**Module-1**

**Introduction to Research Methodology**

Research refers to search of knowledge. It can be defined as scientific and systematic search for pertinent information on a specific topic. Research is an original contribution to the existing stock of knowledge leading to its advancement. It is the pursuit of truth with the help of study, observation, comparison and experiment. Research can also be finding solution to a problem through objective and systematic method. It is systematic approach concerning generalization and formulation of theory.

Research is purposive investigation and organized inquiry of a particular issue. 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 reaching conclusions either in the form of solutions towards a problem or in generalizations for theoretical formulation.

**Definitions:**

Redman and Mory define research as a “systematized effort to gain new knowledge”.

Advanced Learner’s Dictionary of Current English lays down the meaning of research as “a careful investigation or inquiry specially through search for new facts in any branch of knowledge.”

The encyclopedia of Social Sciences defines Research as “the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art.”

According to Kerlinger research is “systematic, controlled, empirical and critical investigation of hypothetical proposition about the presumed relations among natural phenomenon.”

According to Best “ the systematic and objective analysis and recording of controlled observations that may lead to the development of generalizations, principles or theories, resulting in prediction and possibly ultimate control of events.”

**Scientific Method:**

Research is a scientific endeavour. It refers to a procedure or mode of investigation but not just to a specific field of subject matter. It is a step-by-step procedure following logical processes of reasoning. It is an objective, logical and systematic method of analysis of phenomena devised to permit the accumulation of reliable knowledge. It is a systematized form of analysis characterized by an intellectual attitude.

**Basis of Scientific Method:**

Scientific method is based on “articles of faith” which are as follows:

1. Reliance on evidence - any statement, finding or conclusion must be based on proofs/evidence backed by data/facts.
2. Use of relevant concepts - Concepts are logical constructs or abstracts created from secure impressions, thoughts and experiences. Concepts with specific meaning have to be used in research for clarity and correct understanding.
3. Commitment to Objectivity - According to Green “objectivity is the willingness and ability to examine evidence dispassionately” It is about forming judgements or findings or conclusions based upon facts **unbiased by personal impressions.**
4. Ethical Neutrality - science does not pass normal judgement on facts. That is., it does not say they are good or bad.
5. Generalizations – generalizations must be based on commonality of events rather on isolated events. Under the surface layer of diversity must lie the thread of commonality. There is always the danger of Particularistic fallacy which is the danger of generalizing based on insufficient, incomplete or unrelated data. Generalization should be based on large body of data suitably analysed.
6. Verifiability – The findings or conclusions must be exposed to critical scrutiny. The researcher should expose the methods as to how the conclusions were arrived at, so that it can be tested and verified by others.
7. Logical Reasoning Process - Logical Reasoning Process must be followed for drawing inferences.

**Social Science Research**

Social Science research is a scientific method of exploring, analyzing and conceptualizing human life in order to extend, correct or verify knowledge of human behavior and social life. It seeks to find explanations to unexplained social phenomena, to clarify the doubtful and correct the misconceived facts of social life.

Young defines Social Science research as “ a scientific undertaking which by means of logical and systematic techniques aims to:

1. Discover new facts or verify and test old facts.
2. Analyse their sequences, interralationships and causal explanations
3. Develop new scientific tools, concepts and theories which would facilitate reliable and valid study of human behavior.”

This definition specifies all the major aims of research.

**Objectives or Significance of Research**

1. Research extends knowledge of human beings, social life and environment. Scientists and researchers build up the wealth of knowledge through their research findings.
2. Research unravels the mysteries of nature; brings to light hidden information which would not have been revealed otherwise.
3. Research establishes generalizations and general laws and contributes to theory building in various fields of knowledge.
4. Research verifies and tests existing facts and theory and help in improvement of knowledge and ability to handle situations. Merton says “…….Research plays an active role, it performs at least four major functions. It initiates, it formulates, it deflects and it clarifies theory.”
5. General laws developed through research may enable us to make reliable predictions of future events.
6. Research aims to analyse inter-relationships between variables and to derive causal explanations.
7. Applied research aims at finding solutions to problems like socio-economic problems, health problems, human relations problem in organizations etc.
8. Research also aims at developing new tools, concepts and theories for a better study of unknown phenomena.
9. Research aids planning and contributes to national development through following ways

* Research brings out factual data on prevailing situations and problems for drawing up plans and schemes on a realistic basis.
* Research studies enable the planners to evaluate strategies and choose the most appropriate strategies for development of various sectors.
* The cost-benefit evaluation studies provide valuable lessons for improving the formulation of similar projects.
* Evaluation studies of ongoing programmes point out their short comings and implementation problems and enable the planners to revise schemes and implementation strategies appropriately.
* The dissemination of research findings creates a general awareness of the country’s current situations and problems among the public which inspire them to participate in formulation and implementation of development schemes.

1. Analytical studies of internal and external environment of business and non-business organizations provide factual data for rational decision making, formulation of strategies and policies. Studies on operational problems contribute to an improvement in their performance.

**Motivation in Research**

What makes people to undertake research? This is a question of fundamental importance. The

possible motives for doing research may be either one or more of the following:

1. Desire to get a research degree along with its consequential benefits;

2. Desire to face the challenge in solving the unsolved problems, i.e., concern over practical

problems initiates research;

3. Desire to get intellectual joy of doing some creative work;

4. Desire to be of service to society;

5. Desire to get respectability.

However, this is not an exhaustive list of factors motivating people to undertake research studies.

Many more factors such as directives of government, employment conditions, curiosity about new

things, desire to understand causal relationships, social thinking and awakening, and the like may as

well motivate (or at times compel) people to perform research operations.

**Types of Research**

Research is classified into different forms on the basis of intent & methods.

The following are the different types of research.

1. Pure and Applied Research
2. Qualitative ,Quantitative and Mixed Research
3. Exploratory
4. Descriptive
5. Diagnostic
6. Evaluation
7. Action
8. Experimental
9. Historical
10. Surveys
11. Case study
12. Field study

**1. Pure/Fundamental Research or Basic Research and Applied Research**

Research can either be applied (or action) research fundamental (or pure) research Applied Research aims at finding a solution for an immediate problem facing a society or an organisation whereas Fundamental Research is mainly concerned with Generalisation and with the formulation of a theory. 'Gathering knowledge for knowledge' is termed pure research. Research studies concerning natural phenomenon, human behaviour etc are examples of Fundamental Research. But Research aims at certain conclusion facing a concrete social problems is an example of applied Research.

**Fundamental (basic or pure) research**

**1.** The main purpose of basic research is to add to the existing stock of knowledge and thus, it can be intellectually challenging.

**2.** The knowledge produced through pure research is sought in order to add to the existing body of research methods.

**3.** It is not likely to have any practical application at the present time or even in the future.

**Applied research**

**1.** Applied research is done to solve specific, practical questions facing the society.

**2.** It can be used for policy formulation, administration and understanding of a phenomenon.

**3.** It is always done according to basic research and can be carried out by academic or industrial institutions.

For example, an academic institution, such as a university, will have a specific applied research

program funded by an industrial partner interested in that program.

**2. Quantitative and Qualitative Research:**

Quantitative research is based on the measurement of quantity or amount. It is applicable to phenomena that can be expressed in terms of quantity. Qualitative research, on the other hand, is concerned with qualitative phenomenon, i.e., phenomena relating to or involving quality or kind. For instance, when we are interested in investigating the reasons for human behaviour (i.e., why people think or do certain things), we quite often talk of ‘Motivation Research’, an important type of qualitative research. This type of research aims at discovering the underlying motives and desires, using in depth interviews for the purpose. Other techniques of such research are word association tests, sentence completion tests, story completion tests and similar other projective techniques. Attitude or opinion research i.e., research designed to find out how people feel or what they think about a particular subject or institution is also qualitative research. Qualitative research is specially important in the behavioural sciences where the aim is to discover the underlying motives of human behaviour. Through such research we can analyse the various factors which motivate people to behave in a particular manner or which make people like or dislike a particular thing. It may be stated, however, that to apply qualitative research in practice is relatively a difficult job and therefore, while doing such research, one should seek guidance from experimental psychologists.

**Mixed research**

Both quantitative and qualitative researches are not exclusive. Qualitative research may end in a hypothesis that can be quantitatively tested later. Quantitative research may involve qualitative research elements. Quantitative research may answer questions, such as the extent and pattern of poverty in India, but it may not be efficient in answering questions, such as what is the experience of facing poverty, hardships, consequences and circumstances that lead to poverty. This may be answered by qualitative research. As quantitative research is generally well known, it may be useful to outline when qualitative research is needed.

**3. Exploratory Research**

This type of research is carried out at the very beginning when the problem is not clear or is vague. In exploratory research, all possible reasons which are very obvious are eliminated, thereby directing the research to proceed further with limited options.

**1.** It is generally done in the beginning of a research. It is undertaken to explore an area where little is known or to investigate the possibilities of undertaking a particular research study and is akin to feasibility study or pilot study. A ‘small-scale study’ is undertaken to decide whether it is worth carrying out a detailed investigation.

**2.** It attempts to clarify why and how there is a relationship between two or more aspects of a situation or phenomenon.

**3.** The purpose of exploratory research is to gain background information, to define terms, to clarify the problems, to develop hypothesis, to establish research priorities and objectives, and to develop questions to be answered.

**4.** It makes use of secondary data (mainly literature review), experience surveys, case studies, interviews (mainly focus groups’ interviews), projective techniques, and Delphi techniques.

The research executives must examine such questions to identify the most useful avenues for further research. Preliminary investigation of this type is called exploratory research. Expert surveys, focus groups, case studies and observation methods are used to conduct the exploratory survey.

**Uses/Objectives of exploratory research:**

1. Gain background information
2. Define concepts more clearly
3. Develop operational definitions
4. Clarify and formulate a more precise research problem and hypotheses
5. Achieve new insights or ideas into a phenomenon
6. Establish research priorities

**Purpose of Exploratory Research**

1. To generate new ideas.
2. To increase the researcher’s familiarity with the problem
3. To make a precise formulation of the problem
4. To gather information for clarifying concepts
5. To determine whether it is feasible to attempt the study

**4. Descriptive Research**

The term ‘Descriptive’ is self-explanatory and the research that describes a situation, an event and an institution is descriptive research. It describes the nature of a situation as it exists at the time of study. Descriptive research answers the questions who, what, where, when and how.....

Descriptive research is a quantitative research method. In simple words, descriptive research is all about describing the phenomenon, observing and drawing conclusions from it.

In social science and business research we quite often usethe term *Ex post facto research* for descriptive research studies. The main characteristic of this method is that the researcher has no control over the variables; he can only report what has happened or what is happening.

**Objectives of Descriptive Research**

.

* Descriptive research studies are aimed at **describing or portraying the characteristics** of a particular individual, group or a situation.

e.g. users of a product with different age, sex, education, etc.

* It is also concerned with specific predictions, with narration of facts and characteristics concerning individual, group or situation e.g. sales of a company's product in each of the next five years
* **Studying relationships between two or more variables** also falls under the scope of descriptive studies. For example, ‘study the problem of relationship between residential status of learners and their performance in university examination’.
* We can go little beyond and **study the cause and effect relationships among variables**. e.g You may try to find out the causes of drop out among the rural background students.

.

**5.Diagnostic Research**

Diagnostic research studies are carried out to determine the frequency with which something (variables) occurs or its association with something else. All the studies that are aimed to find out whether certain variables are associated by testing the hypothesis is diagnostic in nature.

It is directed towards discovering what is happening, why it is happening and what can be done about.It aims at identifying the causes of a problem and the possible solution for it.

The design concentrates on in-depth analysis of data to identify the factors that contributed to the problem and examine their interrelationships from various angles by bringing as many relevant variables as possible. The diagnostic design is concerned with the case as well as the treatment

.

**Characteristics of Diagnostic Research**

* Existence of a pre-planned research design,
* Design should be carefully planned,
* Design must be rigid and not flexible,
* It should ensure minimization of bias and maximization of reliability of the data collected.
* Done using rigid methods with clear specifications of who, what, when, where, why and how of the research.
* Frequently use pilot studies to test the data collection tool and analysis techniques.
* Data collection often done through structured interviews or questionnaires.

**6. Evaluation Research**

Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object.Evaluation Research is used to provide feedback on an event, organization, program, policy, technology, person, activity, etc. This is an example of applied research, therefore, evaluation research is made for effectiveness of social or economic programmes implemented or for assessing the impact of development projects on the development of project area. It is characterised by the focus on collecting data to ascertain the effects of some form of planned change.

**7. Action** **Research**

There are two dimensions attached to this word ‘Action Research’, where one is action, which is doing something and second is research, which is analyzing. Action research also means ‘learning by doing’.

Action research refers to a wide variety of evaluative, investigative and analytical research methods designed to diagnose problems or weaknesses, and help researchers to develop practical solutions to address them quickly and efficiently. It may also be applied to programs or educational techniques that are not necessarily experiencing any problems, but that researchers simply want to learn more about the techniques and improve their knowledge.

Action research is a process by which change and understanding can be pursued at one time. It is usually described as cyclic, with action and critical reflection taking place in turn. The reflection is used to review the previous action and plan the next one. It is commonly done by a group of people, though sometimes individuals use it to improve their practice

**Purpose of Action Research**

1. Action research aims at an increased understanding of an immediate social situation

2. Action research simultaneously assists in practical problem solving and expands scientific knowledge.

3. Action research is performed collaboratively and enhances the competencies of the respective actors.

4. Action research is primarily applicable for the understanding of change processes in social systems.

**8. Experimental** **Research**

Experimental research is designed for establishing causal relationships. It begins with a question concerning the relationship between two or more variables.

The premise of the design is that something (an independent variable) directly influences the behavior of something else (the dependent variable). The presumed cause is called the independent variable and the presumed effect is called the dependent variable. The researcher develops one or more hypotheses to state the nature of expected relationship. Thus experimental research is also known as hypothesis-testing or causal research.

The application of experimental method yielded better results in physical sciences. Therefore, this method was soon applied to other sciences like biological sciences and medicine.

In its simplest form, an experiment has three characteristics as follows:

**1.** An independent variable is manipulated.

**2.** All other variables except the independent variables are held constant.

**3.** The effect of manipulation of the independent variable on the dependent variable is observed.

The variable upon which the effects of changes are observed is called the dependent variable, which is observed but not manipulated by the experimenter. The dependent variable is so named because its value is hypothesized to depend upon and vary with the value of the independent variable.

For example, to examine the effect of different teaching methods upon achievement in reading, an investigator would manipulate method, the independent variable, by using different teaching methods in order to ascertain their effect upon achievement, the dependent variable.

**9. Historical Research**

Historical study is a study of past records and other information sources with a view to reconstructing the origin and development of an institution or a movement or a system and discovering the trends in the past. It must depend upon inference and logical analysis of record of data and indirect evidences rather than direct observation. It is described as the induction of principles through research into the past and social forces which have shaped the present.

**Objectives**

* Draw explanations and generalisations from the past trends in order to understand the present and to anticipate the future.
* It enables us to grasp our relationship with past and to plan more intelligently for the future.
* The past contains the key to the present and past and the present influences the future.
* Historical study helps us in visualising the society as a dynamic organism and its structures and functions as evolving, steadily growing and undergoing change and transformation.

**Steps in Historical Method**

* Feasibility of the study should examined.
* Selected problem should be formulated and the plan of the study should be designed.
* Sources of data should be located.
* Genuineness of the sources and validity of the facts should be tested.
* Relevant facts should be collected from the authentic sources and they should be checked and cross-verified.
* The facts should be arranged into a logical sequence and this synthesis is the basis for rebuilding of the past situation.
* By adopting induction process meaningful interpretations and generalisations should be made.

**Limitation**

* Reliable and adequate data may not be available.
* It is also difficult to test the genuineness and authenticity of the sources.
* It is difficult to establish the time order of events.
* It is difficult to perceive the real significance of the data because of their remote time period.

**10. Survey Research**

Survey is a fact finding study. It is a method of research involving collection of data directly from a population or sample thereof at particular time. It must not be confused with mere clerical routine of gathering and tabulating figures. It requires expert and imaginative planning careful analysis and rational interpretation of the findings.

Definitions

1.Mark Abraham defines survey as "a social survey is a process by which Quantitative facts are collected about the social aspects of a community composition and activities".

2. Herman N Morse defines It as "a method of analysis on scientific and orderly form for defined purpose of given social situations and activities."

**Characteristics of Survey**

1. It is a field study; It is always conducted in a natural setting

2. It seeks responses directly from the respondents.

3. It can cover a very large population.

4. A survey involves an extensive and intensive study.

5. A survey covers a definite geographical area, city, a district or a state

**Steps involved in a Survey**

The sequences of the task involved in carried out a survey from the 1st stage of planning to the Final stage of preparing the report is presented below.

a) Selection of problem and its formulation

b) Preparation of the research design.

c) Operationalisation of concepts and construction of measuring indexes and states.

d) Sampling

e) Construction of tools for collection of data and there pre-test.

f) Field work and collection of data

g) Processing of data and tabulation

h) Analysis of data

i) Reporting

**Purpose of the Survey**

1. The purpose of survey is to provide information's do government or planners or business enterprises.

2. Many enquiries aim to explain phenomenon

3. Surveys may be designed to make comparison of demographic groups.

4. Surveys are useful for making predictions

**Types of Survey**

**1. General or Specific survey**

When a survey is conducted for collecting general information about population institution or phenomenon without any particular object or hypothesis it is known as general survey. Specific survey are conducted for specific problems or for testing the validity of some theory or hypothesis.

**2. Regular and Adhoc Survey**

Some surveys are regular in nature and must be repeated after regular intervals. Such survey is called Regular Survey.

**3. Preliminary And Final Survey**

A Preliminary survey is generally known as 'Pilot study' and it is a fore run of the Final Survey. Final survey is made after the pilot study has completed.

**4. Senses and Sample Survey**

A survey make our all the units of a given universe then it is called a sense survey. If the survey covers only a fraction of the universe, then it is called sample survey.

**Advantages of Survey**

The major advantages of the survey method are

1. The versatility of the survey method is its greatest strength. It is the only practical way to

collect many types of information’s from individuals, socio-economic data, attitudes,

opinions, experience and expectations.

2. The survey method facilitates drawing generalisations about large populations on the basis

of studies of representative sample.

3. The survey method is flexible to permit the use of various methods of collection of data.

4. The survey help the researches to face unanticipated problems.

5. Survey is useful in verifying theories.

**Limitations of Survey**

1. Survey method is primarily meant for collection of data from primary sources. So its

success depends upon the willingness and co-operations of the respondents.

2. The survey method depends primarily on verbal behaviour. The respondent can give

misleading answers.

3. A sample survey is subject to sampling error.

4. There is a limit of the number of items of information that can be collected in a single

survey. There is an optimal length of time for an interview.

5. A survey is very expensive in terms of time and cost.

**11. Case Study Research:**

A case study is an in-depth comprehensive study of a person, a social group, a process, a situation, a programme, a community, an institution or any other social group. Its purpose is to understand the life cycle of the unit under study or the interaction between the factors that explain the present status or the development over a period of time.

It describes a case in terms of its peculiarities. It provides an opportunity for the intensive analysis of many specific details that are overlooked in other methods. A case study aims at studying everything about something rather than something about everything. In a case study individual representing the wholeness where in statistical study individual disappears.

Case study research is mostly used to study an organization or an entity. This research method has evolved over the years as one of the most valuable qualitative research methods. This type of research is used in the areas of management, education sector, philosophical and psychological. This method involves a deep digging into the developments and collects data.

**Merits**

* Case studies are flexible with respect to data collection methods.
* It can extend virtually to any dimension of the topic studied.
* It may be conducted in practically any kind of social setting.
* Case studies offer specific instances of tests of theories.
* It gives a wide range of insights into the study.

**Limitations**

* The most prominent limitation is its limited generalisability.
* It is inadequate for an analysis of macro-problems.
* The danger of investigator’s over-confidence is more in a case study.

**12. Field Study**

Fields studies are scientific enquiries aimed at discovering the relations and interactions among sociological, psychological and educational variables in social institutions and actual life situations like communities, schools, factories, organisations and institutions.

