lead to false conclusion.
4. It cannot be accurately determined even if one of the values is not known.
5. It is not useful for the study of qualities like intelligence, honesty and character.
6. It cannot be located by observation or the graphic method.
7. It gives greater importance to bigger items of a series and lesser importance to smaller items.

Even though arithmetic mean is subject to various demerits, it is considered to be the best of all averages.

Median

Median is the value of item that goes to divide the series into equal parts. Median may be defined as the value of that item which divides the series into two equal parts, one half containing values greater than it and the other half containing values less that it. Therefore, the series has to be arranged in ascending or descending order, before finding the median. In other words, arranging is necessary to compute median.

As distinct from the arithmetic mean, which is calculated from the value of every item in the series, the median is what is called a positional average. The term position refers to the place of a value in a series.

The definitions of median given by different authors are: 'The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median'. ***– L.R.Connor.***

'Median of a series is the value of the item actual or estimated when a series is arranged in order of magnitude which divides the distribution into two parts'. - ***Secrist***

"The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or great that it".

- ***Croxton and Cowden***

"The median, as its name indicates, is the value of the middle item in a series, when items are arranged according to magnitude". - ***Yau Leun Chou***

Individual series

Step 1. Arrange the data in ascending or descending order.
Step 2. Apply the formula

$$\text{Median} = \text{Size of } \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item.}$$

Odd number of observations

Example
The following are the marks scored by 7 students, find out the median marks.

Roll numbers	1	2	3	4	5	6	7
Marks	45	32	18	57	65	28	46

Solution

Arrange the marks in ascending order.

Roll numbers	1	2	3	4	5	6	7
Marks	18	28	32	45	57	58	65

$$\text{Median} = \text{Size of } \left(\frac{N+1}{2}\right)^{th} \text{ item} = \frac{7+1}{2} = 4^{th} \text{ item}$$

$= \text{Size of } 4^{th} \text{ item} = 45$

Even number of observations

To find the median from even number of items, the median is estimated by the mean of the two middle values. It means add the two values of middle item and divide by two.

Example

Find out the median from the following data:

57 58 61 42 38 65 72 66

Solution

Arrange the marks in ascending order.

38 42 57 58 61 65 66 72

$$\text{Median} = \text{Size of } \left(\frac{N+1}{2}\right)^{th} \text{ item} = \frac{8+1}{2} = 4.5^{th} \text{ item}$$

$= \text{Size of } 4.5^{th} \text{ item}$

$$= \frac{4^{th}\text{item} + 5^{th}\text{item}}{2} = \frac{58+61}{2} = 59.5$$

Discrete series

Step 1. Arrange the data is ascending or descending order.

Step 2. Find the cumulative frequencies.

Step 3. Apply the formula

$$\text{Median} = \text{Size of } (\frac{N+1}{2})^{th} \text{ item}$$

Example
Locate the median from the following:

Size of the shoes	5	5.5	6	6.5	7	7.5	8
frequency	10	16	28	15	30	40	34

Solution
Arrange the data in ascending order.

Size of the shoes (x)	f	Cumulative frequency (cf)
5	10	10
5.5	16	26
6	28	54
6.5	15	69
7	30	99
7.5	40	139
8	34	173

$$\text{Median} = \text{Size of } (\frac{N+1}{2})^{th} \text{ item} = \frac{173+1}{2} = 87^{h} \text{ item}$$

= Size of 87^{th} item = 7

Interpretation

Therefore, median size of the shoe = 7.

Continuous series

Step 1: Find out the median by using $\frac{N}{2}$.

Step 2: Find out the class in which median lies.
Step 3: Apply the formula:

$$\text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times i, \text{ where}$$

L = Lower limit of the median class (middle class)
f = Frequency of median class

cf = Cumulative frequency of the class preceding median class

(Total frequency of all lower class)

i = Class interval of median class.

Example

Calculate the median from the following table:

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Frequency	6	20	44	26	3	1

Solution

Marks (x)	f	Cumulative frequency (cf)
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100

Median item = $\frac{N}{2} = \frac{100}{2} = 50$.

Median item lies in 40-55 mark group. Therefore, median is calculated as follows.

$$\text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times i$$

Here, $L = 40$; $\frac{N}{2} = 50$; $cf = 26$; $f = 44$; $i = 15$.

$$\text{Median} = 40 + \frac{50 - 26}{44} \times 15$$

$$= 40 + 8.18$$

$$= 48.18 \text{ marks.}$$

Merits of median

1. It is easy to understand and easy to compute.
2. It is quite rigidly defined.
3. It eliminates the effect of extreme items.
4. It is amenable to further algebraic process.
5. Since it is positional average, median can be computed even if the items at the extremes are unknown.

6. Median can be calculated even from qualitative phenomena i.e. honesty, character etc.
7. Median can sometimes be known by simple inspection.
8. Its value generally lies in the distribution.

Demerits of median

1. Typical representative of the observations cannot be computed if the distribution of item is irregular. For example, 1, 2, 3, 100 and 500, the median is 3.
2. When the number of items is large, the prerequisite process (Arraying the items) is a difficult process.
3. It ignores the extreme items.
4. In case of continuous series, the median is estimated, but not calculated.
5. It is more affected by fluctuations of sampling than in mean.
6. Median is not amenable to further algebraic manipulation.

Mode

Mode is the most common item of a series. Mode is the value which occurs the greatest number of frequency in a series. According to Croxton and Cowdon, "The mode of a distribution is the value at the point around which the item tend to be most heavily concentrated".

The chief feature of mode is that it is the size of the item which has the maximum frequency and is also affected by the frequencies of the neighbouring items.

Individual observation

Mode can often be found out by mere inspection in case of individual observations. The data have to be arranged in the form of an array so that the value which has the highest frequency can be known. For example, 10 persons have the following income:

Rs. **850**, 750, 600, 825, **850**, 900, 630, **850**, 640, 530, **850**

850 repeats four times, therefore the mode salary is Rs.850.

Discrete series

We cannot depend on the method of inspection to find out the mode. In such situations, it is suggested to prepare a grouping table and an analysis table to find out the mode. First we prepare grouping table and then an analysis table.

Step 1: Prepare a grouping table with 6 columns.

Step 2: Write the size of the item in the margin.

Step 3: In column 1, write the frequencies against the respective items.

Step 4: In column 2, the frequencies are grouped in twos. (1 and 2, 3 and 4, 5 and 6 and so on).