A social or institution is selected and the relations among the attitudes, values, perceptions and behaviours of individuals and groups in the selected situation are studied.

**Steps in Field Study**

Daniel Katz suggests the following steps in the conduct of field study.

* Preliminary planning
* The formulation of research design
* The pretesting of research instruments and procedures
* The full-scale field operation.
* The analysis of materials.

**Advantages**

* In a field study it is possible to maintain continued observation, ascertaining the timing of certain variables.
* It provides opportunity for direct observation of interaction and relationships.
* Possibility of going beyond measures obtained a single instruments.
* It facilitates use of different independent measures.

**Limitations**

* The most serious weakness is its ex post facto character.
* The lack of precision in the measurement of field variables.

**Research Approaches**

The above description of the types of research brings to light the fact that there are two basic approaches to research, viz., *quantitative approach* and the *qualitative approach*. The former involves the generation of data in quantitative form which can be subjected to rigorous quantitative analysis in a formal and rigid fashion. This approach can be further sub-classified into *inferential*, *experimental* and *simulation approaches* to research. The purpose of *inferential approach* to research is to form a data base from which to infer characteristics or relationships of population. This usually means survey research where a sample of population is studied (questioned or observed) to determine its characteristics, and it is then inferred that the population has the same characteristics.

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

*Simulation* *approach* involves the construction of an artificial environment within which relevant information and data can be generated. This permits an observation of the dynamic behaviour of a system (or its sub-system) under controlled conditions. The term ‘simulation’ in the context of business and social sciences applications refers to “the operation of a numerical model that represents the structure of a dynamic process. Given the values of initial conditions, parameters and exogenous variables, a simulation is run to represent the behaviour of the process over time.” Simulation approach can also be useful in building models for understanding future conditions.

*Qualitative approach* to research is concerned with subjective assessment of attitudes, opinions

and behaviour. Research in such a situation is a function of researcher’s insights and impressions.

Such an approach to research generates results either in non-quantitative form or in the form which

are not subjected to rigorous quantitative analysis. Generally, the techniques of focus group interviews,

projective techniques and depth interviews are used. All these are explained at length in chapters

that follow.

**Significance of Research**

“All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry, and

inquiry leads to invention” is a famous Hudson Maxim in context of which the significance of research

can well be understood. Increased amounts of research make progress possible. *Research inculcates*

*scientific and inductive thinking and it promotes the development of logical habits of thinking and organisation*.

*The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times*. The increasingly complexnature of business and government has focused attention on the use of research in solving operationalproblems. Research, as an aid to economic policy, has gained added importance, both for governmentand business.

*Research provides the basis for nearly all government policies in our economic system*. For instance, government’s budgets rest in part on an analysis of the needs and desires of the people and on the availability of revenues to meet these needs. The cost of needs has to be equated to probable revenues and this is a field where research is most needed. Through research we can devise alternative policies and can as well examine the consequences of each of these alternatives.

*Research has its special significance in solving various operational and planning problems of business and industry*. Operations research and market research, along with motivational research,are considered crucial and their results assist, in more than one way, in taking business decisions.Market research is the investigation of the structure and development of a market for the purpose offormulating efficient policies for purchasing, production and sales.

*Research is equally important for social scientists in studying social relationships and in seeking answers to various social problems*. It provides the intellectual satisfaction of knowing afew things just for the sake of knowledge and also has practical utility for the social scientist to knowfor the sake of being able to do something better or in a more efficient manner. Research in socialsciences is concerned both with knowledge for its own sake and with knowledge for what it cancontribute to practical concerns.

In addition to what has been stated above, the significance of research can also be understood keeping in view the following points:

(a) To those students who are to write a master’s or Ph.D. thesis, research may mean a careerism or a way to attain a high position in the social structure;

(b) To professionals in research methodology, research may mean a source of livelihood;

(c) To philosophers and thinkers, research may mean the outlet for new ideas and insights;

(d) To literary men and women, research may mean the development of new styles and creative work;

(e) To analysts and intellectuals, research may mean the generalisations of new theories.

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

**Research Problem**

Without a problem, research cannot proceed, because there is nothing to proceed from and proceed towards. Therefore, the first step in research is to perceive a problem - either practical or theoretical. The recognition or existence of a problem motivates research. It may be noted that research is the process of repeated search for truth/facts. Unless there is a problem to search for, investigation cannot proceed. Thus, a problem sets the goal or direction of research.

**Sources of Research Problems**

If the researcher / research organization has a ready problem on hand, he/she can proceed further in the research process or else you have to search for a problem. Where can you search for research problems? Your own mind, where else? You have to feel the problem and think about it. However, the following sources may help you in identifying the problem / problem areas.

1) **Business Problems:** A research problem is a felt need, the need may be an answer, or a solution or an improvement in facilities / technology eg. Cars Business experiences, various types of problems. They may be business policy problems, operational problems, general management problems, or functional

area problems. The functional areas are Financial Management, Marketing Management, Production Management and Human Resources Management. Every business research problem is expected to solve a management problem by facilitating rational decision-making.

2) **Day to Day Problems:** A research problem can be from the day to day experience of the researcher. Every day problems constantly present something new and worthy of investigation and it depends on the keenness of observation and sharpness of the intellect of the researcher to knit his daily experience into a research problem. For example, a person who travels in city buses every day finds it a problem to get in or get out of the bus. But a Q system (that is the answer to the problem) facilitates boarding and alighting comfortably.

3) **Technological Changes:** Technological changes in a fast changing world are constantly bringing forth new problems and thus new opportunities for research. For example, what is the impact or implications of a new technique or new process or new machine?

4) **Unexplored Areas:** Research problems can be both abstract and of applied interest. The researcher may identify the areas in which much work has been done and the areas in which little work has been done or areas in which no work has been done. He may select those areas which have not been explored so far/explored very little.

5) **Theory of One’s Own Interest:** A researcher may also select a problem for investigation from a given theory in which he has considerable interest. In such situations the researcher must have a thorough knowledge of that theory and should be able to explore some unexplained aspects or assumptions of that theory. His effort should revalidate, or modify or reject the theory.

6) **Books, Theses, Dissertation Abstracts, Articles :** Special assignments in textbooks, research theses, investigative reports, research articles in research journals etc., are rich sources for problem seekers. These sources may suggest some additional areas of needed research. Many of the research theses and articles suggest problems for further investigation which may prove fruitful.

7) **Policy Problems:** Government policy measures give rise to both positive and negative impact. The researcher may identify these aspects for his research. For example, what is the impact of the Government’s new industrial policy on industrial development? What is the impact of Export - Import policy on balance of payments? What is the impact of Securities Exchange Board of India Regulations on stock markets?

8) **Discussions with Supervisor and Other Knowledgeable Persons:** The researcher may find it fruitful to have discussions with his/her proposed supervisor or other knowledgeable persons in the area of the topic.

**Points to be Considered while Selecting a Problem**

The topic or problem which the researcher selects among the many possibilities should meet certain requirements. Every problem selected for research must satisfy the following criteria.

1) The topic selected should be original or at least less explored. The purpose of research is to fill the gaps in existing knowledge or to discover new facts and not to repeat already known facts. Therefore, a preliminary survey of the existing literature in the proposed area of research should be carried out to find out the possibility of making an original contribution. Knowledge about previous research will serve at least three purposes.

a) It will enable the researcher to identify his specific problem for research.

b) It will eliminate the possibility of unnecessary duplication of effort, and

c) It will give him valuable information on the merits and limitations of various research techniques which have been used in the past.

2) It should be of significance and socially relevant and useful.

3) It should be interesting to the researcher and should fit into his aptitude.

4) It should be from an area of the researcher’s specialization.

5) It should correspond to the researcher’s abilities - both acquired and acquirable.

6) It should be big enough to be researchable and small enough to be handled – the topic should be amenable for research with existing and acquirable skills.

7) It should have a clear focus or objective.

8) The feasibility of carrying out research on the selected problem should be checked against the following considerations.

a) Whether adequate and suitable data are available?

b) Whether there is access to the organization and respondents?

c) Whether cooperation will be forth coming from the organization and respondents?

d) What are the resources required and how are they available?

e) Whether the topic is within the resources (money and man power) position of the researcher?

9) It should be completed within the time limits permissible.

**Specification of the Problem**

After going through all the above issues a problem is to be restated in an analytical jargon keeping in view its solution. The best way of understanding the problem is to discuss it with those who first raised it in order to find out how it originally came up and what was in the minds of the people who raised it. The more general the original statement of the problem, the more the necessity of preliminary discussions about its nature. The research problem should define the goal of the researcher in clear terms. It means that along with the problem, the objective of the proposal should adequately be spelled out. Without a clear cut idea of the goal to be reached, research activities would be meaningless.

The first step in the formulation and specification of a research problem is to make it concrete and explicit. There is no foolproof method by which one can do it. However, R.L.Ackoff provides considerable guidance in identifying and specifying a problem of research. He presents five components of a problem.

1) **Research Consumer:** There must be an individual or a group which has difficulty. The individual may be the researcher himself and the group / a group of researchers. For some problems there are also other participants. The researcher, if he/she is different from the research consumer, is a participant in the problem.

2) **Research-Consumer’s Objective:** The research consumer must have something to know or some ends to achieve. Obviously, a person who wants nothing cannot have a problem.

3) **Alternative Means to Achieve the Objective:** The research consumer must have alternative means to achieve his objectives. Means are courses of action. A course of action may involve the use of objects. The objects used thus are instruments. Here an instrument refers to any object, concept or idea which can be effectively used in the pursuit of an objective. It should be remembered that there must be at least two means available to the research consumer. If he/she has no choice of means, he/she cannot have a problem.

4) **Doubt in Regard to Selection of Alternatives:** The existence of alternative courses of action is not enough. To experience a problem, the research consumer must have some doubt as to which alternative to select. Without such a doubt there can be no problem. All problems then get reduced ultimately to the evaluation of efficiency of the alternative means for a given set of objectives.

5) **One or More Environments:** There must be one or more environments to which the difficulty or problem pertains. A problem may exist in one environment and may not in another. Thus a change in environment may produce or remove a problem. A research consumer may have doubts as to which will be the most efficient means. The strategy of marketing a product may be different in the urban market, the semi-urban market and the rural market. The instruments of spreading the family planning message may be different in the case of educated and illiterate people. The range of environments over which a problem may exist may vary from one to many. Some problems are specific to only one environment while others are quite general.

The selection of a topic for research is only half-a-step forward. This general topic does not help a researcher to see what data are relevant to his/her purpose. What are the methods would he/she employ in securing them? And how to organize these? Before he/she can consider all these aspects, he/she has to formulate a specific problem by making the various components of it (as explained above) explicit. A research problem is nothing but a basic question for which an answer or a solution is sought through research. The basic question may be further broken down into specifying questions. These “simple, pointed, limited, empirically verifiable questions are the final result of the phased process, we designate as the formulation of a research problem”. Specification or definition of the problem is therefore a process that involves a progressive narrowing of the scope and sharpening of focus of questions till the specific challenging questions are finally posed. If you can answer the following questions, you have clearly specified/defined the problem.

1) What do you want to know? (What is the problem / what are the questions to be answered).

2) Why do you want to know? (What is the purpose or objective).

3) How do you want to answer or solve it? (What is the methodology we want to adopt to solve it)

4) When do you want to solve it? (Within what time limits)

5) Where do you want to solve it? (Within what spatial limits)

6) Who is your research-consumer? ( to whom are you answering)

Please remember that a problem well put is half solved.

**Theoretical Foundation**

* **Theory.**

The purpose of scientific inquiry is first to understand a phenomenon, then to explain and to predict, and, if possible, to control the occurrence of a phenomenon. A theory explains by incorporating empirically established relationships among variables, and predicts by stating the theory's antecedent-consequent (cause-and-effect) relationships. Theories are developed by combining logic with the results of observations.

**Theoretical Foundation**

A theoretical framework/foundation is analogous to the frame/foundation of the house. Just as the foundation supports a house, a theoretical framework provides a rationale for predictions about the relationships among variables of a research study.

**Importance of Theoretical Framework**

* It provides a context for examining a problem i.e. theoretical rationale
* Developing hypotheses
* A frame of reference/base for
* Observations
* Definitions of concepts
* Research designs
* Interpretations
* Generalizations
* Serves as a guide to systematically identify logical, precisely defined relationships among variables

**Review of Literature**

Literature Review is the documentation of a comprehensive review of the published and unpublished work from secondary sources of data in the areas of specific interest to the researcher. Literature review is an integral part of entire research process. It makes significant contribution to each and every operational step at a later stage. After passing through this stage, a researcher is able to acquaint oneself with the available body of knowledge in the area of interest. The main objectives of literature review are as follows.

* To gain a background knowledge of the research topic.
* To identify the concepts relating to it, potential relationships between them and to formulate researchable hypothesis.
* To identify appropriate methodology, research design, methods of measuring concepts and techniques of analysis.
* To identify data sources used by other researchers.
* To learn how others structured their reports.

The procedure for reviewing literature covers searching the existing literature, reviewing it and developing a theoretical and conceptual framework. The main sources of literature review are books and

journals. In both cases, specifically in journals, there can be a gap of two to three years between the completion of a research project and the publication in a journal.

As with books, the researcher needs to prepare a list of journals for identifying the literatures relevant to his research. Nowadays, researchers make extensive use of the internet sources for literature survey and review, and at the same time, the researcher should be careful about the authenticity of the contents.

Bibliography given at the end of a project gives a clear and complete description of the sources that were used while preparing the report.

**Formulation of objectives**

Objectives are the goals you set out to attain in your study. They inform the reader what the researcher wants to accomplish through the research work. The wording of the objective should be very precise and specific.

Objectives can be written under two headings:

**1.** Main objectives or aims

**2.** Sub-objectives

The main objective is an overall statement of the study. It also states the main associations and relationships that we want to establish. The sub-objectives are the specific aspects of the topic that you want to investigate within the main framework.

* + They should be listed numerically.
  + The wording should be clear, complete and specific.
  + Each objective should contain only one aspect of the study.
  + Use action-oriented words or verbs when writing objectives.

The objectives should start with words, such as ‘to determine’, ‘to find out’, ‘to ascertain’, ‘to measure’, ‘to explore’, etc. The wording of objectives determines the type of research (descriptive, correlational and experimental) and the type of research design you need to adopt to achieve them. For example, in case of descriptive studies, the objective can be stated as, ‘To describe the types of incentives provided by the organizations in Chandigarh to their employees in IT industry’. In correlational studies, it may state, ‘To ascertain the impact of coaching classes on students’ performance’.

Example of main objective and sub-objectives

***Main Objective***

The main objective is to explore the relationship between the use of modern teaching techniques and student performance.

***Sub-objectives***

The sub-objectives are as follows:

**1.** To find out the extent of relationship between the use of modern teaching techniques and student performance.

**2.** To compare the use of modern teaching techniques in government and private schools.

**3.** To study the impact of modern teaching and the level of motivation of students to learn.

**Hypothesis**

We know that research begins with a problem or a felt need or difficulty. The purpose of research is to find a solution to the difficulty. It is desirable that the researcher should propose a set of suggested solutions or explanations of the difficulty which the research proposes to solve. Such tentative solutions formulated as a proposition are called **hypotheses**. The suggested solutions formulated as hypotheses may or may not be the real solutions to the problem. Whether they are or not is the task of research to test and establish.

**Meaning of Hypothesis**

To understand the meaning of a hypothesis, let us see some definitions:

“A hypothesis is a tentative generalization, the validity of which remains to be tested. In its most elementary stage the hypothesis may be any guess, hunch, imaginative idea, which becomes the basis for action or investigation”. (G.A.Lundberg)

“It is a proposition which can be put to test to determine validity”. (Goode and Hatt).

“A hypothesis is a question put in such a way that an answer of some kind can be forth coming” - (Rummel and Ballaine).

These definitions lead us to conclude that a hypothesis is a tentative solution or explanation or a guess or assumption or a proposition or a statement to the problem facing the researcher, adopted on a cursory observation of known and available data, as a basis of investigation, whose validity is to be tested or verified.

**Types of Hypothesis**

Hypotheses can be classified in a variety of ways into different types or kinds.

The following are some of the types of hypotheses:

i) **Explanatory Hypothesis:** The purpose of this hypothesis is to explain a certain fact. All hypotheses are in a way explanatory for a hypothesis is advanced only when we try to explain the observed fact. A large number of hypotheses are advanced to explain the individual facts in life. A theft, a murder, an accident are examples.

ii) **Descriptive Hypothesis:** Sometimes a researcher comes across a complex phenomenon. He/ she does not understand the relations among the observed facts. But how to account for these facts? The answer is a descriptive hypothesis. A hypothesis is descriptive when it is based upon the points of resemblance of something. It describes the **cause** and **effect** relationship of a phenomenon e.g., the current unemployment rate of a state exceeds 25% of the work force. Similarly, the consumers of local made products constitute a significant market segment.

iii) **Analogical Hypothesis:** When we formulate a hypothesis on the basis of similarities (analogy), it is called an analogical hypothesis e.g., families with higher earnings invest more surplus income on long term investments.

iv) **Working hypothesis:** Sometimes certain facts cannot be explained adequately by existing hypotheses, and no new hypothesis comes up. Thus, the investigation is held up. In this situation, a researcher formulates a hypothesis which enables to continue investigation. Such a hypothesis, though inadequate and formulated for the purpose of further investigation only, is called a working hypothesis. It is simply accepted as a starting point in the process of investigation.

v) **Null Hypothesis:** It is an important concept that is used widely in the sampling theory. It forms the basis of many tests of significance. Under this type, the hypothesis is stated negatively. It is null because it may be nullified, if the evidence of a random sample is unfavourable to the hypothesis. It is a hypothesis being tested (H0). If the calculated value of the test is less than the permissible value, Null hypothesis is accepted, otherwise it is rejected. The rejection of a null hypothesis implies that the difference could not have arisen due to chance or sampling fluctuations.

vi) **Statistical Hypothesis:** Statistical hypotheses are the statements derived from a sample. These are quantitative in nature and are numerically measurable. For example, the market share of product X is 70%, the average life of a tube light is 2000 hours etc.

**Criteria for Workable Hypothesis**

A hypothesis controls and directs the research study. When a problem is felt, we require the hypothesis to explain it. Generally, there is more than one hypothesis which aims at explaining the same fact. But all of them cannot be equally good. Therefore, how can we judge a hypothesis to be true or false, good or bad? Agreement with facts is the sole and sufficient test of a true hypothesis. Therefore, certain conditions can be laid down for distinguishing a good hypothesis from bad ones. The formal conditions laid down by thinkers provide the criteria for judging a hypothesis as good or valid. These conditions are as follows:

i) **A hypothesis should be empirically verifiable:** The most important condition for a valid hypothesis is that it should be empirically verifiable. A hypothesis is said to be verifiable, if it can be shown to be either true or false by comparing with the facts of experience directly or indirectly. A hypothesis is true if it conforms to facts and it is false if it does not. Empirical verification is the characteristic of the scientific method.

ii) **A hypothesis should be relevant:** The purpose of formulating a hypothesis is always to explain some facts. It must provide an answer to the problem which initiated the enquiry. A hypothesis is called relevant if it can explain the facts of enquiry.

iii) **A hypothesis must have predictive and explanatory power:** Explanatory power means that a good hypothesis, over and above the facts it proposes to explain, must also explain some other facts which are beyond its original scope. We must be able to deduce a wide range of observable facts which can be deduced from a hypothesis. The wider the range, the greater is its explanatory power.

iv) **A hypothesis must furnish a base for deductive inference on consequences:** In the process of investigation, we always pass from theknown to the unknown. It is impossible to infer any thing from the absolutelyunknown. We can only infer what would happen under supposed conditions byapplying the knowledge of nature we possess. Hence, our hypothesis must bein accordance with our previous knowledge.

v) **A hypothesis does not go against the traditionally established knowledge:** As far as possible, a new hypothesis should not go against anypreviously established law or knowledge. The new hypothesis is expected to beconsistent with the established knowledge.

vi) **A hypothesis should be simple:** A simple hypothesis is preferable to a complex one. It some times happens that there are two or more hypotheses which explain a given fact equally well. Both of them are verified by observable facts. Both of them have a predictive power and both are consistent with established knowledge. All the important conditions of hypothesis are thus satisfied by them. In such cases the simpler one is to be accepted in preference to the complex one.

vii) **A hypothesis must be clear, definite and certain:** It is desirable that the hypothesis must be simple and specific to the point. It must be clearly defined in a manner commonly accepted. It should not be vague or ambiguous.