Step 5; In column 3, the frequencies are grouped in twos, leaving the first frequency. (2 and 3, 4 and 5, 6 and 7 and so on).

Step 6: In column 4, the frequencies are grouped in threes. (1, 2 and 3; 4, 5 and 6; 7,8, and 9; and so on)

Step 7: In column 5, the frequencies are grouped in threes, leaving the first frequency.(2, 3 and 4; 5, 6 and 7; 8, 9 and 10 and so on).

Step 8: In column6, the frequencies are grouped in threes, leaving the first two frequencies. (3, 4 and 5; 6, 7 and 8 so on). In all the processes mark down the maximum frequencies by bold letters or by a circle.

Step 9: After grouping the frequencies table, an analysis table is prepared to show the exact size, which has the highest frequency.

Example

Calculate mode from the following:

Size	10	11	12	13	14	15	16	17	18
frequency	10	12	15	19	20	8	4	3	2

Solution

Grouping Table

size	1	2	3	4	5	6
10	10					
		22				
11	12			37		
			27			
12	15				46	
		34				
13	19					54
			39			
14	20			47		
		28				
15	8				32	
			12			
16	4					15
		7				
17	3			9		
			5			
18	2					

Analysis Table

Column number	Size of item containing maximum frequency					
	10	11	12	13	14	15
1					1	
2			1	1		
3				1	1	
4				1	1	1
5		1	1	1		
6			1	1	1	
Total		1	3	**5**	4	1

The mode is 13, as the size of the item repeats 5 times. But through inspection, it is seen that mode is 14, because the size 14 occurs 20 times. But this wrong decision is revealed by analysis table.

Continuous series

In a continuous series, to find out the mode, we need one step more than those used for discrete series. As explained in the discrete series, modal class is determined by preparing grouping table and analysis tables. Then we apply the following formula:

$$Z = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i \text{ , where}$$

Z = Mode

L = Lower limit of the modal class

f_1 = Frequency of the modal class

f_0 = Frequency of the class preceding the modal class

f_2 = Frequency of the class succeeding the modal class

i = Class interval

Example

Compute the mode from the following data:

Size of item	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Frequency	20	24	32	28	20	16	34	10	8

Solution

Grouping Table

Size of item	1	2	3	4	5	6
0-5	20					
		44				
5-10	24			76		
			56			
10-15	32				84	
		60				
15-20	28					80
			48			
20-25	20			64		
		36				
25-30	16				70	
			50			
30-35	34					60
		44				
35-40	10			52		
			18			
40-45	8					

Analysis Table

Column number	Size of item containing maximum frequency					
	0-5	5-10	10-15	15-20	20-25	25-30
1						1
2			1	1		
3		1	1			
4	1	1	1			
5		1	1	1		
6			1	1	1	
Total	1	3	**5**	3	1	1

From the above analysis table, it is found that the mode is 10-15 as its frequencies occur the maximum times. Mode is estimated by the formula

$$Z = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here, $L = 10$; $f_1 = 32$; $f_0 = 24$; $f_2 = 28$; $i = 5$.

Therefore, Mode = $Z = 10 + \frac{32-24}{2(32)-24-28} \times 5$

$$= 10 + \frac{8}{64 - 24 - 28} \times 5$$

$$= 10 + \frac{40}{12}$$

$$= 10 + 3.33$$

= 13.33

Merits

1. It is easy to understand as well as easy to calculate. In certain cases, it can be found out by inspection.
2. It is usually an actual value as it occurs most frequently in the series.
3. It is not affected by extreme values as in the average.
4. It is simple and precise.
5. It is the most representative average.
6. The value of mode can be determined by the graphic method.
7. Its value can be determined in an open end class-interval without ascertaining the class limits.

Demerits

1. It is not suitable for further mathematical treatment.
2. It may not give weight to extreme items.
3. In a bimodal distribution there are two modal classes and it is difficult to determine the value of the mode.
4. It is difficult to compute; when there are both positive and negative items in a series and when there one or more items are zero.
5. It is stable only when the sample is large.
6. Mode is influenced by magnitude of the class-intervals.
7. It will not give the aggregate value as in average.

Relationship between Mean, Median and Mode

- In a symmetrical distribution, the three averages are equal.

  Mean = Median = Mode

- In a asymmetrical distribution,

  Mean – Mode = 3 (Mean - Median).

Measures of Dispersion

While the measures of central tendency gives us one single figure that represents the entire data, the measures of dispersion, also known as measures of central distendency, attempts to measure the degree of spread, or dispersion around the central position.

According to Connor, "Dispersion is a measure of the extent to which the individual items vary". Brooks and Dick define it "as the degree of scatter or variation of the variables about a central value". Spiegel writes that "the degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data".

Objectives of the measures of dispersion

1. It is used to compare two or more series with regard to their variability.
2. It is used to control the variability itself.
3. It is used to judge the reliability of the measures of the central tendency.
4. It is used to facilitate the use of other statistical measures.
5. It is used in quality control and time series

Characteristics of good measures of dispersion

The measures of dispersion should have all the characteristics prescribed for good measures of central tendency.

1. It should be rigidly defined.
2. It should be easy to comprehend.
3. It should be easy to compute.
4. It should be based on all the observations of a series.
5. It should be capable of further algebraic calculation.

Types of measures of dispersion

Two categories of measures of dispersion are identified. The first type (The Range) is based on the difference between the two extreme scores in a distribution. This is known as the Distance Measures. The second types are those based on average deviation from the measures of central tendency.

The emphasis in this course is given to the following four measures of dispersion.

1. Range
2. Quartile Deviation

3. Mean Deviation
4. Standard Deviation

Range

The Range is an easy way to indicate Dispersion. It is the difference between the largest and the smallest value included in a distribution.

For example the Range of the series 51, 57, 31, 35, 95 is equal to

Largest value – Smallest value = 95 – 51 = 44.

Range is very simple to understand, and easy to compute. It is quite useful as a rough measure of dispersion for a set of limited items.

Limitations

1. It is not rigidly defined.
2. It has no further algebraic properties.
3. It is not stable.
4. It does not take into account the entire distribution.
5. It cannot be computed from frequency distribution with open classes.

In spite of these limitations, it is used in studying quality control, fluctuations in share prices, weather forecasting etc.,

Quartile Deviation

Quartile Deviation, also known as Semi-Inter Quartile Range, is a measure of dispersion based on the difference between the values of the first quartile and third quartile divided by two. This is represented by the following formula.