(viii) **A Hypothesis should be related to available techniques:** If tools and techniques are not available we cannot test the hypothesis. Therefore, the hypothesis should be formulated only after due thought is given to the methods and techniques that can be used to measure the concepts and variables related

to the hypothesis.

**Stages in Hypothesis**

There are four stages. The first stage is feeling of a problem. The observation and analysis of the researcher reveals certain facts. These facts pose a problem. The second stage is formulation of a hypothesis or hypotheses. A tentative supposition/ guess is made to explain the facts which call for an explanation. At this stage some past experience is necessary to pick up the significant aspects of the observed facts. Without previous knowledge, the investigation becomes difficult, if not impossible. The third stage is deductive development of hypothesis using deductive reasoning. The researcher uses the hypothesis as a premise and draws a conclusion from it. And the last stage is the verification or testing of hypothesis. This consists in finding whether the conclusion drawn at the third stage is really true. Verification consists in finding whether the hypothesis agrees with the facts. If the hypothesis stands the test of verification, it is accepted as an explanation of the problem. But if the hypothesis does not stand the test of verification, the researcher has to search for further solutions.

To explain the above stages let us consider a simple example. Suppose, you have started from your home for college on your scooter. A little while later the engine of your scooter suddenly stops. What can be the reason? Why has it stopped? From your past experience, you start guessing that such problems generally arise due to either petrol or spark plug. Then start deducing that the cause could be: (i) that the petrol knob is not on. (ii) that there is no petrol in the tank. (iii) that the spark plug has to be cleaned. Then start verifying them one after another to solve the problem. First see whether the petrol knob is on. If it is not, switch it on and start the scooter. If it is already on, then see whether there is petrol or not by opening the lid of the petrol tank. If the tank is empty, go to the near by petrol bunk to fill the tank with petrol. If there is petrol in the tank, this is not the reason, then you verify the spark plug. You clean the plug and fit it. The scooter starts. That means the problem is with the spark plug. You have identified it. So you got the answer. That means your problem is solved.

**Testing of Hypothesis**

When the hypothesis has been framed in the research study, it must be verified as true or false. Verifiability is one of the important conditions of a good hypothesis. Verification of hypothesis means testing of the truth of the hypothesis in the light of facts. If the hypothesis agrees with the facts, it is said to be true and may be accepted as the explanation of the facts. But if it does not agree it is said to be false. Such a false hypothesis is either totally rejected or modified. Verification is of two types viz., **Direct verification** and **Indirect** **verification**.

**Direct verification** may be either by observation or by experiments. When direct observation shows that the supposed cause exists where it was thought to exist, we have a direct verification. When a hypothesis is verified by an experiment in a laboratory it is called direct verification by experiment. When the hypothesis is not amenable for direct verification, we have to depend on indirect verification.

**Indirect verification** is a process in which certain possible consequences are deduced from the hypothesis and they are then verified directly. Two steps are involved in indirect verification.

(i) Deductive development of hypothesis: By deductive development certain consequences are predicted and (ii) finding whether the predicted consequences follow. If the predicted consequences come true, the hypothesis is said to be indirectly verified. Verification may be done directly or indirectly or through logical methods.

Testing of a hypothesis is done by using statistical methods. Testing is used to accept or reject an assumption or hypothesis about a random variable using a sample from the distribution. The assumption is the null hypothesis (H0), and it is tested against some alternative hypothesis (H1). Statistical tests of hypothesis are applied to sample data. The procedure involved in testing a hypothesis is

A) Select a sample and collect the data.

B) Convert the variables or attributes into statistical form such as mean, proportion.

C) Formulate hypotheses.

D) Select an appropriate test for the data such as t-test, Z-test.

E) Perform computations.

F) Finally draw the inference of accepting or rejecting the null hypothesis

**Uses of Hypothesis**

If a clear scientific hypothesis has been formulated, half of the research work is already done. The advantages/utility of having a hypothesis is summarized here underneath:

i) It is a starting point for many a research work.

ii) It helps in deciding the direction in which to proceed.

iii) It helps in selecting and collecting pertinent facts.

iv) It is an aid to explanation.

v) It helps in drawing specific conclusions.

vi) It helps in testing theories.

vii) It works as a basis for future knowledge.

**Research Design (Research Methodology)**

Research design is also known by different names such as research outline, plan, blue print. In the words of Fred N. Kerlinger, it is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and control variance. The plan includes everything the investigator will do from writing the hypothesis and their operational implications to the final analysis of data. The structure is the outline, the scheme, the paradigms of the operation of the variables. The strategy includes the methods to be used to collect and analyze the data. At the beginning this plan (design) is generally vague and tentative. It undergoes many modifications and changes as the study progresses and insights into it deepen. The working out of the plan consists of making a series of decisions with respect to what, why, where, when, who and how of the research.

**Functions of Research Design**

Regardless of the type of research design selected by the investigator, all plans perform one or more functions outlined below.

i) It provides the researcher with a blue print for studying research questions.

ii) It dictates boundaries of research activity and enables the investigator to channel his energies in a specific direction.

iii) It enables the investigator to anticipate potential problems in the implementation of the study.

iv) The common function of designs is to assist the investigator in providing answers to various kinds of research questions.

A study design includes a number of component parts which are interdependent and which demand a series of decisions regarding the definitions, methods, techniques, procedures, time, cost and administration aspects.

**Components of a Research Design**

A research design basically is a plan of action. Once the research problem is selected, then it must be executed to get the results. Then how to go about it? What is its scope? What are the sources of data? What is the method of enquiry? What is the time frame? How to record the data? How to analyze the data? What are the tools and techniques of analysis? What is the manpower and organization required? What are the resources required? These and many such are the subject matter of attacking the research problem demanding decisions in the beginning itself to have greater clarity about the research study. It is similar to having a building plan before the building is constructed. Thus, according to P.V. Young the various “considerations which enter into making decisions regarding what, where, when, how much, by what means constitute a plan of study or a study design”. Usually the contents or components of a Research design are as follows:

1) **Need for the Study:** Explain the need for and importance of this study and its relevance.

2) **Review of Previous Studies:** Review the previous works done on this topic, understand what they did, identify gaps and make a case for this study and justify it.

3) **Statement of Problem:** State the research problem in clear terms and give a title to the study.

4) **Objectives of Study:** What is the purpose of this study? What are the objectives you want to achieve by this study? The statement of objectives should not be vague. They must be specific and focussed.

5) **Formulation of Hypothesis:** Conceive possible outcome or answers to the research questions and formulate into hypothesis tests so that they can be tested.

6) **Operational Definitions:** If the study is using uncommon concepts or unfamiliar tools or using even the familiar tools and concepts in a specific sense, they must be specified and defined.

7) **Scope of the Study:** It is important to define the scope of the study, because the scope decides what is within its purview and what is outside. Scope includes Geographical scope, content scope, chronological scope of the study. The territorial area to be covered by the study should be decided. E.g. only Delhi or northern states or All India. As far as content scope is concerned according to the problem say for example, industrial relations in so and so organization, what are aspects to be studied, what are the aspects not coming under this and hence not studied. Chronological scope i.e., time period selection and its justification is important. Whether the study is at a point of time or longitudinal say 1991-2003.

8) **Sources of Data:** This is an important stage in the research design. At this stage, keeping in view the nature of research, the researcher has to decide the sources of data from which the data are to be collected. Basically the sources are divided into primary source (field sources) and secondary source (documentary sources). The data from primary source are called as primary data, and data from secondary source are called secondary data. Hence, the researcher has to decided whether to collect from primary source or

secondary source or both sources.

9) **Method of Collection:** After deciding the sources for data collection, the researcher has to determine the methods to be employed for data collection, primarily, either census method or sampling method. This decision may depend on the nature, purpose, scope of the research and also time factor and financial resources.

10) **Tools & Techniques:** The tools and techniques to be used for collecting data such as observation, interview, survey, schedule, questionnaire, etc., have to be decided and prepared.

11) **Sampling Design:** If it is a sample study, the sampling techniques, the size of sample, the way samples are to be drawn etc., are to be decided.

12) **Data Analysis:** How are you going to process and analyze the data and information collected? What simple or advanced statistical techniques are going to be used for analysis and testing of hypothesis, so that necessary care can be taken at the collection stage.

13) **Presentation of the Results of Study:** How are you going to present the results of the study? How many chapters? What is the chapter scheme? The chapters, their purpose, their titles have to be outlined. It is known as chapterisation.

14) **Time Estimates:** What is the time available for this study? Is it limited or unlimited time? Generally, it is a time bound study. The available or permitted time must be apportioned between different activities and the activities to be carried out within the specified time. For example, preparation of research design one month, preparation of questionnaire one month, data collection two months, analysis of data two months, drafting of the report two months etc.,

15) **Financial Budget:** The design should also take into consideration the various costs involved and the sources available to meet them. The expenditures like salaries (if any), printing and stationery, postage and telephone, computer and secretarial assistance etc.

16) **Administration of the Enquiry:** How is the whole thing to be executed? Who does what and when? All these activities have to be organized systematically, research personnel have to be identified and trained. They must be entrusted with the tasks, the various activities are to be coordinated and the whole project must be completed as per schedule. Research designs provide guidelines for investigative activity and not necessarily hard and fast rules that must remain unbroken. As the study progresses, new aspects, new conditions and new connecting links come to light and it is necessary to change the plan / design as circumstances demand. A universal characteristic of any research plan is its flexibility. Depending upon the method of research, the designs are also known as survey design, case study design, observation design and experimental design.

**SAMPLING DESIGN**

**Introduction**

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. Each observation measures one or more properties (weight, location, etc.) of an observable entity enumerated to distinguish objects or individuals. Survey weights often need to be applied to the data to adjust for the sample design. Results from probability theory and statistical theory are employed to guide practice.

**Sampling – An Introduction**

A sample is a part of a target population, which is carefully selected to represent the population. Sampling frame is the list of elements from which the sample is actually drawn. Actually, sampling frame is nothing but the correct list of population.

*Example:* Telephone directory, Product finder, Yellow pages.

The sampling process comprises several stages:

1. Defining the population of concern

2. Specifying a sampling frame, a set of items or events possible to measure

3. Specifying a sampling method for selecting items or events from the frame

4. Determining the sample size

5. Implementing the sampling plan

6. Sampling and data collecting

7. Reviewing the sampling process

**Distinction between Census and Sampling**

Census refers to complete inclusion of all elements in the population. A sample is a sub-group of the population.

**When is a Census Appropriate?**

1. A census is appropriate if the size of population is small.

*Example:* A researcher may be interested in contacting firms in iron and steel or petroleum products industry. These industries are limited in number, so a census will be suitable.

2. Sometimes, the researcher is interested in gathering information from every individual.

*Example:* Quality of food served in a mess.

**When is Sample Appropriate?**

1. When the size of population is large.

2. When time and cost are the main considerations in research.

3. If the population is homogeneous.

4. Also, there are circumstances when a census is not possible.

**Steps in Sample Design**

While developing a sampling design, the researcher must pay attention to the following points:

(i) **Type of universe:** The first step in developing any sample design is to clearly define the set of objects, technically called the Universe, to be studied. The universe can be finite or infinite. In finite universe the number of items is certain, but in case of an infinite universe the number of items is infinite, i.e., we cannot have any idea about the total number of items. The population of a city, the number of workers in a factory and the like are examples of finite universes, whereas the number of stars in the sky, listeners of a specific radio programme, throwing of a dice etc. are examples of infinite universes.

(ii) **Sampling unit:** A decision has to be taken concerning a sampling unit before selecting sample. Sampling unit may be a geographical one such as state, district, village, etc., or a construction unit such as house, flat, etc., or it may be a social unit such as family, club, school, etc., or it may be an individual. The researcher will have to decide one or more of such units that he has to select for his study.

(iii) **Source list:** It is also known as ‘sampling frame’ from which sample is to be drawn. It contains the names of all items of a universe (in case of finite universe only). If source list is not available, researcher has to prepare it. Such a list should be comprehensive, correct, reliable and appropriate. It is extremely important for the source list to be as representative of the population as possible.

(iv) **Size of sample:** This refers to the number of items to be selected from the universe to constitute a sample. This a major problem before a researcher. The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility. While deciding the size of sample, researcher must determine the desired precision as also an acceptable confidence level for the estimate. The size of population variance needs to be considered as in case of larger variance usually a bigger sample is needed. The size of population must be kept in view for this also limits the sample size. The parameters of interest in a research study must be kept in view, while deciding the size of the sample. Costs too dictate the size of sample that we can draw. As such, budgetary constraint must invariably be taken into consideration when we decide the sample size.

(v) **Parameters of interest:** In determining the sample design, one must consider the question of the specific population parameters which are of interest. For instance, we may be interested in estimating the proportion of persons with some characteristic in the population, or we may be interested in knowing some average or the other measure concerning the population. There may also be important sub-groups in the population about whom we would like to make estimates. All this has a strong impact upon the sample design we would accept.

(vi) **Budgetary constraint:** Cost considerations, from practical point of view, have a major impact upon decisions relating to not only the size of the sample but also to the type of sample. This fact can even lead to the use of a non-probability sample.

(vii) **Sampling procedure:** Finally, the researcher must decide the type of sample he will use i.e., he must decide about the technique to be used in selecting the items for the sample. In fact, this technique or procedure stands for the sample design itself. There are several sample designs (explained in the pages that follow) out of which the researcher must choose one for his study. Obviously, he must select that design which, for a given sample size and for a given cost, has a smaller sampling error.

**Criteria of selecting A Sampling Procedure**

In this context one must remember that two costs are involved in a sampling analysis viz., the cost of collecting the data and the cost of an incorrect inference resulting from the data. Researcher must keep in view the two causes of incorrect inferences viz., systematic bias and sampling error. A *systematic bias* result from errors in the sampling procedures, and it cannot be reduced or eliminated by increasing the sample size. At best the causes responsible for these errors can be detected and corrected. Usually a systematic bias is the result of one or more of the following factors:

**1. Inappropriate sampling frame:** If the sampling frame is inappropriate i.e., a biased representation of the universe, it will result in a systematic bias.

**2. Defective measuring device:** If the measuring device is constantly in error, it will result in systematic bias. In survey work, systematic bias can result if the questionnaire or the interviewer is biased. Similarly, if the physical measuring device is defective there will be systematic bias in the data collected through such a measuring device.

**3. Non-respondents:** If we are unable to sample all the individuals initially included in the sample, there may arise a systematic bias. The reason is that in such a situation the likelihood of establishing contact or receiving a response from an individual is often correlated with the measure of what is to be estimated.

**4. Indeterminacy principle:** Sometimes we find that individuals act differently when kept under observation than what they do when kept in non-observed situations. For instance, if workers are aware that somebody is observing them in course of a work study on the basis of which the average length of time to complete a task will be determined and accordingly the quota will be set for piece work, they generally tend to work slowly in comparison to the speed with which they work if kept unobserved. Thus, the indeterminacy principle may also be a cause of a systematic bias.

**5. Natural bias in the reporting of data:** Natural bias of respondents in the reporting of data is often the cause of a systematic bias in many inquiries. There is usually a downward bias in the income data collected by government taxation department, whereas we find an upward bias in the income data collected by some social organization. People in general understate their incomes if asked about it for tax purposes, but they overstate the same if asked for social status or their affluence. Generally in psychological surveys, people tend to give what they think is the ‘correct’ answer rather than revealing their true feelings.

**Types of Sample Designs**

If money, time, trained manpower and other resources were not a concern, the researcher could get most accurate data from surveying the entire population of interest. Since most often the resources are scarce, the researcher is forced to go for sampling. But the real purpose of the survey is to know the characteristics of the population. Then the question is with what level of confidence will the researcher be able to say that the characteristics of a sample represent the entire population. Using a combination of tasks of hypotheses and unbiased sampling methods, the researcher can collect data that actually represents the characteristics of the entire population from which the sample was taken. To ensure a high level of confidence that the sample represents the population it is necessary that the sample is unbiased and sufficiently large.

The various sampling methods can be classified into two categories. These are Probability sampling methods and non- Probability sampling methods. Let us discuss them in detail with help of following chart.

**Probability sampling**

Probability sampling is also known as ‘random sampling’ or ‘chance sampling’. Under this sampling design, every item of the universe has an equal chance of inclusion in the sample.

Probability sampling technique has been classified into 2 categories based on element selection basis,

**1. Simple Random Sampling**

When each sample element is drawn individually from the population at large, then the sample so drawn is known as Simple Random Sampling or ‘unrestricted sample’

**2. Complex Random Sampling**

Sampling are covered under the term Complex Random Sampling are also called as ‘restricted sampling’ because the samples are drawn under some restrictions which are discussed in the following.

**1. Simple Random sampling:** The most commonly used random sampling method is simple random sampling method. A simple random sample is one in which each item in the total population has an equal chance of being included in the sample. In addition, the selection of one item for inclusion in the sample

should in no way influence the selection of another item. Simple random sampling should be used with a homogeneous population, that is, a population consisting of items that possess the same attributes that the researcher is interested in. The characteristics of homogeneity may include such as age, sex, income, social/religious/political affiliation, geographical region etc.

The best way to choose a simple random sample is to use random number table. A random sampling method should meet the following criteria.

a) Every member of the population must have an equal chance of inclusion in the sample.

b) The selection of one member is not affected by the selection of previous members. The random numbers are a collection of digits generated through a probabilistic mechanism. The random numbers have the following properties:

i) The probability that each digit (0,1,2,3,4,5,6,7,8,or 9) will appear at any place is the same. That is 1/10.

ii) The occurrence of any two digits in any two places is independent of each other. Each member of a population is assigned a unique number. The members of the population chosen for the sample will be those whose numbers are identical to the ones extracted from the random number table in succession until the desired sample size is reached.

**Advantages**

i) The simple random sample requires less knowledge about the characteristics of the population.

ii) Since sample is selected at random giving each member of the population equal chance of being selected the sample can be called as unbiased sample. Bias due to human preferences and influences is eliminated.

iii) Assessment of the accuracy of the results is possible by sample error estimation.

iv) It is a simple and practical sampling method provided population size is not large.

**Limitations**

i) If the population size is large, a great deal of time must be spent listing and numbering the members of the population.

ii) A simple random sample will not adequately represent many population characteristics unless the sample is very large. That is, if the researcher is interested in choosing a sample on the basis of the distribution in the population of gender, age, social status, a simple random sample needs to be very large to ensure all these distributions are representative of the population. To obtain a representative sample across multiple population attributes we should use stratified random sampling.

**2. Complex Random Sampling**

Probability sampling under restricted sampling techniques, as stated above, may result in complex random sampling designs. Such designs may as well be called ‘mixed sampling designs’ for many of such designs may represent a combination of probability and non-probability sampling procedures in selecting a sample. Some of the popular complex random sampling designs are as follows:

**(i) Systematic sampling:**

In some instances, the most practical way of sampling is to select every *i*th item on a list. Sampling of this type is known as systematic sampling. An element of randomness is introduced into this kind of sampling by using random numbers to pick up the unit with which to start. For instance, if a 4 per cent sample is desired, the first item would be selected randomly from the first twenty-five and thereafter every 25th item would automatically be included in the sample. Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals. Although a systematic sample is not a random sample in the strict sense of the term, but it is often considered reasonable to treat systematic sample as if it were a random sample.

Systematic sampling has certain plus points. It can be taken as an improvement over a simple random sample in as much as the systematic sample is spread more evenly over the entire population. It is an easier and less costlier method of sampling and can be conveniently used even in case of large populations. But there are certain dangers too in using this type of sampling. If there is a hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling. For instance, every 25th item produced by a certain production process is defective. If we are to select a 4% sample of the items of this process in a systematic manner, we would either get all defective items or all good items in our sample depending upon the random starting position. If all elements of the universe are ordered in a manner representative of the total population, i.e., the population list is in random order, systematic sampling is considered equivalent to random sampling. But if this is not so, then the results of such sampling may, at times, not be very reliable. In practice, systematic sampling is used when lists of population are available and they are of considerable length.