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2}$$

The First Quartile or the Lower Quartile, denoted by Q_1 in the above formula, is the value which leaves one-fourth of the value below it, and the Third Quartile or the Upper Quartile, denoted by Q_3 in the above formula, is the value which leaves three-fourths of the value below it.

Example

The following are the marks obtained by 11 students in a test. The marks are arranged in the ascending order.

25 28 33 37 42 45 48 51 54 57 60

Q_1 = Size of $(\frac{N+1}{4})^{th}$ item $= \frac{11+1}{4} = 3^{rd}$ item

= Size of 3^{rd} item

= 33 marks.

Q_3 = Size of $3(\frac{N+1}{4})^{th}$ item $= 3(\frac{11+1}{4}) = 3 \times 3 = 9^{th}$ item

= Size of 9^{th} item

= 54 marks.

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{54-33}{2}$$

= **10.5**

Merits

1. It is simple to understand and easy to compute.
2. It is not influenced by the extreme values.
3. It can be found out with open end distribution.
4. It is not affected by presence of extreme values.

Demerits

1. It ignores the first 25% of the items and the last 25% of the items.
2. It is a positional average; hence not amenable to further mathematical treatment.
3. Its value is affected by sampling fluctuations.
4. It gives only a rough measure.

Mean Deviation

The Mean Deviation (M. D.) of a series is the arithmetic average of the deviations of various items from a measure of central tendency, may be mean, median or mode. It is also called as 'Average Deviation' as it is the average amount of scatter of the items in a distribution from

either the mean or the median or the mode, ignoring the sign of the deviation. This is represented by the following formula.

$$\text{M. D.} = \frac{\sum|D|}{N}, \text{ where}$$

$\sum|D|$ = sum of the deviations

D = deviations of the items from mean or median or mode, ignoring plus or minus signs.

N = the number of the observations.

Example

Calculate M.D. for the following data. (From the mean)

100 150 200 250 360 490 500 600 671

Solution

x	$\|D\| = x - \bar{x}$ $= x - 369$
100	269
150	219
200	169
250	119
360	9
490	121
500	131
600	231
671	302
$\sum x =$ 3321	$\sum\|D\| = 1570$

$$\text{Mean } \bar{x} = \frac{\sum x}{N}$$

$$= \frac{3321}{9} = 369$$

$$\text{Mean Deviation from Mean} = \frac{\sum|D|}{N} = \frac{1570}{9} = 174.44$$

Merits

1. It is simple to understand and easy to compute.
2. M.D. is a calculated value.
3. It is not much affected by the fluctuations of sampling.
4. It is based on all the items of the series and gives weight according to their size.

5. It is less affected by extreme items.
6. It is rigidly defined.
7. It is flexible, because it can be calculated from any measure of central tendency.
8. It is a better measure for comparison.

Demerits

1. It is a non-algebraic treatment.
2. Algebraic positive and negative signs are ignored. In mean deviation, +5 and -5 have the same meaning. It is mathematically unsound and illogical.
3. It is not a very accurate measure of dispersion.
4. It is not suitable for further mathematical calculation.
5. It is rarely used. It is not as popular as standard deviation.

Standard Deviation

Standard Deviation (S.D) is the most important and widely used measure of dispersion. It is defined as the square root of the arithmetic mean of the squares of deviations from the arithmetic mean. This measure is calculated from the arithmetic mean alone and not from median or mode like the mean deviation. It is to be noted that in the case of S.D. the algebraic signs are not ignored, and therefore it is mathematically sound. The S.D. is denoted by the Greek letter σ (Sigma).

Individual Observation

Steps for calculating S.D.

Step 1: Find out the actual mean of the series. ($\overline{X}$)

Step 2: Find out the deviation of each value from the mean. ($x = X - \overline{X}$)

Step 3: Square the deviations and take the total of squared deviations.($\sum x^2$)

Step 4: Divide the total $\sum x^2$ by the number of observations (N). The square root of the quotient is standard deviation. Thus apply the following formula:

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

Example

Calculate the standard deviation from the following data.

14 22 9 15 20 17 12 11

Solution

Values (X)	$x = X - \overline{X}$ (X-15)	x^2
14	-1	1
22	7	49
9	-6	36
15	0	0
20	5	25
17	2	4
12	-3	9
11	-4	16
$\sum X$ =120		$\sum x^2$ =140

Mean: $\overline{X} = \frac{\sum X}{N} = \frac{120}{8} = 15$

Standard Deviation: $\sigma = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\frac{140}{8}} = \sqrt{17.5} = \mathbf{4.18}$

Discrete series

Step 1: Calculate the mean of the series.

Step 2: Find deviations for various items from the mean. That is, $X - \overline{X} = d$

Step 3: Square the deviations (d^2) and multiply by the respective frequencies (f). We get fd^2.

Step 4: Total the product $\sum fd^2$. Then apply the formula:

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$$

Example

Calculate standard deviation from the following:

Marks	10	20	30	40	50	60
No. of students	8	12	20	10	7	3

Solution

Marks (X)	f	fX	$X-\overline{X}=d$ $\overline{X}=30.8$	d^2	fd^2
10	8	80	- 20.8	432.64	3461.12
20	12	240	- 10.8	116.64	1399.68
30	20	600	- 0.8	0.64	12.80
40	10	400	9.2	84.64	846.40
50	7	350	19.2	368.64	2580.48
60	3	180	29.2	852.64	2557.92
$\sum X$ =210	N =60	$\sum fX$ =1850			$\sum fd^2$ =10858.40

Mean: $\overline{X}=\frac{\sum fX}{N}=\frac{1850}{60}$ =30.8 marks

Standard Deviation: $\sigma=\sqrt{\frac{\sum fd^2}{N}}=\sqrt{\frac{10858.40}{60}}=\mathbf{13.45}$ marks

Continuous Series

In the continuous series, the method of calculating standard deviation is almost the same as in a discrete series. But in a continuous series, mid values of the class intervals are to be found out. The formula is

$$\sigma=\sqrt{\frac{\sum fd'^2}{N}-\left(\frac{\sum fd'}{N}\right)^2}\times C$$

where $d'=\frac{m-A}{C}$, C = interval length (common factor).

Step 1: Find out the mid value of each class (m).

Step 2: Assume one of the mid values as an average and denote it by A.

Step 3: Find out deviation of each mid value from the assumed average A and denote these deviations by d.

Step 4: If the class intervals are equal, then take a common factor. Divide each deviation by the common factor and denote this column by d'.