**Advantages:**

i) The main advantage of using systematic sample is that it is more expeditious to collect a sample systematically since the time taken and work involved is less than in simple random sampling. For example, it is frequently used in exit polls and store consumers.

ii) This method can be used even when no formal list of the population units is available. For example, suppose if we are interested in knowing the opinion of consumers on improving the services offered by a store we may simply choose every kth (say 6th) consumer visiting a store provided that we know how many consumers are visiting the store daily (say 1000 consumers visit and we want to have 100 consumers as sample size).

**Limitations:**

i) If there is periodicity in the occurrence of elements of a population, the selection of sample using systematic sample could give a highly un-representative sample. For example, suppose the sales of a consumer store are arranged chronologically and using systematic sampling we select sample for 1st of every month. The 1st day of a month can not be a representative sample for the whole month. Thus in systematic sampling there is a danger of order bias.

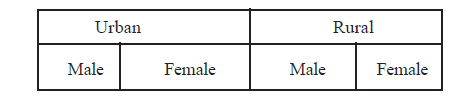
ii) Every unit of the population does not have an equal chance of being selected and the selection of units for the sample depends on the initial unit selection. Regardless how we select the first unit of sample, subsequent units are automatically determined lacking complete randomness.

**2. Stratified Random Sampling:**

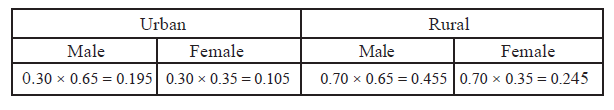
The stratified sampling method is used when the population is heterogeneous rather than homogeneous. A heterogeneous population is composed of unlike elements such as male/female, rural/urban, literate/illiterate, high income/low income groups, etc. In such cases, use of simple random sampling may not always provide a representative sample of the population. In stratified sampling, we divide the population into relatively homogenous groups called strata. Then we select a sample using simple random sampling from each stratum. There are two approaches to decide the sample size from each stratum, namely, proportional stratified sample and disproportional stratified sample. With either approach, the stratified sampling guarantees that every unit in the population has a chance of being selected. We will now discuss these two approaches of selecting samples.

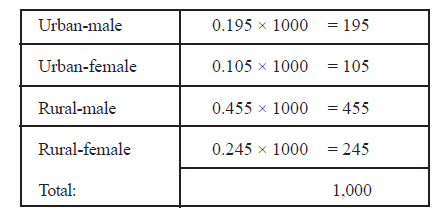
i) ***Proportional Stratified Sample:*** If the number of sampling units drawn from each stratum is in proportion to the corresponding stratum population size, we say the sample is proportional stratified sample. For example, let us say we want to draw a stratified random sample from a heterogeneous population (on some characteristics) consisting of rural/urban and male/female respondents.

So we have to create 4 homogeneous sub groups called stratums as follows:



To ensure each stratum in the sample will represent the corresponding stratum in the population we must ensure each stratum in the sample is represented in the same proportion to the stratums as they are in the population. Let us assume that we know (or can estimate) the population distribution as follows: 65% male, 35% female and 30% urban and 70% rural. Now we can determine the approximate proportions of our 4 stratums in the population as shown below.

Thus a representative sample would be composed of 19.5% urban-males, 10.5% urban-females, 45.5% rural-males and 24.5% rural females. Each percentage should be multiplied by the total sample size needed to arrive at the actual sample size required from each stratum. Suppose we require 1000 samples then the required sample in each stratum is as follows:

****

ii) ***Disproportional Stratified Sample:*** In a disproportional stratified sample, sample size for each stratum is not allocated on a proportional basis with the population size, but by analytical considerations of the researcher such as stratum variance, stratum population, time and financial constraints etc. For example, if the researcher is interested in finding differences among different stratums, disproportional sampling should be used. Consider the example of income distribution of households. There is a small percentage of households within the high income brackets and a large percentage of households within the low income brackets. The income among higher income group households has higher variance than the variance among the lower income group households. To avoid under-representation of higher income groups in the sample, a disproportional sample is taken. This indicates that as the variability within the

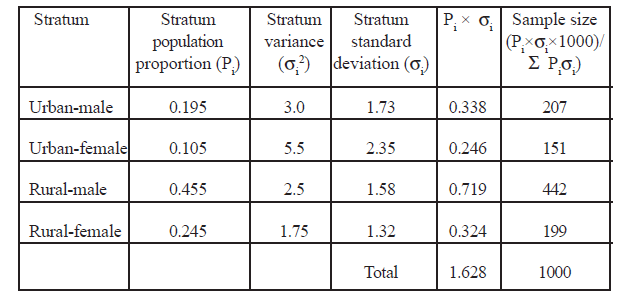
stratum increases sample size must increase to provide accurate estimates and vice-versa.

Suppose in our example of urban/rural and male/female stratum populations, the stratum estimated variances (s2) are as follows.

Urban-male 3.0; Urban-female 5.5; Rural-males 2.5; Rural-females 1.75.

The above figures are, normally, estimated on the basis of previous knowledge of a researcher.

Then the allocation of sample size of 1000 for each strata using disproportional stratified sampling method will be as shown in the following table:



**Advantages:**

a) Since the sample are drawn from each of the stratums of the population, stratified sampling is more representative and thus more accurately reflects characteristics of the population from which they are chosen.

b) It is more precise and to a great extent avoids bias.

c) Since sample size can be less in this method, it saves a lot of time, money and other resources for data collection.

**Limitations:**

a) Stratified sampling requires a detailed knowledge of the distribution of attributes or characteristics of interest in the population to determine the homogeneous groups that lie within it. If we cannot accurately identify the homogeneous groups, it is better to use simple random sample since improper stratification can lead to serious errors.

b) Preparing a stratified list is a difficult task as the lists may not be readily available.

**3. Cluster Sampling:**

In cluster sampling we divide the population into groups having heterogenous characteristics called clusters and then select a sample of clusters using simple random sampling. We assume that each of the clusters is representative of the population as a whole. This sampling is widely used for geographical studies of many issues. For example if we are interested in finding the consumers’ (residing in Delhi) attitudes towards a new product of a company, the whole city of Delhi can be divided into 20 blocks. We assume that each of these blocks will represent the attitudes of consumers of Delhi as a whole, we might use cluster sampling treating each block as a cluster. We will then select a sample of 2 or 3 clusters and obtain the information from consumers covering all of them.

The principles that are basic to the cluster sampling are as follows:

i) The differences or variability within a cluster should be as large as possible. As far as possible the variability within each cluster should be the same as that of the population.

ii) The variability between clusters should be as small as possible. Once the clusters are selected, all the units in the selected clusters are covered for obtaining data.

**Advantages:**

a) The cluster sampling provides significant gains in data collection costs, since traveling costs are smaller.

b) Since the researcher need not cover all the clusters and only a sample of clusters are covered, it becomes a more practical method which facilitates fieldwork.

**Limitations:**

a) The cluster sampling method is less precise than sampling of units from the whole population since the latter is expected to provide a better cross-section of the population than the former, due to the usual tendency of units in a cluster to be homogeneous.

b) The sampling efficiency of cluster sampling is likely to decrease with the decrease in cluster size or increase in number of clusters.

The above advantages or limitations of cluster sampling suggest that, in practical situations where sampling efficiency is less important but the cost is of greater significance, the cluster sampling method is extensively used. If the division of clusters is based on the geographic sub-divisions, it is known as area sampling.

In cluster sampling instead of covering all the units in each cluster we can resort to sub-sampling as two-stage sampling. Here, the clusters are termed as primary units and the units within the selected clusters are taken as secondary units.

**4. Multistage Sampling:**

We have already covered two stage sampling. Multi stage sampling is a generalisation of two stage sampling. As the name suggests, multi stage sampling is carried out in different stages. In each stage

progressively smaller (population) geographic areas will be randomly selected. A political pollster interested in assembly elections in Uttar Pradesh may first divide the state into different assembly units and a sample of assembly constituencies may be selected in the first stage. In the second stage, each of

the sampled assembly constituents are divided into a number of segments and a second stage sampled assembly segments may be selected. In the third stage within each sampled assembly segment either all the house-holds or a sample random of households would be interviewed. In this sampling method, it is possible to take as many stages as are necessary to achieve a representative sample. Each stage results in a reduction of sample size. In a multi stage sampling at each stage of sampling a suitable method of sampling is used. More number of stages are used to arrive at a sample of desired sampling units.

**Advantages**

a) Multistage sampling provides cost gains by reducing the data collection on costs.

b) Multistage sampling is more flexible and allows us to use different sampling procedures in different stages of sampling.

c) If the population is spread over a very wide geographical area, multistage sampling is the only sampling method available in a number of practical situations.

**Limitations**

a) If the sampling units selected at different stages are not representative multistage sampling becomes less precise and efficient.

**5. Area sampling:**

If clusters happen to be some geographic subdivisions, in that case cluster sampling is better known as area sampling. In other words, cluster designs, where the primary sampling unit represents a cluster of units based on geographic area, are distinguished as area sampling. The plus and minus points of cluster sampling are also applicable to area sampling.

**6. Sampling with probability proportional to size:**

In case the cluster sampling units do not have the same number or approximately the same number of elements, it is considered appropriate to use a random selection process where the probability of each cluster being included in the sample is proportional to the size of the cluster. For this purpose, we have to list the number of elements in each cluster irrespective of the method of ordering the cluster. Then we must sample systematically the appropriate number of elements from the cumulative totals. The actual numbers selected in this way do not refer to individual elements, but indicate which clusters and how many from the cluster are to be selected by simple random sampling or by systematic sampling. The results of this type of sampling are equivalent to those of a simple random sample and the method is less cumbersome and is also relatively less expensive.

**7. Sequential sampling:**

This sampling design is some what complex sample design. The ultimate size of the sample under this technique is not fixed in advance, but is determined according to mathematical decision rules on the basis of information yielded as survey progresses. This is usually adopted in case of acceptance sampling plan in context of statistical quality control. When a particular lot is to be accepted or rejected on the basis of a single sample, it is known as single sampling; when the decision is to be taken on the basis of two samples, it is known as double sampling and in case the decision rests on the basis of more than two samples but the number of samples is certain and decided in advance, the sampling is known as multiple sampling. But when the number of samples is more than two but it is neither certain nor decided in advance, this type of system is often referred to as sequential sampling. Thus, in brief, we can say that in sequential sampling, one can go on taking samples one after another as long as one desires to do so.

**Non- Probability sampling**

The non-random sampling methods are also often called non-probability sampling methods. In a non-random sampling method the probability of any particular unit of the population being chosen is unknown. Here the method of selection of sampling units is quite arbitrary as the researchers rely heavily on personal judgment. Non-random sampling methods usually do not produce samples that are representative of the general population from which they are drawn. The greatest error occurs when the researcher attempts to generalise the results on the basis of a sample to the entire population. Such an error is insidious because it is not at all obvious from merely looking at the data, or even from looking at the sample. The easiest way to recognise whether a sample is representative or not is to determine whether the sample is selected randomly or not. Nevertheless, there are occasions where non-random samples are best suited for the researcher’s purpose. The various non-random sampling methods commonly used are:

1) Convenience Sampling

2) Judgement Sampling

3) Quota Sampling.

4) Accidental sampling ; and

5) Snowball sampling

Let us discuss these methods in detail.

1) **Convenience Sampling:**

Convenience sampling refers to the method of obtaining a sample that is most conveniently available to the researcher. For example, if we are interested in finding the overtime wage paid to employees working in call centres, it may be convenient and economical to sample employees of call centres in a nearby area. Also, on various issues of public interest like budget, election, price rise etc., the television channels often present on-the-street interviews with people to reflect public opinion. It may be

cautioned that the generalisation of results based on convenience sampling beyond that particular sample may not be appropriate. Convenience samples are best used for exploratory research when additional research will be subsequently conducted with a random sample. Convenience sampling is also useful in testing the questionnaires designed on a pilot basis. Convenience sampling is extensively used in marketing studies.

2) **Judgement Sampling:**

Judgement sampling method is also known as purposive sampling. In this method of sampling the selection of sample is based on the researcher’s judgment about some appropriate characteristic required of the sample units. For example, the calculation of consumer price index is based on a judgment sample of a basket of consumer items, and other related commodities and services which are expected to reflect a representative sample of items consumed by the people. The prices of these items are collected from selected cities which are viewed as typical cities with demographic profiles matching the national profile.

In business judgment sampling is often used to measure the performance of salesmen/saleswomen. The salesmen/saleswomen are grouped into high, medium or low performers based on certain specified qualities. Then the sales manager may actually classify the salesmen/saleswomen working under him/her who in his/her opinion will fall in which group. Judgment sampling is also often used in forecasting election results. We may often wonder how a pollster can predict an election based on only 2% to 3% of votes covered. It is needless to say the method is biased and does not have any scientific basis. However, in the absence of any representative data, one may resort to this kind of non-random sampling.

**3)** **Quota Sampling:**

The quota sampling method is commonly used in marketingresearch studies. The samples are selected on the basis of some parameterssuch as age, sex, geographical region, education, income, occupation etc, inorder to make them as representative samples. The investigators, then, areassigned fixed quotas of the sample meeting these population characteristics.The purpose of quota sampling is to ensure that various sub-groups of thepopulation are represented on pertinent sample characteristics to the extent thatthe investigator desires. The stratified random sampling also has this objectivebut should not be confused with quota sampling. In the stratified samplingmethod the researcher selects a random sample from each group of thepopulation, where as, in quota sampling, the interviewer has a quota fixed forhim/her to achieve. For example, if a city has 10 market centres, a soft drinkcompany may decide to interview 50 consumers from each of these 10 marketcentres to elicit information on their products. It is entirely left to theinvestigator whom he/she will interview at each of the market centres and thetime of interview. The interview may take place in the morning, mid day, orevening or it may be in the winter or summer.

Quota sampling has the advantage that the sample confirms the selected characteristics of the population that the researcher desires. Also, the cost and time involved in collecting the data are also greatly reduced.

**Limitations:**

a) In quota sampling the respondents are selected according to the convenience of the field investigator rather than on a random basis. This kind of selection of sample may be biased. Suppose in our example of soft drinks, after the sample is taken it was found that most of the respondents belong to the lower income

group then the purpose of conducting the survey becomes useless and the results may not reflect the actual situation.

b) If the number of parameters, on which basis the quotas are fixed, are larger then it becomes difficult for the researcher to fix the quota for each sub-group.

c) The field workers have the tendency to cover the quota by going to those places where the respondents may be willing to provide information and avoid those with unwilling respondents. For example, the investigators may avoid places where high income group respondents stay and cover only low income group areas.

**4. Accidental sampling:**

It is akin to quota sampling, but it is used in market research (in market places) where a researcher can come across any person and they may not have any information.

**5. Snowball sampling:**

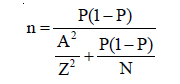
In this kind of sampling, the information may be selected from few individuals and they may identify other people for the purpose of gathering information. They may also become a part of the sample. It creates a network of sample elements.

**Determination of Sample Size**

The question of how large a sample should be is a difficult one. Sample size can be determined by various factors (like time, funds, manpower, population size, purpose of study etc. For example, if the available funds for study are limited then the researcher may not be able to spend more than a fixed proportion of the total fund available with him/her. In general, sample size depends on the nature of the analysis to be performed, the desired precision of the estimates one wishes to achieve, number of variables that have to be examined simultaneously and how heterogeneous is the population spread. Moreover, technical considerations suggest that the required sample size is a function of the precision of the estimates one wishes to achieve, the variance of the population and statistical level of confidence one wishes to use. The higher the precision and confidence level required, the larger the sample size should be. Typical confidence levels are 95% and 99%, while a typical precision (significance) value is 1% or 5%.

Once the researcher determines the desired degree of precision and confidence level, there are several formulas he/she can use to determine the sample size and interpretation of results depending on the plan of the study. Here we will discuss three of them.

1) If the researcher wishes to report the results as proportions of the sample responses, use the following formula.



Where, n = Sample size.

P = Estimated percentage of the population possessing attribute of interest.

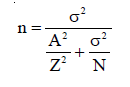
A = Accuracy desired, usually expressed as a decimal (i.e. 0.01, 0.05, etc.)

Z = Standardization value indicating a confidence level (Z=1.96 at 95% confidence level and Z = 2.56 at 99% confidence level.

N = Population size (known or estimated)

2) If the researcher wishes to report the results as means of the sample responses,

use the following formula.



Where, n = Sample size.

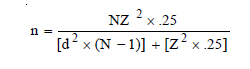
P = Estimated percentage of the population possessing attribute of interest.

A = Accuracy desired, usually expressed as a decimal (i.e. 0.01, 0.05, etc.)

Z = Standardization value indicating a confidence level (Z=1.96 at 95% confidence level and Z = 2.56 at 99% confidence level).

N = Population size (known or estimated)

3) If the researcher plans the results in a variety of ways or if he/she has difficulty in estimating the proportion or standard deviation of the attribute of interest, the following formula may be more useful.



Where, n = Sample size required

d = Accuracy precision level (i.e. 0.01, 0.05, 0.10 etc.)

Z = Standardization value indicating a confidence level (Z = 1.96 at 95% confidence level and Z = 2.56 at 99% confidence level.

N = Population size (known or estimated).

For example, if the population size (N) is 1000 and you wish a 95% confidence level and ±5% precision level (d=0.05 and Z=1.96) then the sample size (n):



**SAMPLING AND NON-SAMPLING ERRORS**

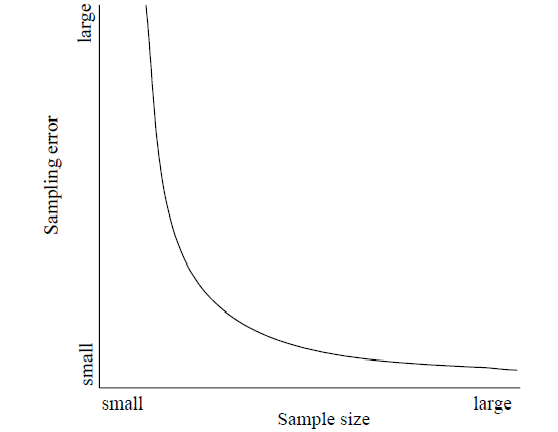
The quality of a research project depends on the accuracy of the data collected and its representation to the population. There are two broad sources of errors. These are sampling errors and non-sampling errors.

**Sampling Errors**

The principal sources of sampling errors are the sampling method applied, and the sample size. This is due to the fact that only a part of the population is covered in the sample. The magnitude of the sampling error varies from one sampling method to the other, even for the same sample size. For example, the sampling error associated with simple random sampling will be greater than stratified random sampling if the population is heterogeneous in nature.

Intuitively, we know that the larger the sample the more accurate the research. In fact, the sampling error varies with samples of different sizes. Increasing the sample size decreases the sampling error.

The following Figure gives an approximate relationship between sample size and sampling error. Study the following figure carefully.



**Non-Sampling Errors**

The non-sampling errors arise from faulty research design and mistakes in executing research. There are many sources of non-sampling errors which may be broadly classified as:

(a) respondent errors, and

(b) administrative errors.

a) **Respondent Errors**:

If the respondents co-operate and give the correct information the objectives of the researcher can be easily accomplished. However, in practice, this may not happen. The respondents may either refuse to provide information or even if he/she provides information it may be biased. If the respondent fails to provide information, we call it as ***non-response error****.*

Although this problem is present in all types of surveys, the problem is more acute in mailed surveys. Non-response also leads to some extreme situations like those respondents who are willing to provide information are over represented while those who are indifferent are under-represented in the sample. In order to minimise the non-response error the researcher often seeks to re-contact with the non-respondents if they were not available earlier.