Step 5: Multiply these deviations d' by the respective frequencies and get $\sum fd'$.

Step 6: Square the deviations and get d^2

Step 7: Multiply the squared deviation d^2 by the respective frequencies (f). Then obtain the total $\sum fd^2$.

Step 8: Substitute the values in the following formula to get the standard deviation.

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C$$

Example

Compute the standard deviation from the following data.

Class (x)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency (f)	8	12	17	14	9	7	4

Solution

Class (x)	Mid-value (m)	$d' = \frac{m-A}{C} = \frac{m-35}{10}$	f	fd'	fd'^2
0-10	5	$\frac{5-35}{10} = -3$	8	-24	72
10-20	15	-2	12	-24	48
20-30	25	-1	17	-17	17
30-40	35	0	14	0	0
40-50	45	1	9	9	9
50-60	55	2	7	14	28
60-70	65	3	4	12	36
			$N = 71$	$\sum fd = -30$	$\sum fd'^2 = 210$

Standard Deviation: $\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C$

$$= \sqrt{\frac{210}{71} - \left(\frac{-30}{71}\right)^2} \times 10$$

$$= \sqrt{2.957 - (-0.4225)^2} \times 10$$

$$= \sqrt{2.957 - 0.1785} \times 10$$

$$= \sqrt{2.7785} \times 10$$

$$= 1.667 \times 10 = \mathbf{16.67}$$

Correlation Analysis

Correlation refers to the relationship of two or more variables. We can find some relationship between two variables; for example, there exists some relationship between the height of a father and the height of the son, price and demand, wage and price index, yield and rainfall, height and weight and so on. Correlation is the statistical analysis which measures and analyses the degree or extent to which two variable fluctuate with reference to each other. The word 'relationship' is of important and indicates that there is some connection between the variables under observation. The correlation measures the closeness of the relationship between the variable. For instance, the height and the weight are correlated because both will be increasing or decreasing.

Definition

According to Ya Lun Chou, "Correlation analysis attempts to determine the degree of relationship between variables."

According to W.I.King, "Correlation means that between two series or group of data, there exists some casual connection."

According to L.R.Connon, "If two or more quantities vary in sympathy, so that movements in one tend to be accompanied by corresponding movements in the other(s), they are said to be correlated."

The degree of correlation is measured by the Coefficient of Correlation. The coefficient of correlation is a ratio which expresses the extent to which the changes in one variable are accompanied by the changes in concerned variable.

Characteristics of Correlation

a) Changes in one series of data is accompanied by or followed by corresponding change in other series

b) The change may be in same direction, that is, positive change, or may be in the opposite direction, that is, the negative direction.

c) The change is based on cause and Effect relationship.

Types of Correlation

a) Positive and negative correlation;

b) Simple and multiple correlation;

c) Linear and Non-linear correlation;

d) Partial and Total correlation;

Positive Correlation

When the change in one phenomena results in change in another phenomena in the same direction then the relationship is known as positive correlation.

Negative Correlation

On the otherhand if the change in one phenomena results in the change in some other phenomena in the opposite direction or the reverse direction, then, the relationship is said to be negative correlation.

Simple Correlation

Simple correlation shows the correlation between more than two variables. It may be positive or negative.

Multiple Correlation

Multiple Correlation shows the correlation between more than two variables. It may be positive or negative.

Linear Correlation

Linear correlation is said to exist when the amount of change in one variable, followed by the amount of change in other variable tends to bear a constant ratio.

Non-Linear Correlation

In non-linear correlation the ratio of the amount of change in one variable to the amount of change in another variable is not constant.

Partial and Total Correlation

In partial correlation the relationship of two or more variable is examined, while other variables included for calculation of total correlation are excluded. The total correlation is based on all the relevant variables.

Measurement of Coefficient of Correlation

The degree of coefficient of correlation can be measured by the following mathematical methods:

a) Karl Pearson's Co-efficient of correlation
b) Spearman's Ranking Method
c) Concurrent Deviation Method
d) Product Moment Method

As the Pearson Correlation Coefficient and the Spearman's Correlation are commonly used measures we will study only these two methods.

Karl Pearson's Coefficient of Correlation

According to this method, the Correlation Coefficient between any two series of data is computed by dividing the product of deviations from the Arithmetic Mean by the product of the two Standard Deviations to the number of pairs in it.

The following is the formula for the Karl Pearson's Coefficient of Correlation applicable to two sets of variables (say X and Y) normally distributed.

$$r = \frac{\left(\sum XY \times N\right) - \left(\sum X \times \sum Y\right)}{\sqrt{\sum X^2 \times N - \left(\sum X\right)^2} \times \sqrt{\sum Y^2 \times N - \left(\sum Y\right)^2}}$$

where r – Correlation coefficient

X and Y – Two sets of variables

N – Number of observations.

Example

Calculate the coefficient of correlation from the following data:

X	12	9	8	10	11	13	7
Y	14	8	6	9	11	12	3

Solution

X	Y	X^2	Y^2	XY
12	14	144	196	168
9	8	81	64	72
8	6	64	36	48
10	9	100	81	90
11	11	121	121	121
13	12	169	144	156
7	3	49	9	21
$\sum X$ =70	$\sum Y$ =63	$\sum X^2$ = 728	$\sum Y^2$ =651	$\sum XY$ =676

Therefore, $$r = \frac{\left(\sum XY \times N\right) - \left(\sum X \times \sum Y\right)}{\sqrt{\sum X^2 \times N - \left(\sum X\right)^2} \times \sqrt{\sum Y^2 \times N - \left(\sum Y\right)^2}}$$

$$r = \frac{(676 \times 7) - (70 \times 63)}{\sqrt{728 \times 7 - (70)^2} \times \sqrt{651 \times 7 - (63)^2}}$$

$$= \frac{4732 - 4410}{\sqrt{5096 - 4900} \times \sqrt{4557 - 3969}}$$

$$= \frac{322}{\sqrt{196 \times 588}}$$

$$= \frac{322}{339.48} = \mathbf{+\ 0.95}$$

Spearman's Rank Correlation Coefficient

In 1904, Charles Edward Spearman, a British psychologist found out the method of ascertaining the coefficient of correlation by ranks. This method is based on rank. This measure is useful in dealing with qualitative characteristics, such as intelligence, beauty, morality, character, etc. It cannot be measured quantitatively, as in the case of pearson's coefficient of correlation; but it is based on the ranks given to the observations. It can be used when the data are irregular or extreme items are erratic or inaccurate, because rank correlation coefficient is not based on the assumption of formality of data.