If the researcher finds that the non-response rate is more in a particular group of respondents (for example, higher income groups) additional efforts should be made to obtain data from these under-represented groups of the population. For example, for these people who are not responding to the mailed questionnaires, personal interviews may be conducted to obtain data. In a mailed questionnaire the researcher never knows whether the respondent really refused to provide data or was simply indifferent. There are several techniques which help to encourage respondents to reply.

***Response bias*** occurs when the respondent may not give the correct information and try to mislead the investigator in a certain direction. The respondents may consciously or unconsciously misrepresent the truth. For example, if the investigator asks a question on the income of the respondent he/ she may not give the correct information for obvious reasons. Or the investigator may not be able to put a question that is sensitive (thus avoiding embarrassment). This may arise from the problems in designing the questionaire and the content of questions. Respondents who must understand the questions may unconsciously provide biased information.

The response bias may also occur because the interviewer’s presence influences respondents to give untrue or modified answers. The respondents interviewers tendency is to please the other person rather than provide/elicit the correct information.

**b) Administrative Errors:** The errors that have arisen due to improper administration of the research process are called administrative errors. There are four types of administrative errors. These are as follows:

i) Sample selection error,

ii) Investigator error,

iii) investigator cheating, and

iv) Data processing error.

i) **Sample Selection Error:** It is difficult to execute a sampling plan. For example, we may plan to use systematic sampling plan in a market research study of a new product and decide to interview every 5th

customer coming out of a consumer store. If the day of interview happened to be a working day then we are excluding all those consumers who are working. This may lead to an error because of the unrepresentative sample selection.

ii) **Investigator Error:** When the investigator interviews the respondent, he/ she may fail to record the information correctly or may fail to cross check the information provided by the respondent. Therefore, the error may arise due to the way the investigator records the information.

iii) **Investigator Cheating:** Some times the investigator may try to fake the data even without meeting the concerned respondents. There should be some mechanism to crosscheck this type of faking by the investigator.

iv) **Data Processing Error:** Once the data is collected the next job the researcher does is edit, code and enter the data into a computer for further processing and analysis. The errors can be minimised by careful editing, coding and entering the data into a computer.

**Control of Errors**

In the above two sections we have identified the most significant sources of errors. It is not possible to eliminate completely the sources of errors. However, the researcher’s objective and effort should be to minimise these sources of errors as much as possible. There are ways of reducing the errors.

Some of these are:

(a) Designing and executing a good questionnaire;

(b) Selection of appropriate sampling method;

(c) Adequate sample size;

(d) Employing trained investigators to collect the data; and

(e) Care in editing, coding and entering the data into the computer.

**Module 2**

**Data Sources, Scaling Techniques, Methods of Data Collection & Analysis**

**Meaning and Need for Data**

Data is required to make a decision in any business situation. The researcher is faced with one of the most difficult problems of obtaining suitable, accurate and adequate data. Utmost care must be exercised while collecting data because the quality of the research results depends upon the reliability of the data.

Suppose, you are the Director of your company. Your Board of Directors has asked you to find out why the profit of the company has decreased since the last two years. Your Board wants you to present facts and figures. What are you going to do?

The first and foremost task is to collect the relevant information to make an analysis for the above mentioned problem. It is, therefore, the information collected from various sources, which can be expressed in quantitative form, for a specific purpose, which is called data. The rational decision maker seeks to evaluate information in order to select the course of action that maximizes objectives. For decision making, the input data must be appropriate. This depends on the appropriateness of the method chosen for data collection. The application of a statistical technique is possible when the questions are answerable in quantitative nature, for instance; the cost of production, and profit of the company measured in rupees, age of the workers in the company measured in years. Therefore, the first step in statistical activities is to gather data. The data may be classified as primary and secondary data. Let us now discuss these two kinds of data in detail.

**Primary and Secondary Data**

The **Primary data** are original data which are collected for the first time for a specific purpose. Such data are published by authorities who themselves are responsible for their collection. The **Secondary data** on the other hand, are those which have already been collected by some other agency and which have

already been processed. Secondary data may be available in the form of **published** or **unpublished** sources. For instance, population census data collected by the Government in a country is primary data for that Government. But the same data becomes secondary for those researchers who use it later.

In case you have decided to collect primary data for your investigation, you have to identify the sources from where you can collect that data. For example, if you wish to study the problems of the workers of X Company Ltd., then the workers who are working in that company are the source. On the other hand,

if you have decided to use secondary data, you have to identify the secondary source who have already collected the related data for their study purpose.

With the above discussion, we can understand that the difference between primary and secondary data is only in terms of degree. That is that the data which is primary in the hands of one becomes secondary in the hands of another.

**Sources of Secondary Data**

We have discussed above the meaning of primary and secondary data. Sometimes, it is not possible to collect primary data due to time, cost and human resource constraints. Therefore, researchers have to take the help of secondary data.

Now let us discuss, (a) various sources from where, one can get secondary data,

(b) precautions while using secondary data, its merits and Demerits and some documentary and electronic sources of data in India.

**Documentary Sources of Data**

This category of secondary data source may also be termed as **Paper Source**.

The main sources of documentary data can be broadly classified into two categories:

a) Published sources, and

b) Unpublished sources.

Let us discuss these two categories in detail.

a) **Published Sources**

There are various national and international institutions, semi-official reports of various committees and commissions and private publications which collect and publish statistical data relating to industry, trade, commerce, health etc. These publications of various organisations are useful sources of secondary data.

These are as follows:

1) **Government Publications:** Central and State Governments publish current information alongwith statistical data on various subjects, quarterly and annually.

For example, Monthly Statistical Abstract, National Income Statistics, Economic Survey, Reports of National Council of Applied Economic Research (NCEAR), Federation of Indian Chambers of Commerce and Industry (FICCI), Indian Council of Agricultural Research (ICAR), Central Statistical Organisation (CSO), etc.

2) **International Publications:** The United Nations Organisation (UNO), International Labour Organisation (ILO), International Monetary Fund (IMF), World Bank, Asian Development Bank (ADB) etc., also publish relevant data and reports.

3) **Semi-official Publications:** Semi-official organisations like Corporations, District Boards, Panchayat etc. publish reports.

4) **Committees and Commissions:** Several committees and commissions appointed by State and Central Governments provide useful secondary data. For example, the report of the 10th Financial Commission or Fifth Pay Commissions etc.

5) **Private Publications:** Newspapers and journals publish the data on different fields of Economics, Commerce and Trade. For example, Economic Times, Financial Express etc. and Journals like Economist, Economic and Political Weekly, Indian Journal of Commerce, Journal of Industry and Trade, Business Today etc. Some of the research and financial institutions also publish their reports annually like Indian Institute of Finance. In addition to this, reports prepared by research scholars, universities etc. also provide secondary source of information.

b) **Unpublished Sources**

It is not necessary that all the information/data maintained by the institutions or individuals are available in published form. Certain research institutions, trade associations, universities, research scholars, private firms, business institutions etc., do collect data but they normally do not publish it. We can get this information from their registers, files etc.

**Measurement and Scaling Techniques**

In our daily life we are said to measure when we use some yardstick to determine weight, height, or some other feature of a physical object. We also measure when we judge how well we like a song, a painting or the personalities of our friends. We, thus, measure physical objects as well as abstract concepts.

Measurement is a relatively complex and demanding task, specially so when it concerns qualitative or abstract phenomena. By measurement we mean the process of assigning numbers to objects or observations, the level of measurement being a function of the rules under which the

numbers are assigned.

Before we proceed further it will be worthwhile to understand the following two terms:

(a) Measurement, and (b) Scaling.

a) **Measurement:** Measurement is the process of observing and recording the observations that are collected as part of research. The recording of the observations may be in terms of numbers or other symbols to characteristics of objects according to certain prescribed rules. The respondent’s, characteristics are feelings, attitudes, opinions etc. For example, you may assign ‘1’ for Male and ‘2’ for Female respondents. In response to a question on whether he/she is using the ATM provided by a particular bank branch, the respondent may say ‘yes’ or ‘no’. You may wish to assign the number ‘1’ for the response yes and ‘2’ for the response no. We assign numbers to these characteristics for two reasons. First, the numbers facilitate further statistical analysis of data obtained. Second, numbers facilitate the communication of measurement rules and results. The most important aspect of measurement is the specification of rules for assigning numbers to characteristics. The rules for assigning numbers should be

standardised and applied uniformly. This must not change over time or objects.

b) **Scaling:** Scaling is the assignment of objects to numbers or semantics according to a rule. In scaling, the objects are text statements, usually statements of attitude, opinion, or feeling. For example, consider a scale locating customers of a bank according to the characteristic “agreement to the satisfactory quality of

service provided by the branch”. Each customer interviewed may respond with a semantic like ‘strongly agree’, or ‘somewhat agree’, or ‘somewhat disagree’, or ‘strongly disagree’. We may even assign each of the responses a number. For example, we may assign strongly agree as ‘1’, agree as ‘2’ disagree as ‘3’,

and strongly disagree as ‘4’. Therefore, each of the respondents may assign 1, 2, 3 or 4.

**Measurement Scales**

From what has been stated above, we can write that scales of measurement can be considered in

terms of their mathematical properties. The most widely used classification of measurement scales

are: (a) nominal scale; (b) ordinal scale; (c) interval scale; and (d) ratio scale.

**(a) Nominal scale:**

Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved. For example, one cannot usefully average the numbers on the back of a group of football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one group with the numbers assigned to another. The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measure of dispersion for nominal scales.

Chi-square test is the most common test of statistical significance that can be utilized, and for the

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

**(b) Ordinal scale:**

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

Thus, the use of an ordinal scale implies a statement of ‘greater than’ or ‘less than’ (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

**(c) Interval scale:**

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

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

**(d) Ratio scale:**

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

Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, etc. are examples. Generally, all statistical techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated.

Thus, proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly. If the nature of the variables permits, the researcher should use the scale that provides the most precise description. Researchers in physical sciences have the advantage to describe variables in ratio scale form but the behavioural sciences are generally limited to describe variables in interval scale form, a less precise type of measurement.

**Sources of Error in Measurement**

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

**(a) Respondent:**

At times the respondent may be reluctant to express strong negative feelings or it is just possible that he may have very little knowledge but may not admit his ignorance. All this reluctance is likely to result in an interview of ‘guesses.’ Transient factors like fatigue, boredom, anxiety, etc. may limit the ability of the respondent to respond accurately and fully.

**(b) Situation:**

Situational factors may also come in the way of correct measurement. Any condition which places a strain on interview can have serious effects on the interviewer-respondent rapport. For instance, if someone else is present, he can distort responses by joining in or merely by being present. If the respondent feels that anonymity is not assured, he may be reluctant to express certain feelings.

**(c) Measurer:**

The interviewer can distort responses by rewording or reordering questions. His behaviour, style and looks may encourage or discourage certain replies from respondents. Careless mechanical processing may distort the findings. Errors may also creep in because of incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data-analysis stage.

**(d) Instrument:**

Error may arise because of the defective measuring instrument. The use of complex words, beyond the comprehension of the respondent, ambiguous meanings, poor printing, inadequate space for replies, response choice omissions, etc. are a few things that make the measuring instrument defective and may result in measurement errors. Another type of instrument deficiency is the poor sampling of the universe of items of concern.

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

**Tests of Sound Measurement**

Sound measurement must meet the tests of validity, reliability and practicality. In fact, these are the three major considerations one should use in evaluating a measurement tool. “Validity refers to the extent to which a test measures what we actually wish to measure. Reliability has to do with the accuracy and precision of a measurement procedure ... Practicality is concerned with a wide range of factors of economy, convenience, and interpretability ...”1 We briefly take up the relevant details concerning these tests of sound measurement.

**1. Test of Validity**

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

(i) Content validity;

(ii) Criterion-related validity and

(iii) Construct validity.

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

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

*Relevance:* (A criterion is relevant if it is defined in terms we judge to be the proper measure.)

*Freedom from bias:* (Freedom from bias is attained when the criterion gives each subject an equal opportunity to score well.)

*Reliability:* (A reliable criterion is stable or reproducible.)

*Availability:* (The information specified by the criterion must be available.)

In fact, a Criterion-related validity is a broad term that actually refers to (i) *Predictive validity* and *(ii) Concurrent validity.* The former refers to the usefulness of a test in predicting some futureperformance whereas the latter refers to the usefulness of a test in closely relating to other measuresof known validity. Criterion-related validity is expressed as the coefficient of correlation betweentest scores and some measure of future performance or between test scores and scores on anothermeasure of known validity.

**(iii) *Construct validity***is the most complex and abstract. A measure is said to possess construct validity to the degree that it confirms to predicted correlations with other theoretical propositions. Construct validity is the degree to which scores on a test can be accounted for by the explanatory constructs of a sound theory. For determining construct validity, we associate a set of other propositions with the results received from using our measurement instrument. If measurements on our devised scale correlate in a predicted way with these other propositions, we can conclude that there is some construct validity. If the above stated criteria and tests are met with, we may state that our measuring instrument is valid and will result in correct measurement; otherwise we shall have to look for more information and/or resort to exercise of judgement.

**2. Test of Reliability**

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

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

(i) By standardising the conditions under which the measurement takes place i.e., we must ensure that external sources of variation such as boredom, fatigue, etc., are minimised to the extent possible. That will improve stability aspect.

(ii) By carefully designed directions for measurement with no variation from group to group, by using trained and motivated persons to conduct the research and also by broadening the sample of items used. This will improve equivalence aspect.

**3. Test of Practicality**

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

**Methods of Data Collection**

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

**Collection of Primary Data**

We collect primary data during the course of doing experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interviews.\* This, in other words, means that there are several methods of collecting primary data, particularly in surveys and descriptive researches. Important ones are: (i) observation method, (ii) interview method, (iii) through questionnaires, (iv) through schedules, and (v) other methods which include (a) warranty cards; (b) distributor audits; (c) pantry audits; (d) consumer panels; (e) using mechanical devices; (f) through projective techniques; (g) depth interviews, and (h) content analysis. We briefly take up each method separately.

**Observation Method**

The observation method is the most commonly used method specially in studies relating to behavioural sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under the observation method, the information is sought by way of investigator’s own direct observation without asking from the respondent. For instance, in a study relating to consumer behaviour, the investigator instead of asking the brand of wrist watch used by the respondent, may himself look at the watch. The main advantage of this method is that subjective bias is eliminated, if observation is done accurately. Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behaviour or future intentions or attitudes. Thirdly, this method is independent of respondents’ willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method. This method is particularly suitable in studies which deal with subjects (i.e., respondents) who are not capable of giving verbal reports of their feelings for one reason or the other.

However, observation method has various limitations. Firstly, it is an expensive method. Secondly, the information provided by this method is very limited. Thirdly, sometimes unforeseen factors may interfere with the observational task. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively. While using this method, the researcher should keep in mind things like: What should be observed? How the observations should be recorded? Or how the accuracy of observation can be ensured? In case the observation is characterised by a careful definition of the units to be observed, the style of recording the observed information, standardised conditions of observation and the selection of pertinent data of observation, then the observation is called as *structured observation.* But when observation is to take place without these characteristics to be thought of in advance, the same is termed *as* *unstructured observation.* Structured observation is considered appropriate in descriptive studies, whereas in an exploratory study the observational procedure is most likely to be relatively unstructured.

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

There are several merits of the participant type of observation: (i) The researcher is enabled to record the natural behaviour of the group. (ii) The researcher can even gather information which could not easily be obtained if he observes in a disinterested fashion. (iii) The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule. But there are also certain demerits of this type of observation viz., the observer may lose the objectivity to the extent he participates emotionally; the problem of observation-control is not solved; and it may narrow-down the researcher’s range of experience.

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

**Interview Method**

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

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

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

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

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

(i) More information and that too in greater depth can be obtained.

(ii) Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.

(iii) There is greater flexibility under this method as the opportunity to restructure questions is always there, specially in case of unstructured interviews.

(iv) Observation method can as well be applied to recording verbal answers to various questions.

(v) Personal information can as well be obtained easily under this method.

(vi) Samples can be controlled more effectively as there arises no difficulty of the missing returns; non-response generally remains very low.

(vii) The interviewer can usually control which person(s) will answer the questions. This is not possible in mailed questionnaire approach. If so desired, group discussions may also be held.

(viii) The interviewer may catch the informant off-guard and thus may secure the most spontaneous reactions than would be the case if mailed questionnaire is used.

(ix) The language of the interview can be adopted to the ability or educational level of the person interviewed and as such misinterpretations concerning questions can be avoided.

(x) The interviewer can collect supplementary information about the respondent’s personal characteristics and environment which is often of great value in interpreting results.

But there are also certain weaknesses of the interview method. Among the important weaknesses, mention may be made of the following:

(i) It is a very expensive method, specially when large and widely spread geographical sample is taken.

(ii) There remains the possibility of the bias of interviewer as well as that of the respondent; there also remains the headache of supervision and control of interviewers.

(iii) Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.

(iv) This method is relatively more-time-consuming, specially when the sample is large and recalls upon the respondents are necessary.

(v) The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.

(vi) Under the interview method the organisation required for selecting, training and supervising the field-staff is more complex with formidable problems.

(vii) Interviewing at times may also introduce systematic errors.

(viii) Effective interview presupposes proper rapport with respondents that would facilitate free and frank responses. This is often a very difficult requirement.

***Pre-requisites and basic tenets of interviewing:***For successful implementation of the interview method, interviewers should be carefully selected, trained and briefed. They should be honest, sincere, hardworking, impartial and must possess the technical competence and necessary practical experience.

Occasional field checks should be made to ensure that interviewers are neither cheating, nor deviating from instructions given to them for performing their job efficiently. In addition, some provision should also be made in advance so that appropriate action may be taken if some of the selected respondents refuse to cooperate or are not available when an interviewer calls upon them.

In fact, interviewing is an art governed by certain scientific principles. Every effort should be made to create friendly atmosphere of trust and confidence, so that respondents may feel at ease while talking to and discussing with the interviewer. The interviewer must ask questions properly and intelligently and must record the responses accurately and completely. At the same time, the interviewer must answer legitimate question(s), if any, asked by the respondent and must clear any doubt that the latter has. The interviewers approach must be friendly, courteous, conversational and unbiased. The interviewer should not show surprise or disapproval of a respondent’s answer but he must keep the direction of interview in his own hand, discouraging irrelevant conversation and must make all possible effort to keep the respondent on the track.

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

1. It is more flexible in comparison to mailing method.

2. It is faster than other methods i.e., a quick way of obtaining information.

3. It is cheaper than personal interviewing method; here the cost per response is relatively low.

4. Recall is easy; callbacks are simple and economical.

5. There is a higher rate of response than what we have in mailing method; the non-response is generally very low.

6. Replies can be recorded without causing embarrassment to respondents.

7. Interviewer can explain requirements more easily.

8. At times, access can be gained to respondents who otherwise cannot be contacted for one reason or the other.

9. No field staff is required.

10. Representative and wider distribution of sample is possible.

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

1. Little time is given to respondents for considered answers; interview period is not likely to exceed five minutes in most cases.

2. Surveys are restricted to respondents who have telephone facilities.

3. Extensive geographical coverage may get restricted by cost considerations.

4. It is not suitable for intensive surveys where comprehensive answers are required to various questions.

5. Possibility of the bias of the interviewer is relatively more.

6. Questions have to be short and to the point; probes are difficult to handle.

**Collection of Data through Questionnaires**

This method of data collection is quite popular, particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organisations and even by governments. In this method a questionnaire is sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire. A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own.

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

1. There is low cost even when the universe is large and is widely spread geographically.

2. It is free from the bias of the interviewer; answers are in respondents’ own words.

3. Respondents have adequate time to give well thought out answers.

4. Respondents, who are not easily approachable, can also be reached conveniently.

5 Large samples can be made use of and thus the results can be made more dependable and reliable.

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

1. Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.

2. It can be used only when respondents are educated and cooperating.

3. The control over questionnaire may be lost once it is sent.

4. There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been despatched.

5. There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.

6. It is difficult to know whether willing respondents are truly representative.

7. This method is likely to be the slowest of all.

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

***Main aspects of a questionnaire:***

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

**1. *General form:***

So far as the general form of a questionnaire is concerned, it can either be structured or unstructured questionnaire. Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questions are presented with exactly the same wording and in the same order to all respondents. Resort is taken to this sort of standardization to ensure that all respondents reply to the same set of questions. The form of the question may be either closed (i.e., of the type ‘yes’ or ‘no’) or open (i.e., inviting free response) but should be stated in advance and not constructed during questioning. Structured questionnaires may also have fixed alternative questions in which responses of the informants are limited to the stated alternatives. Thus a highly structured questionnaire is one in which all questions and answers are specified and comments in the respondent’s own words are held to the minimum. When these characteristics are not present in a questionnaire, it can be termed as unstructured or non-structured questionnaire. More specifically, we can say that in an unstructured questionnaire, the interviewer is provided with a general guide on the type of information to be obtained, but the exact question formulation is largely his own responsibility and the replies are to be taken down in the respondent’s own words to the extent possible; in some situations tape recorders may be used to achieve this goal.

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

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

The following type of questions should generally be avoided as opening questions in a questionnaire:

1. Questions that put too great a strain on the memory or intellect of the respondent;

2. Questions of a personal character;

3. Questions related to personal wealth, etc.

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

**3. *Question formulation and wording:***

With regard to this aspect of questionnaire, the researcher should note that each question must be very clear for any sort of misunderstanding can do irreparable harm to a survey. Question should also be impartial in order not to give a biased picture of the true state of affairs. Questions should be constructed with a view to their forming a logical part of a well thought out tabulation plan. In general, all questions should meet the following standards—(a) should be easily understood; (b) should be simple i.e., should convey only one thought at a time; (c) should be concrete and should conform as much as possible to the respondent’s way of thinking. For instance, instead of asking. “How many razor blades do you use annually?” The more realistic question would be to ask, “How many razor blades did you use last week?” Concerning the form of questions, we can talk about two principal forms, viz., multiple choice question and the open-end question. In the former the respondent selects one of the alternative possible answers put to him, whereas in the latter he has to supply the answer in his own words. The question with only two possible answers (usually ‘Yes’ or ‘No’) can be taken as a special case of the multiple choice question, or can be named as a ‘closed question.’ There are some advantages and disadvantages of each possible form of question. Multiple choice or closed questions have the advantages of easy handling, simple to answer, quick and relatively inexpensive to analyse. They are most amenable to statistical analysis. Sometimes, the provision of alternative replies helps to make clear the meaning of the question. But the main drawback of fixed alternative questions is that of “putting answers in people’s mouths” i.e., they may force a statement of opinion on an issue about which the respondent does not infact have any opinion. They are not appropriate when the issue under consideration happens to be a complex one and also when the interest of the researcher is in the exploration of a process. In such situations, open-ended questions which are designed to permit a free response from the respondent rather than one limited to certain stated alternatives are considered appropriate. Such questions give the respondent considerable latitude in phrasing a reply. Getting the replies in respondent’s own words is, thus, the major advantage of open-ended questions. But one should not forget that, from an analytical point of view, open-ended questions are more difficult to handle, raising problems of interpretation, comparability and interviewer bias.\*

In practice, one rarely comes across a case when one questionnaire relies on one form of questions alone. The various forms complement each other. As such questions of different forms are included in one single questionnaire. For instance, multiple-choice questions constitute the basis of a structured questionnaire, particularly in a mail survey. But even there, various open-ended questions are generally inserted to provide a more complete picture of the respondent’s feelings and attitudes. Researcher must pay proper attention to the wordings of questions since reliable and meaningful returns depend on it to a large extent. Since words are likely to affect responses, they should be properly chosen. Simple words, which are familiar to all respondents should be employed. Words with ambiguous meanings must be avoided. Similarly, danger words, catch-words or words with emotional connotations should be avoided. Caution must also be exercised in the use of phrases which reflect upon the prestige of the respondent. Question wording, in no case, should bias the answer. In fact, question wording and formulation is an art and can only be learnt by practice.

***Essentials of a good questionnaire:***

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

**Collection of Data through Schedules**

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

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

This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organisations. Population census all over the world is conducted through this method.

**Case Study Method**

**Meaning**: The case study method is a very popular form of qualitative analysis and involves a careful and complete observation of a social unit, like a person, a family, an institution, a cultural group or even the entire community. It is a method of study in depth rather than breadth. The case study places more emphasis on the full analysis of a limited number of events or conditions and their interrelations. The case study deals with the processes that take place and their interrelationship. Thus, case study is essentially an intensive investigation of the particular unit under consideration. The object of the case study method is to locate the factors that account for the behaviour-patterns of the given unit as an integrated totality.

**Characteristics:** The important characteristics of the case study method are as under:

1. Under this method the researcher can take one single social unit or more of such units for his study purpose; he may even take a situation to study the same comprehensively.

2. Here the selected unit is studied intensively i.e., it is studied in minute details. Generally, the study extends over a long period of time to ascertain the natural history of the unit so as to obtain enough information for drawing correct inferences

3. In the context of this method we make complete study of the social unit covering all facets. Through this method we try to understand the complex of factors that are operative within a social unit as an integrated totality.

4 Under this method the approach happens to be qualitative and not quantitative. Mere quantitative information is not collected. Every possible effort is made to collect information concerning all aspects of life. As such, case study deepens our perception and gives us a clear insight into life. For instance, under this method we not only study how many crimes a man has done but shall peep into the factors that forced him to commit crimes when we are making a case study of a man as a criminal. The objective of the study may be to suggest ways to reform the criminal.

5. In respect of the case study method an effort is made to know the mutual inter-relationship of causal factors.

6. Under case study method the behaviour pattern of the concerning unit is studied directly and not by an indirect and abstract approach.

7. Case study method results in fruitful hypotheses along with the data which may be helpful in testing them, and thus it enables the generalised knowledge to get richer and richer. In its absence, generalised social science may get handicapped.

**Assumptions:** The case study method is based on several assumptions. The important assumptions may be listed as follows:

(i) The assumption of uniformity in the basic human nature in spite of the fact that human behaviour may vary according to situations.

(ii) The assumption of studying the natural history of the unit concerned.

(iii) The assumption of comprehensive study of the unit concerned.

**Major phases involved:** Major phases involved in case study are as follows:

(i) Recognition and determination of the status of the phenomenon to be investigated or the unit of attention.

(ii) Collection of data, examination and history of the given phenomenon.

(iii) Diagnosis and identification of causal factors as a basis for remedial or developmental treatment.

(iv) Application of remedial measures i.e., treatment and therapy (this phase is often characterized as case work).

(v) Follow-up programme to determine effectiveness of the treatment applied.

**Advantages:** There are several advantages of the case study method that follow from the various characteristics outlined above. Mention may be made here of the important advantages.

(i) Being an exhaustive study of a social unit, the case study method enables us to understand fully the behaviour pattern of the concerned unit. In the words of Charles Horton Cooley, “case study deepens our perception and gives us a clearer insight into life…. It gets at behaviour directly and not by an indirect and abstract approach.”

(ii) Through case study a researcher can obtain a real and enlightened record of personal experiences which would reveal man’s inner strivings, tensions and motivations that drive him to action along with the forces that direct him to adopt a certain pattern of behaviour.

(iii) This method enables the researcher to trace out the natural history of the social unit and its relationship with the social factors and the forces involved in its surrounding environment.

(iv) It helps in formulating relevant hypotheses along with the data which may be helpful in testing them. Case studies, thus, enable the generalised knowledge to get richer and richer.

(v) The method facilitates intensive study of social units which is generally not possible if we use either the observation method or the method of collecting information through schedules. This is the reason why case study method is being frequently used, particularly in social researches.

(vi) Information collected under the case study method helps a lot to the researcher in the task of constructing the appropriate questionnaire or schedule for the said task requires thorough knowledge of the concerning universe.

(vii) The researcher can use one or more of the several research methods under the case study method depending upon the prevalent circumstances. In other words, the use of different methods such as depth interviews, questionnaires, documents, study reports of individuals, letters, and the like is possible under case study method.

(viii) Case study method has proved beneficial in determining the nature of units to be studied along with the nature of the universe. This is the reason why at times the case study method is alternatively known as “mode of organising data”.

(ix) This method is a means to well understand the past of a social unit because of its emphasis of historical analysis. Besides, it is also a technique to suggest measures for improvement in the context of the present environment of the concerned social units.

(x) Case studies constitute the perfect type of sociological material as they represent a real record of personal experiences which very often escape the attention of most of the skilled researchers using other techniques.

(xi) Case study method enhances the experience of the researcher and this in turn increases his analysing ability and skill.

(xii) This method makes possible the study of social changes. On account of the minute study of the different facets of a social unit, the researcher can well understand the social change then and now. This also facilitates the drawing of inferences and helps in maintaining the continuity of the research process. In fact, it may be considered the gateway to and at the same time the final destination of abstract knowledge.

(xiii) Case study techniques are indispensable for therapeutic and administrative purposes. They are also of immense value in taking decisions regarding several management problems. Case data are quite useful for diagnosis, therapy and other practical case problems.

**Limitations:** Important limitations of the case study method may as well be highlighted.

(i) Case situations are seldom comparable and as such the information gathered in case studies is often not comparable. Since the subject under case study tells history in his own words, logical concepts and units of scientific classification have to be read into it or out of it by the investigator.

(ii) Read Bain does not consider the case data as significant scientific data since they do not provide knowledge of the “impersonal, universal, non-ethical, non-practical, repetitive aspects of phenomena.” Real information is often not collected because the subjectivity of the researcher does enter in the collection of information in a case study.

(iii) The danger of false generalisation is always there in view of the fact that no set rules are followed in collection of the information and only few units are studied.

(iv) It consumes more time and requires lot of expenditure. More time is needed under case study method since one studies the natural history cycles of social units and that too minutely.

(v) The case data are often vitiated because the subject, according to Read Bain, may write what he thinks the investigator wants; and the greater the rapport, the more subjective the whole process is.

(vi) Case study method is based on several assumptions which may not be very realistic at times, and as such the usefulness of case data is always subject to doubt.

(vii) Case study method can be used only in a limited sphere. It is not possible to use it in case of a big society. Sampling is also not possible under a case study method.

(viii) Response of the investigator is an important limitation of the case study method. He often thinks that he has full knowledge of the unit and can himself answer about it. In case the same is not true, then consequences follow. In fact, this is more the fault of the researcher rather than that of the case method.

**Module-3**

**Data Processing and Data Analysis**

The data, after collection, has to be processed and analysed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis.

Technically speaking, processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis. The term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. Thus, “in the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions”.

**Processing Operations**

With this brief introduction concerning the concepts of processing and analysis, we can now proceed with the explanation of all the processing operations.

**1. Editing:**

Editing of data is a process of examining the collected raw data (specially in surveys) to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaires and/or schedules. Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as completed as possible and have been well arranged to facilitate coding and tabulation.

With regard to points or stages at which editing should be done, one can talk of field editing and central editing. *Field editing* consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated and/or in illegible form at the time of recording the respondents’ responses. This type of editing is necessary in view of the fact that individual writing styles often can be difficult for others to decipher. This sort of editing should be done as soon as possible after the interview, preferably on the very day or on the next day.While doing field editing, the investigator must restrain himself and must not correct errors of omission by simply guessing what the informant would have said if the question had been asked.

*Central editing* should take place when all forms or schedules have been completed and returned to the office. This type of editing implies that all forms should get a thorough editing by a single editor in a small study and by a team of editors in case of a large inquiry. Editor(s) may correct the obvious errors such as an entry in the wrong place, entry recorded in months when it should have been recorded in weeks, and the like. In case of inappropriate on missing replies, the editor can sometimes determine the proper answer by reviewing the other information in the schedule. At times, the respondent can be contacted for clarification. The editor must strike out the answer if the same is inappropriate and he has no basis for determining the correct answer or the response. In such a case an editing entry of ‘no answer’ is called for. All the wrong replies, which are quite obvious, must be dropped from the final results, especially in the context of mail surveys.

Editors must keep in view several points while performing their work:

(a) They should be familiar with instructions given to the interviewers and coders as well as with the editing instructions supplied to them for the purpose.

(b) While crossing out an original entry for one reason or another, they should just draw a single line on it so that the same may remain legible.

(c) They must make entries (if any) on the form in some distinctive colur and that too in a standardised form.

(d) They should initial all answers which they change or supply.

(e) Editor’s initials and the date of editing should be placed on each completed form or schedule.

**2. Coding:**

Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristic of exhaustiveness (i.e., there must be a class for every data item) and also that of mutual exclusively which means that a specific answer can be placed in one and only one cell in a given category set. Another rule to be observed is that of unidimensionality by which is meant that every class is defined in terms of only one concept.

Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. This makes it possible to precode the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires. But in case of hand coding some standard method may be used. One such standard method is to code in the margin with a coloured pencil. The other method can be to transcribe the data from the questionnaire to a coding sheet. Whatever method is adopted, one should see that coding errors are altogether eliminated or reduced to the minimum level.

**3. Classification:**

Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

**(a) *Classification according to attributes:***As stated above, data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.). Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain attributes are known as *statistics of attributes* and their classification is said to be classification according to attributes.

Such classification can be simple classification or manifold classification. In simple classification we consider only one attribute and divide the universe into two classes—one class consisting of items possessing the given attribute and the other class consisting of items which do not possess the given attribute. But in manifold classification we consider two or more attributes simultaneously, and divide that data into a number of classes (total number of classes of final order is given by 2*n*, where *n* = number of attributes considered).\*

Whenever data are classified according to attributes, the researcher must see that the attributes are defined in such a manner that there is least possibility of any doubt/ambiguity concerning the said attributes.

**(b) *Classification according to class-intervals:***

Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomenon which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as *statistics of variables* and are classified on the basis of class intervals. For instance, persons whose incomes, say, are within Rs 201 to Rs 400 can form one group, those whose incomes are within Rs 401 to Rs 600 can form another group and so on. In this way the entire data may be divided into a number of groups or classes or what are usually called, ‘class-intervals.’ Each group of class-interval, thus, has an upper limit as well as a lower limit which are known as class limits. The difference between the two class limits is known as class magnitude. We may have classes with equal class magnitudes or with unequal class magnitudes. The number of items which fall in a given class is known as the frequency of the given class. All the classes or groups, with their respective frequencies taken together and put in the form of a table, are described as group frequency distribution or simply frequency distribution.

**4. Tabulation:**

When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise and logical order. This procedure is referred to as tabulation. Thus, tabulation is the process of summarising raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis. In a broader sense, tabulation is an orderly arrangement of data in columns and rows.

Tabulation is essential because of the following reasons.

1. It conserves space and reduces explanatory and descriptive statement to a minimum.

2. It facilitates the process of comparison.

3. It facilitates the summation of items and the detection of errors and omissions.

4. It provides a basis for various statistical computations.

Tabulation can be done by hand or by mechanical or electronic devices. The choice depends on the size and type of study, cost considerations, time pressures and the availaibility of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation if other factors are favourable and necessary facilities are available. Hand tabulation is usually preferred in case of small inquiries where the number of questionnaires is small and they are of relatively short length. Hand tabulation may be done using the direct tally, the list and tally or the card sort and count methods. When there are simple codes, it is feasible to tally directly from the questionnaire. Under this method, the codes are written on a sheet of paper, called tally sheet, and for each response a stroke is marked against the code in which it falls. Usually after every four strokes against a particular code, the fifth response is indicated by drawing a diagonal or horizontal line through the strokes. These groups of five are easy to count and the data are sorted against each code conveniently. In the listing method, the code responses may be transcribed onto a large work-sheet, allowing a line for each questionnaire. This way a large number of questionnaires can be listed on one work sheet. Tallies are then made for each question. The card sorting method is the most flexible hand tabulation. In this method the data are recorded on special cards of convenient size and shape with a series of holes. Each hole stands for a code and when cards are stacked, a needle passes through particular hole representing a particular code. These cards are then separated and counted.

In this way frequencies of various codes can be found out by the repetition of this technique. We can as well use the mechanical devices or the computer facility for tabulation purpose in case we want quick results, our budget permits their use and we have a large volume of straight forward tabulation involving a number of cross-breaks.

Tabulation may also be classified as simple and complex tabulation. The former type of tabulation gives information about one or more groups of independent questions, whereas the latter type of tabulation shows the division of data in two or more categories and as such is designed to give information concerning one or more sets of inter-related questions. Simple tabulation generally results in one-way tables which supply answers to questions about one characteristic of data only. As against this, complex tabulation usually results in two-way tables (which give information about two inter-related characteristics of data), three-way tables (giving information about three interrelated characteristics of data) or still higher order tables, also known as manifold tables, which supply information about several interrelated characteristics of data. Two-way tables, three-way tables or manifold tables are all examples of what is sometimes described as cross tabulation.

**Transcription:**

Transcription is the process of providing a written account of audio information/spoken words. Usually in qualitative research interviews, focus group discussions, brain storming etc will be recorded through audio-visual aids. Transcription is the process of transferring the audio information into written form. Generally it is done verbatim that is exactly word to word information is written.

**Types of Analysis:**

Univariate data means data of a single variable. For a single variable data, the measures of central tendencies like mean median and mode can be estimated along with the measures of dispersion.

Bivariate data means data of two variables. Along with measures of central tendencies and dispersion for both the variables, both the variables can be analyzed with the help of correlation, simple regression and also coefficient of variation.

Multivariate data means the data of more than two variables. The popular techniques that are available for multivariate data analysis are multiple regression, one way ANOVA, multiple way ANOVA, factor analysis, principal component analysis etc.

Descriptive Vs Inferential Analysis

* Descriptive Statistics deals with data in the following ways
  + Collect
  + Organize
  + Summarize
  + Display
  + Analyze/Estimation
* Inferential Statistics
  + Predict and forecast values of population parameters
  + Test hypotheses about values of population parameters
  + Make decisions
* **Inferential statistics** are methods used to make conclusions or inferences concerning some unknown aspect of a **population** based on data from a **sample**. Statistical inference is the process of making prediction, or decision about a population based on a sample.

**Normal Distribution**

A **continuous random variable** is a random variable that can take on any value in an interval of numbers.

The probabilities associated with a continuous random variable X are determined by the **probability density function** of the random variable.

The function, denoted *f(x)*, has the following properties.

1. *f(x) 0* for all *x*.

2. The probability that *X* will be between two numbers *a* and *b* is equal to the area under

*f(x)* between *a* and *b*.

3. The total area under the curve of *f(x)*is equal to 1.00.

* **The normal is a *family* of**
  + ***Bell-shaped* and *symmetric* distributions. because the distribution is symmetric, one-half (.50 or 50%) lies on either side of the mean.**
  + **Each is characterized by a different pair of *mean, µ,* and *variance, σ2* . That is: [X~N(*µ*, *σ2*)].**
  + **Each is *asymptotic* to the horizontal axis.**
  + **The area under any normal probability density function within *k*** σ**of** µ **is the same for any normal distribution, regardless of the mean and variance.**

**Normal Probability Density Function:**

**Normal distribution and its properties**

The important properties of the normal distribution are:

1. The normal curve is “bell shaped” and symmetrical in nature. The distribution of the frequencies on either side of the maximum ordinate of the curve is similar with each other.

2. The maximum ordinate of the normal curve is at x = µ. Hence the mean, median and mode of the normal distribution coincide.

3. It ranges between - ∞ to + ∞

4. The value of the maximum ordinate is.

5. The points where the curve change from convex to concave or vice versa is at X = µ ± σ.

6. The first and third quartiles are equidistant from median.

7. The area under the normal curve distribution are:

a. µ ± 1σ covers 68.27% area;

b. µ ± 2σ covers 95.45% area.

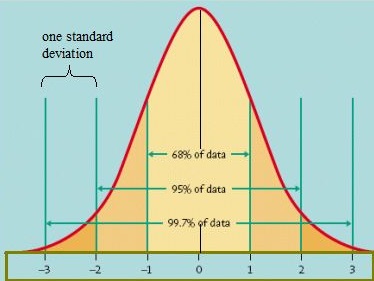
c. µ ± 3σ covers 99.73% area.