Rank correlation is applicable only to individual observations. The result we get from this method is only an approximate one, because under ranking method original value are not taken into account. The formula for spearman's rank correlation which is denoted by *P*is;

$$P = 1 - \frac{6\sum D^2}{N(N^2 - 1)} \text{ or } P = 1 - \frac{6\sum D^2}{N^3 - N}$$

where

P = Rank coefficient of correlation

D^2 = Sum of the squares of the differences of two ranks

N = Number of paired observations.

Like the Karl Pearson's coefficient of correlation, the value of P lies between +1 and -1. If P = +1, then there is complete agreement in the order of ranks and the direction of the rank is also the same. When P = -1, then there is complete disagreement in the order of ranks and they are in opposite directions.

There are two types of problems:

- Where ranks are given
- Where ranks are not given

(1) Where ranks are given

When the actual ranks are given, the following are the steps to be followed in calculating rank correlation coefficient.

Step 1: Compute the difference of the two ranks (R_1 and R_2) and denote by *D*.

Step 2: Square the *D* and get $\sum D^2$.

Step 3: Substitute the figures in the formula.

Example

Following are the rank obtained by 10 students in two subjects, Statistics and Mathematics. To what extent the knowledge of the students in the two subjects is related?

Statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

Solution

Rank of Statistics (x)	Rank of Mathematics (y)	$D = x - y$	D^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4
			$\sum D^2 = 40$

Therefore, rank correlation $P = 1 - \frac{6\sum D^2}{N(N^2-1)}$

$$= 1 - \frac{6 \times 40}{10(10^2-1)}$$

$$= 1 - \frac{240}{10(100-1)}$$

$$= 1 - \frac{240}{990}$$

$$= 1 - 0.24 = \mathbf{+0.76}$$

(2) Where ranks are not given

When no rank is given, but actual data are given, then the ranks must be given. Ranks are given by taking the highest as 1 or lowest value as 1, next to the highest (lowest) as 2 and follow the same procedure for both variables.

Example

A random sample of 5 college students is selected and their grades in Mathematics and Statistics are found to be:

	1	2	3	4	5
Mathematics	85	60	73	40	99
Statistics	93	75	65	50	80

Calculate Spearman's rank correlation coefficient.

Solution

Marks in Mathematics (X)	Ranks (x)	Marks in Statistics (Y)	Ranks (y)	Difference $D = x - y$	D^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1
					$\sum D^2 = 4$

Therefore, rank correlation $P = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$

$= 1 - \frac{6 \times 4}{5(5^2 - 1)}$

$= 1 - \frac{24}{5 \times 24}$

$= 1 - \frac{4}{5}$

$= 1 - 0.2$

$= +\mathbf{0.8}$

Equal or repeated ranks

When two or more items have equal value, it is difficult to rank them. In that case, the items are given the average of the ranks they would have received, if they are not tied. For example, if two individuals are placed in seventh place they are each given the rank $\frac{7+8}{2} = 7.5$ which is the common rank to be assigned – and the next will be 9 – and if three are ranked equal at the seventh place, they are given the rank $\frac{7+8+9}{3} = 8$ which is the common rank to be assigned to each – and the next rank will be 10, in this case. The formula is

$$P = 1 - 6\left\{\frac{\sum D^2 + \frac{1}{12}\left(m^3 - m\right) + \frac{1}{12}\left(m^3 - m\right) + \mathrm{K}}{N(N^2 - 1)}\right\}$$

Example

From the following data, calculate the rank correlation coefficient after making adjustment for tied ranks.

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	6	19

Solution

X	Rank (x)	Y	Rank (y)	Difference $D = x - y$	D^2
48	8	13	5.5	2.5	6.25
33	6	13	5.5	0.5	0.25
40	7	24	10	-3	9.00
9	1	6	2.5	-1.5	2.25
16	3	15	7	4	16.00
16	3	4	1	2	4.00
65	10	20	9	1	1.00
24	5	9	4	1	1.00
16	3	6	2.5	0.5	0.25
57	9	19	8	1	1.00
					$\sum D^2 = 41$

Therefore, $$P = 1 - 6\left\{\frac{\sum D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)}{N(N^2 - 1)}\right\}$$

$$= 1 - 6\left\{\frac{41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)}{10(10^2 - 1)}\right\}$$

$$= 1 - 6\left\{\frac{41 + 2 + 0.5 + 0.5}{990}\right\}$$

$$= 1 - \frac{264}{990}$$

$$= 1 - 0.267$$

$$= + \mathbf{0.733}$$

Merits

1. It is simple to understand and easy to calculate.
2. It is very useful in the case of data which are of qualitative nature, like intelligence, honesty, beauty, efficiency, etc.
3. No other method can be used when the ranks are given, except this.
4. When the actual data are given, this method can also be applied.

Demerits

1. It cannot be used in the case of bi-variate distribution.
2. If the number of items is greater than, say 30, the calculation becomes tedious and requires a lot of time. If we are given the ranks, then we can apply this method even though *N* exceeds 30.

Regression Analysis

In Correlation, two variables are related and are studied in terms of existing casual relationship between the two variables. However, we may be also interested in estimating or predicting the value of one variable given the value of another. For example, if we know that advertising and sales are correlated, we may find out the expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales. Similarly, if we know that the yield of rice and rainfall are closely related, we may find out the amount of rain required to achieve a certain production figure. Statistical tool with the help of which we are in a position to estimate the unknown values of one variable is called "regression." Thus regression reveals the average relationship between two variables and this makes possible estimation or prediction.

Definition

According to Blair, "Regression is the measure of the average relationship between two or more variable in terms of the original units of the data."

According to Taro Yamane, “One of the most frequently used techniques in economics and business research, to find a relation between two or more variable that are related casually, is regression analysis.”

According to Wallis and Robert, “It is often more important to find out what the relation actually is, in order to estimate or predict one variable (the dependent variable), and statistical technique appropriate in such a case is called Regression Analysis.”

According to Y-Lun-Chow, “Regression analysis is a statistical device. With the help of the regression analysis, we can estimate or predict the unknown values of one variable from the known values of another variable. In the regression analysis, the independent variable is also known as the “regressor” or “predictor” or “explanatory and the dependent variable is known as ‘regressed’ or ‘explained’ variable.