**This property helps in the decision making of whether to reject or not to reject a null hypothesis.**

8. When µ = 0 and σ = 1, then the normal distribution will be a standard normal curve. The **standard normal random variable**, Z, is the normal random variable with mean µ = 0 and standard deviation σ = 1: Z~N(0,12).

9. Any normal random variable x can be transformed to standard normal variable with the help of following transformation

10. Central limit theorem- When sampling from a population with mean µ and finite standard deviation σ, the sampling distribution of the sample mean will tend to a normal distribution with mean µ and finite standard deviation as the sample size becomes large (*n* >30). For large n. the mean of the sample is distributed normally with mean µ and variance .



**Testing of Hypothesis**

Hypothesis testing determines the validity of the assumption (technically described as null hypothesis) with a view to choose between two conflicting hypotheses about the value of a population parameter. Hypothesis testing helps to decide on the basis of a sample data, whether a hypothesis about the population is likely to be true or false.

The actual test begins by considering two hypotheses. They are called the null hypothesis and the alternate hypothesis. These hypotheses contain opposing viewpoints.

Ho: The null hypothesis: It is a statement about the population that will be assumed to be true unless it can be shown to be incorrect beyond a reasonable doubt.

Ha: The alternate hypothesis: It is a claim about the population that is contradictory to Ho and what we conclude when we reject Ho.

A hypothesis test involves collecting data from a sample and evaluating the data. Then, the statistician makes a decision as to whether or not there is sufficient evidence based upon analyses of the data, to reject the null hypothesis.

Statisticians have developed several tests of hypotheses (also known as the tests of significance) for the purpose of testing of hypotheses which can be classified as:

(a) Parametric tests or standard tests of hypotheses; and

(b) Non-parametric tests or distribution-free test of hypotheses.

Parametric tests usually assume certain properties of the parent population from which we draw samples. Assumptions like observations come from a normal population, sample size is large, assumptions about the population parameters like mean, variance, etc., must hold good before parametric tests can be used.

But there are situations when the researcher cannot or does not want to make such assumptions. In such situations we use statistical methods for testing hypotheses which are called non-parametric tests because such tests do not depend on any assumption about the parameters of the parent population. Besides, most non-parametric tests assume only nominal or ordinal data, whereas parametric tests require measurement equivalent to at least an interval scale. As a result, non-parametric tests need more observations than parametric tests to achieve the samesize of Type I and Type II errors.

**Important Parametric Tests**

The important parametric tests are: (1) *z*-test; (2) *t*-test; (\*3) 2-test, and (4) *F*-test. All these tests

are based on the assumption of normality i.e., the source of data is considered to be normally distributed.

***z-test***is based on the normal probability distribution and is used for judging the **significance of** several statistical measures, **particularly the mean.** The relevant test statistic\*, *z*, is worked out and compared with its probable value (to be read from table showing area under normal curve) at a specified level of significance for judging the significance of the measure concerned. This is most frequently used test in research studies. This test is used even when binomial distribution or *t*-distribution is applicable on the presumption that such a distribution tends to approximate normal distribution as ‘*n*’ becomes larger.

*z*-test is generally used for comparing the mean of a sample to some hypothesised mean for the population in case of large sample, or when population variance is known. *z*-test is also used for judging the significance of difference between means of two independent samples in case of large samples, or when population variance is known. ***z*-test** is also used for comparing the sample proportion to a theoretical value of population proportion or **for judging the difference in proportions of two independent samples when *n* happens to be large.**

Besides, this test may be used for judging the significance of median, mode, coefficient of correlation and several other measures.

***t-test***is based on *t*-distribution and is considered an appropriate test for judging the significance of a sample mean or for judging **the significance of difference between the means of two samples** in case of small sample(s) when population variance is not known (in which case we use variance of the sample as an estimate of the population variance).

In case two samples are related, we use ***paired t-test***(or what is known as difference test) **for judging the significance of the mean of difference between the two related samples**. It can also be used for judging the significance of the **coefficients of simple and partial correlations**. The relevant test statistic, *t*, is calculated from the sample data and then compared with its probable value based on *t*-distribution (to be read from the table that gives probable values of *t* for different levels of significance for different degrees of freedom) at a specified level of significance for concerning degrees of freedom for accepting or rejecting the null hypothesis. It may be noted that *t*-test applies only in case of small sample(s) when population variance is unknown.

**2 -*test***is based on chi-square distribution and as a parametric test is used for comparing a sample variance to a theoretical population variance.

***F-test***is based on *F*-distribution and is used to compare the variance of the two-independent samples. This test is also used in the context of analysis of variance (ANOVA) for judging the significance of more than two sample means at one and the same time. It is also used for judging the significance of multiple correlation coefficients. Test statistic, *F*, is calculated and compared with its probable value (to be seen in the *F*-ratio tables for different degrees of freedom for greater and smaller variances at specified level of significance) for accepting or rejecting the null hypothesis.

**Steps in Testing of Hypothesis**

To perform a hypothesis test, a statistician will:

1. Set up two contradictory hypotheses.

2. Collect sample data.

3. Determine the correct distribution/statistic to perform the hypothesis test.

4. Estimate the required statistic by performing the calculations that ultimately will allow you to reject or fail to reject the null hypothesis.

5. Make a decision and write a meaningful conclusion.

Since the null and alternate hypotheses are contradictory, you must examine evidence to decide if you have enough evidence to reject the null hypothesis or not. The evidence is in the form of sample data. After you have determined which hypothesis the sample supports, you make a decision. There are two options for a decision. They are "reject Ho" if the sample information favors the alternate hypothesis or "do not reject Ho" or "fail to reject Ho" if the sample information is insufficient to reject the null hypothesis.

**Decision and Conclusion**

A systematic way to make a decision of whether to reject or not reject the null hypothesis is to compare the p-value and a preset or preconceived α (also called a "significance level"). A preset α is the probability of a Type I error (rejecting the null hypothesis when the null hypothesis is true). It may or may not be given to you at the beginning of the problem.

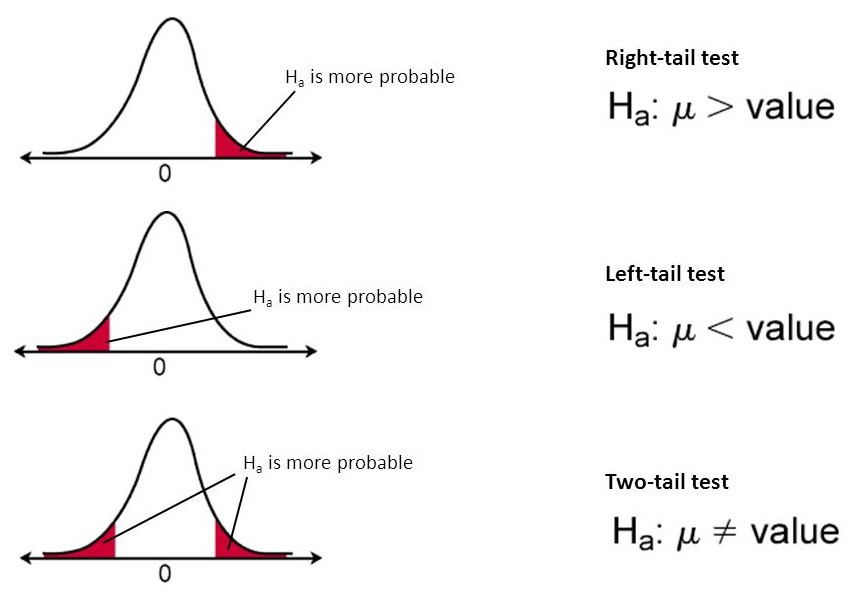
When you make a decision to reject or not reject Ho, do as follows:

• If α > p-value, reject Ho. The results of the sample data are significant. There is sufficient evidence to conclude that Ho is an incorrect belief and that the alternative hypothesis, Ha, may be correct.

• If α ≤ p-value, do not reject Ho. The results of the sample data are not significant. There is not sufficient evidence to conclude that the alternative hypothesis, Ha, may be correct.

• When you "do not reject Ho", it does not mean that you should believe that Ho is true. It simply means that the sample data have failed to provide sufficient evidence to cast serious doubt about the truthfulness of Ho.

Conclusion: After you make your decision, write a thoughtful conclusion about the hypotheses in terms of the given problem.



**Type I and Type II Errors:** As regards the testing of hypotheses, a researcher can make basically two types of errors. He/she may reject H0 when it is true, or accept H0 when it is not true. The former is called as Type I error and the latter is known as Type II error. In other words, Type I error implies the rejection of a hypothesis when it must have been accepted, while Type II error implies the acceptance of a hypothesis which must have been rejected. Type I error is denoted by α (alpha) and is known as α error, while Type II error is usually denoted by β (beta) and is known as β error.

|  |  |  |
| --- | --- | --- |
| **Hypothesis** | **Decision** | |
| Accept Ho | Reject Ho |
| Ho (True) | Correct Decision | Type I Error (α error) |
| Ho (False) | Type II Error (β error) | Correct Decision |

**One-Tailed and Two-Tailed Tests:** These two types of tests are very important in the context of hypothesis testing.

A two-tailed test rejects the null hypothesis, when the sample mean is significantly greater or lower than the hypothesized value of the mean of the population. Such a test is suitable when the null hypothesis is some specified value; the alternative hypothesis is a value that is not equal to the specified value of the null hypothesis. This leads to use of both sides of the distribution and hence the name two tailed test

Null Hypothesis Ho : = 100 Alternative Hypothesis Ha : 100

In One-Tailed test, the sample mean is **either** greater or lower than the hypothesized value of the mean of the population. This leads to the use of only one side of the distribution and hence called one tailed test.

Null Hypothesis Ho: = 100 Alternative Hypothesis Ha : > 100

Null Hypothesis Ho: = 100 Alternative Hypothesis Ha : < 100

**Estimation of Mean and Variance**

Let us take the following ungrouped data and obtain the mean, variance & Standard deviation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.No** | 1 | 2 | 3 | 4 |
| **x Income** | 2 | 5 | 7 | 6 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 2  5  7  6 | (2-5)=3  (5-5)=0  (7-5)=2  (6-5)=1 | 9  0  4  1 |
|  | **0** | **14** |

=5

**Variance σ² = for population**

**Variance σ² = for sample**

**Standard deviation is the positive square root of variance.**

**Tests of Single Sample Mean**

As mentioned earlier Z and t test are applied for testing the significance of mean/average of a sample (Univariate data).

**Testing hypothesis about population mean**

If there is a random sample froma normal population with mean µ and variance **σ² , we can test about the** µ when **σ² is known or unknown.**

* When **σ² is known: The test statistic is given as**

**Z =**

The Z follows standard normal distribution. The null hypothesis is specified as Ho: µ= µo

Eg: A random sample of 1000 members is found to have a mean of 342cm.Could it be reasonably regarded as a sample from a large population whose mean is 3.30 cm and standard deviation is 2.6cm?

Z =  **= ==1.46**

Since 1.46 < 1.96 where 1.96 is the z table/critical value at 5% level, of significance, the null hypothesis is not rejected. It can be concluded that sample is drawn from the given population.

* When σ² is not known : When the population variance is not known, the variance and standard deviation of the sample are estimated. The test statistic is given as

t = where s is the standard deviation of the sample.

Eg: The weight in grams of plums consumed by 4 people everyday are 27.1, 28.1, 27.0, and 28.0. Can it be regarded as a sample drawn from a population with mean of 25 grams.

Ho: =25, Ha:

The test statistic is t =

S is sample standard deviation which comes to 0.580

t = = 8.79

The critical value of t statistic at 5% level of significance is 2.353 for 3 degrees of freedom(n-1=4-1=3 df). Since 8.79>2.353, the null hypothesis is rejected. It can be concluded that the sample does not come from a population with mean of 25 grams.

**Test of Significance for the Difference of Means (Two Samples)**

**Let** 1 and 2 be the means of two independent samples of size n1 and n2 from two different populations with standard deviations σ1 and σ2 respectively.

The variance of1 - 2 is σ1²/ n1 + σ2²/ n2

The Ho is there is no difference between the population means. The test statistic is given as

Z=

The Ho is accepted at 5% level of significance if the estimated Z value is lesser than 1.96 otherwise its rejected.

If σ1² = σ2² = σ², then the

Z=

Eg: The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches?

n1= 1000 and n2 =2000, 1=67.5and 2=68.0

Z= = = -5.1

Since the absolute value of Z that is 5.1 is greater than 1.96, the null hypothesis is rejected. It can be concluded that the samples are not drawn from the population with standard deviation 2.5 inches.

Eg: In a survey of buying habits, 400 women shoppers are chosen at random in super market ‘A’. Their average weekly food expenditure is Rs.250 with a standard deviation of Rs.40. For another 400 women shoppers chosen at random in super market ‘B’, the average weekly food expenditure is Rs.220 with a standard deviation of Rs.55.Test at 1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.

n1= 400, n2 =400, 1=250, 2=200, σ1 = 40and σ2 = 55

Z= = = = = 14.70

The estimated Z value is greater than the Z critical value 3(2.97) at 1% level of significance. Hence the null hypothesis of average weekly food expenditure of two population shoppers being equal stands to be rejected.

**Independent Samples Test/ Testing the Differences in Two Means**

When the samples are small and population variances are not known but assumed to be equal for both the samples, then independent samples t test is applied to test the differences between two samples. The test statistic t is estimated as given below

t= with degrees of freedom.

Eg: Sample of Sales in similar shops in two towns are taken for a new product with following results:

|  |  |  |  |
| --- | --- | --- | --- |
| Town | Mean sales | Variance | Sample Size |
| 1 | 57 | 5.3 | 5 |
| 2 | 61 | 4.8 | 7 |

Is there evidence of difference in sales in two towns? Use 5% level of significance for testing this difference between the means of two samples.

Ho: µ1 = µ2

Ha: µ1 : µ2

n1= 5, n2 =7, 1=57, 2=61, σ1²= 5.3and σ2²= 4.8

t= = =

= -3.053 with 10 df (5+7-2)

The estimated t value 3.053 is greater than the t critical value 2.228 for 10df at 5% level of significance. Hence the null hypothesis of no difference is rejected. It can be concluded that there is significant difference in the sales of two towns.

**Testing the Means of Related Samples/Paired t test**

Paired t test is a parametric test for comparing two related samples, involving small values of n that does not require the variances of two populations to be equal. It is necessary that the observations in the two samples be collected in the form of what is called as matched pairs. That is one observation in one sample must be paired with an observation in another sample. The differences are calculated for matched pairs. Such a test is useful for befor and after treatment study or in impact studies where the impact of interventions are being examined.

The t statistic that has to be estimated is

t = with df n-1

Mean of Differences

Variance of Differences = or

n = number of matched pairs.

Eg: A certain stimulus was administered to each of the 12 patients resulted in the following increase of blood pressure: 5, 2, 8, -1, 3, 0, -2, 1,5, 0,4 and 6. Can it be concluded that the stimulus is generally accompanied by an increase in blood pressure.

|  |  |
| --- | --- |
| Di | Di2 |
| 5 | 25 |
| 2 | 4 |
| 8 | 64 |
| -1 | 1 |
| 3 | 9 |
| 0 | 0 |
| -2 | 4 |
| 1 | 1 |
| 5 | 25 |
| 0 | 0 |
| 4 | 16 |
| 6 | 36 |
| Total 31 | 185 |

= = = 9.55

= 3.09

t = = = = = 2.89

The table/critical value of t for 11 df (12-1) at 5% level of significance is 2.201. The estimated value of t statistic 2.89 is greater than the table value 2.201. Hence the null hypothesis of no difference is rejected. It can be concluded that the stimulus is accompanied by an increase in blood pressure.

**Test of Single Sample Variance with a Hypothesised Population Variance**

Chi-square test is used to compare the sample variance to some theoretical or hypothesised variance of population. The chi-square statistic used is

(n-1) with n-1 is the degrees of freedom

where σs² = Variance of Sample

σP² = Variance of the Population

n-1 is the degrees of freedom

Eg: A sample of 10 is drawn randomly from a certain population. The sum of the squared deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at

5 perc ent level of significance.

The data provided is n=10,

The Variance of Sample σs² =

(n-1) = (10-1) = = 10

Degrees of freedom = (10 – 1) = 9.

The table value of c2 at 5 per cent level for 9 df. is 16.92. The calculated value of c2 is less than this table value, so we accept the null hypothesis and conclude that the variance of the population is 5 as given in the question.

**Testing the Equality of Variances of Two Normal Populations**

For testing the equality of variances of two normal populations, F test is employed. The null hypothesis happens to be

Ho: σP1² = σP2² where σP1² and σP2² represent the variances of two normal populations. This hypothesis is tested on the basis of sample data and F statistic is estimated using σs1² and σs2².

**F =** with v1 and v2 df where v1 is n1-1 and v2 is n2-2

v1 and v2 are df for numerator and denominator respectively. n1and n2 are sample observations in sample1 and sample 2.

**While calculating F statistic, we should always keep in mind that >. That is the sample with greater variance should be in the numerator. Since F test is parametric test, it is assumed that**

* **The populations are normal;**
* **Samples have been drawn randomly; and**
* **Observations are independent**

**The F test is used frequently in Analysis of Variance or ANOVA. In ANOVA , the ratio of variances follows an F-distribution:**

****

* **The F-test tests the hypothesis that two variances are equal.**
* **F will be close to 1 if sample variances are equal.**

Eg: The following table table gives the number of units produced per day by two workers A and B for a number of days:

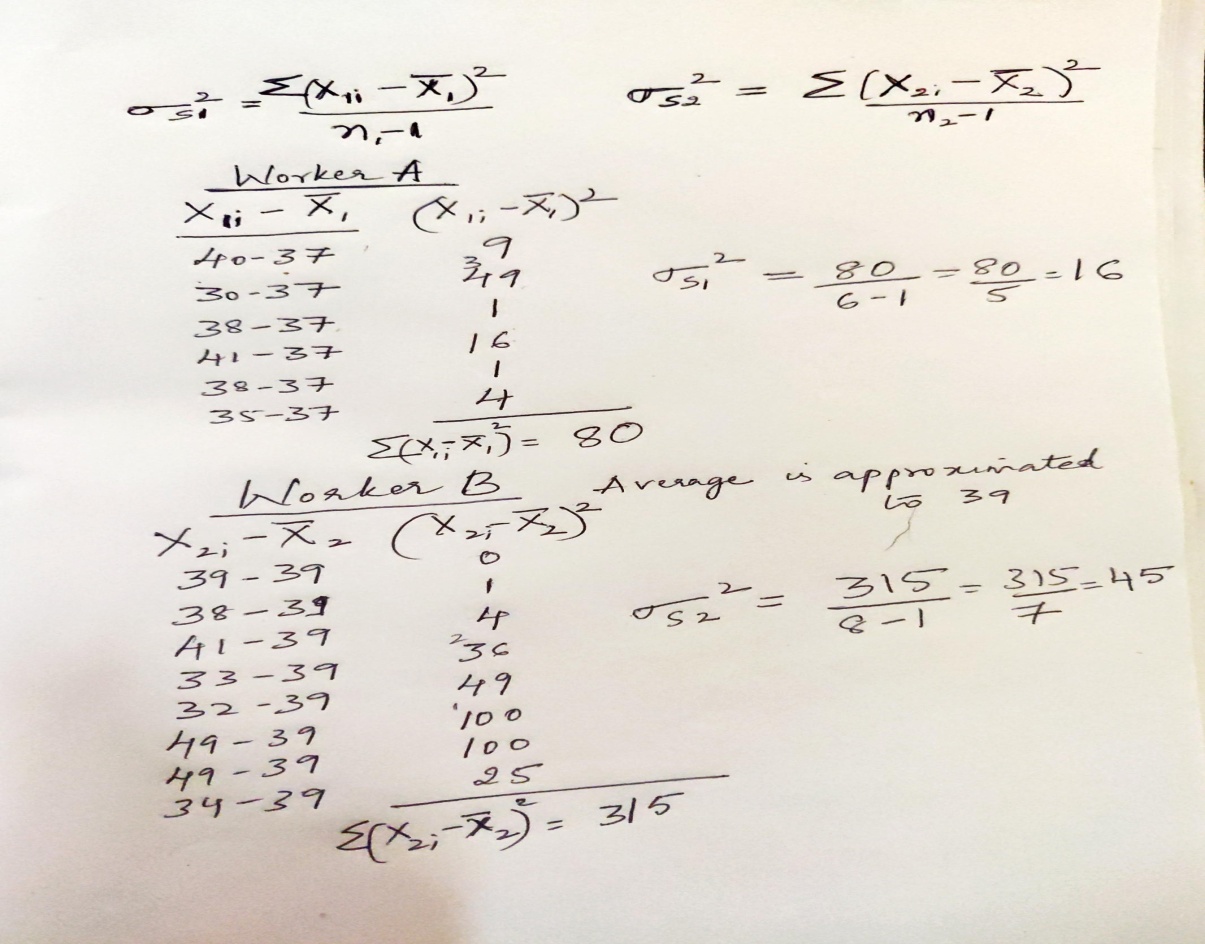
Worker A - 40, 30, 38, 41, 38, 35

Worker B – 39, 38, 41, 33, 32, 49, 49, 34

Should these results be accepted as evidence that B is the more stable worker? Use F test at 5% significance level.