Uses of Regression Analysis

1. Regression analysis is used in statistics in all those fields where two or more relative variables are having the tendency to go back to the average. It is used more than the correlation analysis in many scientific studies. It is widely used in social sciences like economics, natural and physical sciences. It is used to estimate the relation between two economic variables like income and expenditure. Thus it is highly valuable tool in economics and business. Most of the economics are based on cause and effect relationship. It is very useful for prediction purposes.

   In business also, it is very helpful to study business predictions. Cost of production is affected by the sale of production. Economists have discovered many predictions and theories on the basis of the regression.
2. Regressionanalysis predicts the value of dependent variablesfrom the values of independent variables.
3. The regression analysis is highly useful and the regression line equation helps to estimates the valueof dependent variable, when the values of independent variables are used in the equation.
4. We can calculate coefficient of correlation (r)and coefficient of determination (r^2) with the help of regression coefficient.
5. Regression analysis in statistical estimation of demand curves, supply curves, production function, cost function, consumption functions etc., can be predicted.

Correlation and Regression

The difference between correlation and regression are:

Sl.No.	Correlation	Regression
1.	Correlation is the relationship between two or more variable, which vary in sympathy with the other in the same or the opposite direction.	Regression means going back and it is a mathematical measure showing the average relationship between two variables.
2.	Both the variables x and y are random variables.	Here x is the random variable and y is a fixed variable. Sometimes both the variables may be random variables.
3.	It finds out the degree of relationship between two variables and not the cause and the effect of the variable.	It indicates the cause and effect relationship between the variables and establishes a function relationship.
4.	It is used for testing and verifying the relation between two variables and gives limited information.	Besides verification, it is used for the prediction of one value, in relationship to the other given value.
5.	The coefficient of correlation is a relative measure. The range of relationship lies between 1.	Regression coefficient is an absolute figure. If we know the value of the independent variable, we can find the value of the dependent variable.
6.	There may be nonsense correlation between two variables.	In regression there is no such nonsense regression.
7.	It has limited application, because it is confined only to linear relationship between the variables.	It has wider application, as it studies linear and non-linear relationship between the variables.
8.	It is not very useful for further mathematical treatment.	It is widely used for further mathematical treatment.
9.	If the coefficient of correlation is positive, then the two variables are positives correlated and vice versa.	The regression coefficient explains that the decrease in one variable is associated with the increase in the other variable.
10.	It is immaterial whether X depends upon Y or Y depends upon X.	There is a functional relationship between the two variables so that we may identify between the independent and dependent variables.

Regression Line

In the graphical jargon, a regression line is a straight line fitted to the data by the method of least squares. It indicates the best possible mean value of one variable corresponding to the mean value of the other. Since a regression line is the line of best fit, it

cannot be used conversely: therefore, there are always two regression lines constructed for the relationship between two variables, say, *X* and *Y*. Thus, one regression line shows regression of *X* upon *Y*, and the other shows the regression of *Y* upon *X*.

Regression Equations

Regression equation is an algebraic method. It is an algebraic expression of the regression line. It can be classified into regression equation, regression coefficient, individual observation and group distribution.

Regression Equation of X on Y

Regression equation of *X* on *Y* is given by

$$X = a + b\,Y.$$

To find out the values of *a* and *b*, the following equations, called normal equations, can be used.

$$\sum Y = Na + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2 .$$

Regression Equation of Y on X

Regression equation of *Y* on *X* is given by

$$Y = a + b\,X.$$

To find out the values of *a* and *b*, the following equations, called normal equations, can be used.

$$\sum X = Na + b\sum Y$$

$$\sum XY = a\sum Y + b\sum Y^2 .$$

Example

A librarian wants to find out a suitable criteria – a convenient measure to fix the monthly wages of, let us say semi-professionals working in the library. He collects the data about seven such semi-professionals as to their present monthly income and their years of service.

Years of service(*X*)	11	7	9	5	8	6	10
Income (*Y*) (in 100's)	10	8	6	5	9	7	11

Solution

Regression Equation of Income (Y) on Years of service(X) :

Regression equation of Y on X is given by

$$Y = a + b\,X. \qquad \text{……..(A)}$$

X	Y	X^2	Y^2	XY
11	10	121	100	110
7	8	49	64	56
9	6	81	36	54
5	5	25	25	25
8	9	64	81	72
6	7	36	49	42
10	11	100	121	110
$\sum X = 56$	$\sum Y = 56$	$\sum X^2 = 476$	$\sum Y^2 = 476$	$\sum XY = 469$

To find a and b:

The normal equations are $\sum Y = Na + b\sum X$

$$\sum XY = a\sum X + b\sum X^2 .$$

Here, the number of observations $N = 7$.

Substituting the values from the table, we have

$$56 = 7a + 56b \qquad \text{……} \quad (1)$$

$$469 = 56a + 476b \qquad \text{……} \quad (2)$$

Multiplying both sides of the equation (1) by 8, we get

$$448 = 56a + 448b \qquad \text{……} \quad (3)$$

Subtracting the equation (3) from the equation (2), we get

$$21 = 28b$$

$$b = \frac{21}{28} = 0.75 \qquad \mathbf{b = 0.75}$$

Substituting the value of b, in equation (1), we get

$$56 = 7a + 56 \times 0.75$$

$$= 7a + 42$$

This implies, $56 - 42 = 7a$

$14 = 7a$

$a = \frac{14}{7} = 2$ **a= 2**

Substituting the values of a and b in the regression equation (A), we get

$$Y = a + b\,X$$

$$Y = 2 + 0.75\,X \qquad \ldots\ldots\ldots\ldots\ldots(4)$$

Suppose the Librarian wants to find out the wage for a semi-professional whose service is 13 years, then $X = 13$.

Therefore, substituting $X = 13$ in equation (4), we have

$$Y = 2 + 0.75 \times 13$$

$Y = 2 + 9.75$

Y = 11.75

Hence, the wage to be fixed = $11.75 \times$ Rs.100

= **Rs.1175.00**

Similarly, the Librarian can use this method to predict the years of service, when he knows the income a particular semi-professional using the regression equation of X on Y. Further, it can be used to prepare the budget, estimate the number of users in future.

CHAPTER – X

Research Report Writing

Preparation and presentation of the research report is the most important part of the research process. The purpose of a research is to find out the truth or some thing new, and make it known to others who are interested. Reporting of the research is the final step in the process of research. A research report is a detailed description of 'what has been done' and 'how it has been done' with respect to a research.

The purpose of the report is "to convey to the interested persons the whole result of the study in detail and so arranged as to enable each reader to comprehend the data and to determine for himself the validity of the conclusion.

Need for Research Report

Research is not restricted to an institution, community, group or society. Hence, it should be exhibited in such a way that any reader will be interested in it to foster his knowledge. Research reports are added to their concerned subject literature and stored properly for future timely retrieval. Scientists need to communicate with other so that the research experience is to reach its target users. Its aim is the dissemination of knowledge and helping further research. The new findings, new methods or new data are made known through reports. The report may help a researcher by some new light, leading to new hypothesis and new theories.

Target Users

While writing research report, it is essential to keep in mind the level of knowledge of the readers. Those who are find more useful of the results of the particular research named as target users. Target users for any research report may be any of the following:

1. Academic community
2. Sponsors of research
3. General public

While writing the report it is essential to keep in mind the level of the readers, experience and the interest of the target audience.

Types of Research Reports

Depending upon the users, the research report may be divided into two categories. These are as follows.

1. Technical Report
2. Popular Report

Technical Report:

It is generally intended for other researchers or for research managers. The report should enable another researcher to critique methodology, check calculations and accuracy and to follow everything which is done on a step-by-step basis.

Popular Report:

The popular report is intended for more general audience. Compared to the technical report, the presentation will be more attention to headlines, flow diagrams, charts, tables and summaries for the purpose of stressing major points.

Guidelines for Writing a Report

While writing a report, the following guidelines may be considered for good report.

Consider the Audience:

Make the report clear; use only words familiar to the readers and define all technical terms. To make the comparison of figures easier, use percentages, rounded off figures, ranks or ratios; put the exact data in a table within the text or in the appendix. Use graphic aids wherever they help clarify the presentation of data.

Address of the Information Needs:

Remember the research report is designed to communicate information to decision makers. Make sure that it clearly relates the research findings to the objectives of the research

Be Concise, Yet complete:

Most of the supervisors are not to read about the details of a research report. It is difficult task what to include and what to leave out. So, the researcher should provide necessary items while writing the reports.

Be Objective:

The findings may be conflict with the practical experience. In these circumstances, the researcher may present the findings in an objective manner without bias.

Contents and form of reporting

The contents and the form of reporting may not be the same for all the target audience. According to Wilkinson and Bhandarkar, a research report for an academic community, should have the following contents.

1. the statement of the problem with which the study is concerned
2. the research procedures:
    a. the study design
    b. the method of manipulating the independent variables
    c. the universe and nature of the sample
    d. data collection techniques
    e. method of statistical analysis
3. the results of major findings
4. the implications of research results for theory and practice

The above format gives the minimum requirements of a research report and does not give comprehensive coverage to all aspects in research reporting.

Shah gives the comprehensive list of Contents for a research report as shown below:

a. Title page
b. Foreword
c. Table of Contents
d. Introduction
e. Identification, selection and formulation of the problem
f. Research design and data collection
g. Data processing and analysis
h. Findings
i. Summary
j. Appendices
k. Bibliographical references

Out of the many more prescriptions available for research reporting, the prescriptions of Shah is more comprehensive list of contents, though he himself states that, "this is only a brief indication of the organization of various items to be included".

The united Nations Statistical Office has suggested the following major topics should contain in the research report:

A. Introduction

1. Clear-cut statement as to the nature of the study
2. Aims
3. Sources of information
4. Scope of the study

B. Brief statement of the working hypothesis

C. Definitions of units of study

D. Brief statement of the techniques followed in the study

1. types of observation used and conditions under which observations were made
2. types of schedules formulated and conditions under which information was secured
3. types of case history secured; their sources, manner of presentation and preliminary analysis made

4. sampling procedures and conditions of selection and testing for appropriateness, representative ness and errors
5. statistical procedures, sources of statistical data, conditions under which they were obtained
6. types of scaling techniques used

E. Brief description of experimental treatment of data and techniques used in experiments

F. Major findings of the study

G. Major conclusions reached

H. Special remarks. These are as follows:

1. problems encountered in gathering data, classifying them and analyzing them
2. possible discrepancies in the data collected
3. suggestions for further research

I. Bibliographical references with annotation

J. Appendices:

1. Sample questionnaires
2. Transcription sheets
3. Sample interview etc

The above three outlines give you a fairly good idea of the components of a good research report. In addition to that, there are some techniques such as the way of presentation, style of writing and the spirit with which it is presented.

Components of reporting

Introduction:

Usually, introduction contains a brief narration about the organization of the materials in the report. The readers are briefly introduced to the research problem, its scope, its theoretical and practical importance etc

Materials and methods:

Any investigation is necessary employed a suitable method. Choice of a suitable method is one of the responsible in the research design. This is the primary importance in reporting the research. Full details regarding materials and methods should be given. The main purpose of materials and methods may provide enough details to another competent worker. Sources should be listed and special characteristics such as age, sex, genetic and physiological status

should be described. The methodology adopted for social investigations should be given elaborately.

Style of Writing:

The basic qualities of good scientific writing are accuracy and clarity. The following points should be remembered while writing the report

1. Write clearly. Sentences must be simple. Avoid complicated sentences. Writing long paragraph should be avoided. Appropriate subheadings should be provided wherever necessary
2. Defined scientific terms and use them consistently. The target audience and their knowledge of technical terms should be considered
3. Adequate attention should be paid to use the correct spelling and grammar
4. As far as possible, present tense should be used
5. Direct and positive sentences should be used. Long, technical or unusual words or phrases should be avoided
6. All chapters and sections, subsections, tables, diagrams and charts should be labeled adequately.
7. Use footnotes and labeled serially.

Construction of Tables

In a research report, it is necessary to arrange and present the data in a certain logical manner. The collected data must be edited, coded and checked so as to develop arrays of values particularly for variables. Tabulation is an art of arranging the data in columns and rows in such a way as to facilitate comparison of data and show the relationships between them graphically. It is a great help in analysis and interpretation of data. There are two types of tables.

1. General purpose table
2. Special purpose table

A general purpose table is a large amount of data presented in an easily accessible convenient form. The repository tables, original tables, primary tables and reference tables are come under this category. A census report is a good example of a general purpose table.

A special purpose table is a secondary table derived from the general purpose table. The data of a special purpose table may be grouped, averaged, rounded, and derived for the purpose

of classification and emphasis. These tables are known as presentation tables, analytical summary tables, interpretive, derivative or secondary tables are come under this category.

Format of a Table

A good statistical table is an art. The following parts must be present in all tables.

1. Table number
2. Title
3. Head note
4. Caption
5. Stubs
6. Body of the table
7. Foot-note
8. Source-note

Table number: A table should always be numbered for identification and reference in the future. Each column should also be numbered as shown in the illustration.

Title of the table: Each table should be given a suitable title. It must be written on the top of table. It must describe the contents of the table. It must explain

1. what the data are
2. where the data are
3. time or period of data
4. how the data are classified etc

Head note: It is a statement, given below the title and enclosed in brackets. For example, the unit of measurement is written as a head-note, such as 'in millions' or 'in crores'.

Captions: These are headings for the vertical columns. They must be brief and self-explanatory. They have main heading and sub-headings are wider than columns.

Stubs: These are the headings or designation for the horizontal rows. Stubs are wider than columns.

Body of the title: It contains the numerical information. It is the most important part of the table. The arrangement in the body is generally from left to right in rows and from top to bottom in columns.

Foot-note: If any explanation or elaboration regarding any item is necessary, foot notes should be given.

Source-note: It refers to the source from where information has been taken. It is useful to the reader to check the figures and gather additional information.

STRUCTURE OF A TABLE

Number

Title

(Head-note if any)

Stub Heading	Caption	Total
	Col. Heading Col. Heading Col. Heading	
Stub entries Total	Body	

Foot-note: *Source*

General Rules for Tabulation

1. The table should be simple and compact. It should not be overloaded with details.
2. The captions and stubs in the tables should be arranged in a systematic manner. It must be easy to read the important items. There are many types. They are alphabetical, chronological, geographical, conventional etc.
3. It should suit the purpose of investigation.
4. The unit of measurement should be clearly defined and given in the tables; for example, height in metres, weight in kilograms, etc.
5. figures may be rounded off to avoid unnecessary details in the table. But a foot-note must be given to this effect.
6. Suitable approximation may be adopted.
7. A miscellaneous column should be added to include unimportant items.
8. A table should be complete and self-explanatory.
9. A table should be attractive to draw the attention of readers.
10. As it forms a basis for statistical analysis, it should be accurate and free from all sorts of errors.

11. Abbreviations should be avoided.
12. Do not use ditto marks that may be mistaken.
13. Proper lettering will help to adjust the size of the tables.
14. If it is a big table, it will lose its simplicity and understandability; and in such a case break it into two or three tables.

Pictorial representation

Most commonly used pictorial devices are as follows.

1. Bar Charts
2. Line Charts (Histogram), Frequency Polygons, and Curves.
3. Segmental representation such as pie charts and component bar charts.
4. Pictographs.

Check list for Tables

1. Is the warranted?
2. Have the data been checked?
3. Does the table follow its mention in the text?
4. Should the table be included in the text or appendix?
5. Are tables numbered consecutively?
6. Are sufficient details given to interpret the table?
7. Is the table caption sufficiently detailed?
8. Does the wording of the caption correspond to that given in the list of Tables?
9. Has a consistent format been used for all tables?
10. Is the wording for the Stub box head contained in the table caption?
11. Are units of measurement stated?
12. Are abbreviations explained in the table?
13. Could able be presented more simply?
14. Are column entries correctly aligned?
15. Is footnote usage consistent?
16. Is the top of sidewise presented tables at the binding edge?
17. Is the table correctly positioned on the page?

Discussion, Interpretation and Generalization

The final stage of discussing, interpreting, and generalizing the findings follows that of analyzing data and reporting in the appropriate and tabular and/or pictorial forms. It is not usual and perhaps even not desirable to strictly compartmentalize these closely related 3 operations.

Interpretation and generalization of findings otherwise called conclusions and recommendations summarize the data so as to yield answers to the research questions. The conclusions should be logical and clear. Readers should not be left to infer their own conclusions.

Implications and Suggestions for further Research

Research stimulated by theoretical considerations may lead to new theoretical issues. This is in turn may lead to further research. Keeping this in mind, attempt should be made to spell out the implications of the findings and research necessary for further test.

Proof Reading

Your manuscript may be neat and perfect but it is the typed or printed version that counts. If the final printed report contains serious (Syntax and Semantics) your reputation may suffer a heavy damage. Sometimes a wrong punctuation or a missing letter may change the meaning of the concept. The word burrow means a hole in the ground. If you don't add 'w' at the end, 'burro' means an ass. Hence proof reading is considered more important in the process of research reporting.

Marking Corrections

Corrections should be twice, once at the point where it occurs and the other in the margin opposite where it occurs. It is the margin marks that a typist/printer works to identify the errors.

#		-	Leave space
~	trs.	-	Transpose (Change the prder of letters, words and sentences)
φ	delt	-	delete (omit) letter, word, line etc.,
Ξ		-	To bring letters or words together (join together)
N.P/F.P/P.		-	Next para/New para/paragraph
Stet (-)		-	Let is stand
L.C.		-	Change to small letter/Lower case
U.C.		-	Change to capital letter/Upper case

Rom.fig.	-	Roman figure
Cap./Caps.	-	Capital/Capitals – CHENNAI
∧	-	Insert letter, word, figure etc.,
Sp.	-	Spelling
.	-	Insert fullstop
“	-	Insert double quotes
(-)/	-	Insert hyphen
/	-	The

Conclusion

The ultimate aim of any research may find out solutions to problems or to find out the truth. But, of course, communication is the goal of research reporting. A poorly reported research result may be misinterpreted and may go unnoticed. Hence, reporting should follow as well as establish a definite standard. The grammar of research reporting discussed above should able to give proper guidance. The papers on various aspects of research in the previous should be supplementary to and complement this chapter.

Questions for Revision

1. Explain the Logic of Scientific Method.
2. Write an essay on case study.
3. What are the sources and factors influencing selection of a research problem?
4. Define hypothesis. Describe the formulation of hypothesis.
5. Emphasize the relevance of observation method in library science research.
6. Explain the techniques of report writing.
7. Discuss briefly various sampling techniques.
8. Define Correlation and explain the various types of correlation.
9. Explain the Measures of Central Tendency.
10. Define the Research Design and their importance.
11. Write Short notes:
    a. Interview method
    b. Regression Scatter
    c. Mean, Median, Mode
    d. Case Study
    e. Statistical Survey
    f. Questionnaire method
    g. Experimental design.