Average/mean of worker A is

Average/mean of worker B is

****

The F statistic is estimated by

F =

The table/critical value of F for 7 and 5 df at 5% level of significance is 4.9. The estimated F value is less than the critical value. Hence the null hypothesis of no difference between sample variances is accepted. It can be stated that

**Chi-Square Test of Association/ Test of Independence**

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom for using this test. As a non-parametric test, chi-square can be used

1. as a test of goodness of fit and
2. as a test of independence.

*As a test of goodness of fit*, c2 test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of chi square is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the calculated value of chi square is greater than its table value, the fit is not considered to be a good one.

*As a test of independence*, c2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, chi square test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., gender and cardiac problem) are independent which means that there is no association between gender and suffering from cardiac problem.

On this basis we first calculate the expected frequencies and then work out the value of chi square statistic. If the calculated value of chi square is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated. But if the calculated value of 2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality.

.

The main reasons for the increasing use of non-parametric tests in research are:-

i. These statistical tests are distribution-free

ii. They are usually computationally easier to handle and understand

than parametric tests; and

iii. They can be used with type of measurements that prohibit the use of

parametric tests.

iv. This test can also be applied to a complex contingency table with several classes and as

such is a very useful test in research work.

In order to apply the chi-square test either as a test of goodness of fit or as a test the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ² is then calculated as follows:

where

*Oij* = observed frequency of the cell in *i*th row and *j*th column.

*Eij* = expected frequency of the cell in *i*th row and *j*th column.

In the case of a contingency table, the degrees of freedom df is worked out as follows:

d.f. = (*c* – 1) (*r* – 1)

where ‘*c*’ means the number of columns and ‘*r*’ means the number of rows.

Conditions for the Application of Test

The following conditions should be satisfied before c2 test can be applied:

(i) Observations recorded and used are collected on a random basis.

(ii) All the items in the sample must be independent.

(iii) No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.

(iv) The overall number of items must also be reasonably large. It should normally be at least

50, howsoever small the number of groups may be.

(v) The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known as linear constraints.

**Steps involved in Applying Chi-Square Test**

The chisquare statistic is estimated in the following way:

(i) The expected frequencies are calculated on the basis of given hypothesis or on the basis of null ypothesis. Usually in case of a 2 × 2 or any contingency table, the expected frequency for any given cell is worked out as:

Expected frequency of any cell =

(ii) Obtain the difference between observed and expected frequencies and find out the squares of such differences i.e., calculate (*Oij* – *Eij*)².

(iii) Divide the quantity (*Oij* – *Eij*)² obtained as stated above by the corresponding expected frequency to get (*Oij* – *Eij*)²/*Eij* and this should be done for all the cell frequencies or the group frequencies.

(iv) Find the summation of (*Oij* – *Eij*)²/*Eij* values or what we call

This is the required 2 value.

Eg: Two research workers classified some people in income groups on the basis of sampling studies.

Their results are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Investigators* | *Income groups* | | | Total |
|  | *Poor* | *Middle* | *Rich* |  |
| A | 160 | 30 | 10 | 200 |
| B | 140 | 120 | 40 | 300 |
| Total | 300 | 150 | 50 | 500 |

Let us take the hypothesis that the sampling techniques adopted by research workers are similar (i.e., there is no difference between the techniques adopted by research workers). This being so, the expectation of *A* investigator classifying the people in

(i) Poor income group = = 120

(ii) Middle income group = = 60

(iii) Rich income group = = 20

Similarly the expectation of *B* investigator classifying the people in

(i) Poor income group = = 180

(ii) Middle income group = = 90

(iii) Rich income group = = 30

With expected frequencies calculated as above, the chi square statistic can be estimated as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Groups | *Observed frequency*  *Oij* | *Expected frequency*  *Eij* | *(Oij – Eij)* | *(Oij – Eij)²* |  |
| Investigator A classification as |  |  |  |  |  |
| Poor | 160 | 120 | 40 | 1600 | 1600/120 =  13.33 |
| Middle Class | 30 | 60 | -30 | 900 | 900/60=  15 |
| Rich | 10 | 20 | -10 | 100 | 100/20=  5 |
| Investigator B classification as |  |  |  |  |  |
| Poor | 140 | 180 | -40 | 1600 | 1600/180=  8.88 |
| Middle Class | 120 | 90 | 30 | 900 | 900/90=  10 |
| Rich | 40 | 30 | 10 | 100 | 100/30=  3.33 |

**= 13.33 + 15 + 5 + 8.88 + 10 + 3.33= 55.54**

The df will (c-1) X ( r-1) = ( 3-1) X( 2-1) = 2 X1 = 2

The table value of for two degrees of freedom at 5 per cent level of significance is 5.991. The calculated value of is much higher than this table value which means that the calculated value cannot be said to have arisen just because of chance. It is significant. Hence, the hypothesis does not hold good. This means that the sampling techniques adopted by two investigators differ and are not similar.

**Module 4**

**Analysis of Variance and Factor Analysis**

Univariate data means data of a single variable. For a single variable data, the measures of central tendencies like mean median and mode can be estimated along with the measures of dispersion.

Bivariate data means data of two variables. Along with measures of central tendencies and dispersion for both the variables, both the variables can be analyzed with the help of correlation, simple regression and also coefficient of variation.

Multivariate data means the data of more than two variables. The popular techniques that are available for multivariate data analysis are multiple regression, one way ANOVA, multiple way ANOVA, factor analysis, principal component analysis etc.

**Analysis of Variance (ANOVA)**

If the data is univariate or bivariate, then one can employ t or z tests for testing the significance or differences in sample means. But if there is multivariate data that is data on more than two variables, then it would be difficult to employ t or z tests. Comparing means for equality or differences would be difficult as t or z test allow comparisons of only two samples at a time. The technique available for comparing the means of more than two samples at once is ANOVA.

ANOVA is essentially a procedure for testing the difference among different groups of data for homogeneity. The essence of ANOVA is that the total amount of variation in a set of data is broken down into two types, that amount which can be attributed to specific causes there may be variation between samples and also within sample items. ANOVA consists in splitting the variance for analytical purposes hence it is a method of analysing the variance to which a response is subject into it’s various components corresponding to various sources of variation.

The ANOVA technique is important in the context of all those situations where we want to compare more than two populations such as in comparing the yield of crop from several varieties of seeds, the gasoline mileage of four automobiles etc. ANOVA can investigate the differences among the means of all the populations simultaneously.

Through ANOVA technique we can investigate any number of factors which are hypothesised or set to influence the dependent variable. one can also investigate the differences amongst various categories within each of these factors which may have large number of possible values. if we take only one factor and investigate the differences amongst its various categories having numerous possible values, we are set to use one way ANOVA and in case we investigate two factors at the same time, then we use two way ANOVA. In a two or multiple way ANOVA [ MANOVA ], the interaction that is the interrelation between two or more independent variables or factors if any, affecting a dependent variable can as well be studied for better decisions.

Basic principle of ANOVA

The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples.

In terms of variation within the given population it is assumed that the values (Xij) differ from the mean of this population only because of random effects that is there are influences on (Xij) which are unexplainable, whereas in examining the differences between populations we assume that the difference between the mean of jth population and the grand mean (mean of all observations together) is attributable to what is called a specific factor or what is technically described as treatment effect.

The F statistic is estimated as follows

F =Estimate of population variance based on between samples variance/Estimate of population variance based on within samples variance

If F value equals or exceeds the critical F value we say that there are significant differences between sample means.

If there are k samples with n1, n2, n3…..nk observations, then F satatistic is calculated as follows

F = MSS between samples/ MSS within samples

MSS is Mean Sum of Squares

MSS between samples = SS between samples/k-1 where k-1 is the degrees of freedom

MSS within samples = SS within samples/n-k where n- k is the degrees of freedom

SS is Sum of Squares

SS between samples is calculated as follows

SS between samples = n1 (Ẋ1 -XG)² + n2 (Ẋ2 -XG)² + n3 (Ẋ3 -XG)²……. + nk (Ẋk -XG)²

where Ẋ1 , Ẋ2 , Ẋ3 ……Ẋk are means of k samples and XG is the Grand mean that is the mean of all samples together.

SS within samples is calculated as follows

SS within samples = Ʃ(X1i - Ẋ1)² + Ʃ(X2i - Ẋ2)² + Ʃ(X3i - Ẋ3)²………..+ Ʃ(Xki - Ẋk)²

Where X1i , X2i , X3i…….. Xki are individual observations from 1, 2, 3….k samples.

Total variance in the samples is given by

Total sum of squares (TSS) = SS Between + SS Within

Example:

Set up an ANOVA table for the following per acre production dataof three varieties of wheat each grown on four plots and state if the variety differences are significant.

|  |  |  |  |
| --- | --- | --- | --- |
| Plot of Land | Per Acre production of  wheat variety | | |
|  | A | B | C |
| 1 | 6 | 5 | 5 |
| 2 | 7 | 5 | 4 |
| 3 | 3 | 3 | 3 |
| 4 | 8 | 7 | 4 |

ẊA = 6+7+3+8 /4 = 24/4 = 6

ẊB = 5+5+3+7/4 =20/4 = 5 ẊG = 6+5+4/3 = 15/3 = 5

ẊC = 5+4+3+4/4 =16/4 = 4

SS between = 4(6-5)² + 4(5-5)² + 4(4-5)² = 4+0+4 = 8

MSS between = SS Between /k-1 = 8/3-1 =8/2 = 4

SS within = (6-6)² + (7-6)² + (3-6)² + (8-6)² + 0+1+9+4+

(5-5)² + (5-5)² + (3-5)² + (7-5)² + = 0+0+4+4+ = 24

(5-4)² + (4-4)² + (3-4)² + (4-4)² 1+0+1+0

MSS within = SS within/n-k = 24/12-3 = 24/9 =2.67

F = MSS between/ MSS within = 4/2.67 = 1.5

The Degrees of Freedom (df) of F Statistic is the respective df of MSS between and MSS within that is k-1 and n-k.

The critical value of F2,9 at 5% level of significance is 4.26 where 2 (3-1) and 9 (12-3) are df . Since the estimated value of F statistic 1.5 is less than critical value the H0 of no difference in sample means is accepted. The difference in wheat output due to varieties is insignificant and just a matter of chance.

One Way ANOVA Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source of Variation | Sum of Squares | Degrees of Freedom df | MSS | F-statistic | F-Critical Value |
| Between Samples | 8 | k-1 =2 | 8/2=4 | 4/2.67=1.5 | F2,9  4.26 |
| Within Samples | 24 | n-k = 9 | 24/9 = 2.67 |

**Factor Analysis**

Factor Analysis is -

* Data reduction tool
* Removes redundancy or duplication from a set of correlated variables
* Represents correlated variables with a smaller set of “derived” variables.
* Factors are formed that are relatively independent of one another.
* Two types of “variables”:
  + latent variables: factors
  + observed variables

*Some Applications of Factor Analysis*

1. Identification of Underlying Factors:

* + clusters variables into homogeneous sets
  + creates new variables (i.e. factors)
  + allows us to gain insight to categories

2. Screening of Variables:

* + identifies groupings to allow us to select one variable to represent many
  + useful in regression (recall multicollinearity)

3. Summary:

* + Allows us to describe many variables using a few factors

4. Clustering of objects:

* + Helps us to put objects (people) into categories depending on their factor scores

One of the ***primary goals of factor analysis*** is often to identify a measurement model for a latent variable.

This includes

* + identifying the items to include in the model
  + identifying how many ‘factors’ there are in the latent variable
  + identifying which items are “associated” with which factors

***Steps in Exploratory Factor Analysis***

1. Collect and explore data: choose relevant variables

(2) Determine the number of factors

(3) Estimate the model using predefined number of factors

(4) Rotate and interpret

(5) (a) Decide if changes need to be made (e.g. drop item(s), include item(s))

(b) repeat (3)-(4)

(6) Construct scales and use in further analysis

***Steps in Factor Analysis***

Factor analysis usually proceeds in four steps:

1st Step: the correlation matrix for all variables is computed

2nd Step: Factor extraction

3rd Step: Factor rotation

4th Step: Make final decisions about the number of underlying factors

* Data Matrix - Factor analysis is **totally dependent** on correlations between variables.
* Factor analysis summarizes correlation structure

***Assumptions*** underlying factor analysis include

* + The measured variables are linearly related to the factors + errors.
    - This assumption is likely to be violated if items limited response scales (two-point response scale like True/False, Right/Wrong items).
  + The data should have a bi-variate normal distribution for each pair of variables.
  + Observations are independent.
  + The factor analysis model assumes that variables are determined by common factors and unique factors. All unique factors are assumed to be **uncorrelated** with each other and with the common factors.

Choosing Number of Factors

* Intuitively: The number of uncorrelated constructs that are jointly measured by the X’s.
* Only useful if number of factors is less than number of X’s (recall “data reduction”).

Use “principal components” to help decide

* + type of factor analysis
  + number of factors is equivalent to number of variables
  + each factor is a weighted combination of the input variables:

**F1 = a11X1 + a12X2 + ….**

* + Recall: For a factor analysis, generally,

**X1 = a11F1 + a12F2 +...**

* To select how many factors to use, consider *eigenvalues* from a principal components analysis
* Two interpretations:
  + eigenvalue ≅ equivalent number of variables which the factor represents
  + eigenvalue ≅ amount of variance in the data described by the factor.
* Rules to go by:
  + number of eigenvalues > 1
  + scree plot
  + % variance explained
  + comprehensibility
* Note: sum of eigenvalues is equal to the number of items

(i) *Factor:* A factor is an underlying dimension that account for several observed variables. There can be one or more factors, depending upon the nature of the study and the number of variables involved in it.

(ii) *Factor-loadings:* Factor-loadings are those values which explain how closely the variables are related to each one of the factors discovered. They are also known as factor-variable correlations. In fact, factor-loadings work as key to understanding what the factors mean. It is the absolute size (rather than the signs, plus or minus) of the loadings that are important in the interpretation of a factor.

(iii) *Communality* (*h*2)*:* Communality, symbolized as *h*2, shows how much of each variable is accounted for by the underlying factor taken together. A high value of communality means that not much of the variable is left over after whatever the factors represent is taken into consideration. It isworked out in respect of each variable as under:

*h*2 of the *i*th variable = (*i*th factor loading of factor *A*)2

+ (*i*th factor loading of factor *B*)2 + …

(iv) *Eigen value* (*or latent root*)*:* When we take the sum of squared values of factor loadings relating to a factor, then such sum is referred to as Eigen Value or latent root. Eigen value indicates the relative importance of each factor in accounting for the particular set of variables being analysed.

(v) *Total sum of squares:* When eigen values of all factors are totalled, the resulting value is termed as the total sum of squares. This value, when divided by the number of variables (involved in a study), results in an index that shows how the particular solution accounts for what all the variables taken together represent. If the variables are all very different from each other, this index will be low. If they fall into one or more highly redundant groups, and if the extracted factors account for all the groups, the index will then approach unity.

(vi) *Rotation:* Rotation, in the context of factor analysis, is something like staining a microscope slide. Just as different stains on it reveal different structures in the tissue, different rotations reveal different structures in the data. Though different rotations give results that appear to be entirely different, but from a statistical point of view, all results are taken as equal, none superior or inferior to others. However, from the standpoint of making sense of the results of factor analysis, one must select the right rotation. If the factors are independent orthogonal rotation is done and if the factors are correlated, an oblique rotation is made. Communality for each variables will remain undisturbed regardless of rotation but the eigen values will change as result of rotation.

(vii) *Factor scores:* Factor score represents the degree to which each respondent gets high scores on the group of items that load high on each factor. Factor scores can help explain what the factors mean. With such scores, several other multivariate analyses can be performed.

**Module 5**

**Correlation and Regression Analysis**

***Correlation Analysis***

If the two variables data is given for all individuals we have a bivariate population. We come across relationship of two or more variables in Economics. Any economic theory deals with set of causes and effect. The degree of association between variables or among variables is called correlation. Demand for commodity is correlated with its price. The relationship is called correlation. Similarly, consumption depends on disposable income.

The Correlation indicates the degree of relationship. It does not indicate which the dependent and independent variable is.

If for every value of a variable X there is a corresponding value of another variable Y. There are n observations.

X Y

X1 Y1

X2 Y2

X3 Y3

**…. ….**

**…. ….**

**…. ….**

**…. ….**

**Xn Yn**

The above series are called Bivariate population. The set of pairs of observation (X1, Y1), (X2, Y2) ………….( **Xn,Yn)** constitute a bivariate population.

Correlation - Describes the relationship between two or more variables;

Describes the strength of the relationship in terms of a number from -1.0 to +1.0;and

Describes whether the direction of the relationship as positive or negative.

***Types of Correlations***

Variable X increases

Variable Y increases

Positive Correlation - Value ranging from .00 to 1.00

Example: the more you eat, the more weight you will gain

Variable X increases

Variable Y decreases

Negative Correlation - Value ranging from -1.00 to .00

Example: the older you are, the less flexible your body is

Variable X decreases

Variable Y increases

Negative Correlation - Value ranging from -1.00 to .00

Example: the less time you study, the more errors you will make

**Correlation does not indicate causation.**

**Correlation only establishes that a relationship exists; it reflects the amount of variability that is shared between two variables and what they have in common.**

Examples:

Amount of ice sold and number of bee stings.

A scattergram or scatter plot visually represents a correlation

The X axis is on the horizontal

The Y axis is on the vertical.



Let us consider the following demand and supply functions are:

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| **Demand** | **Price** | **Supply** | **Price** |
| X  16  14  12  10  8  6 | Y  3  4  5  6  7  8 | X  20  22  24  26  28  30 | Y  3  6  9  12  15  18 |

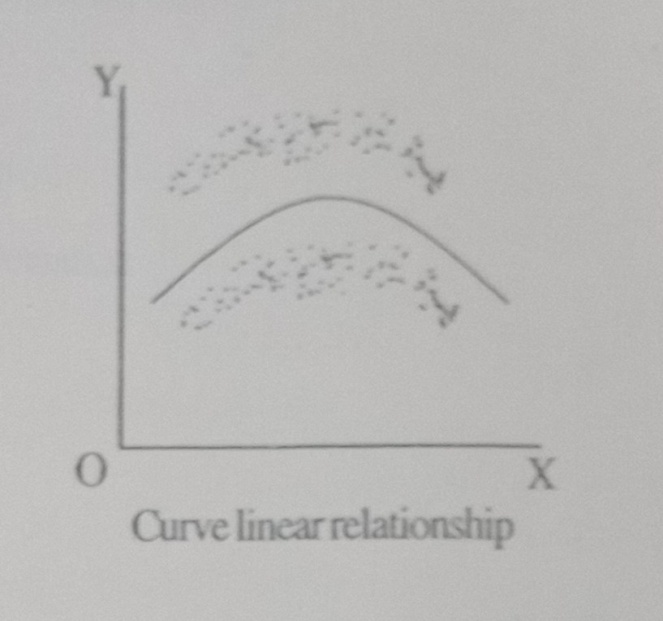
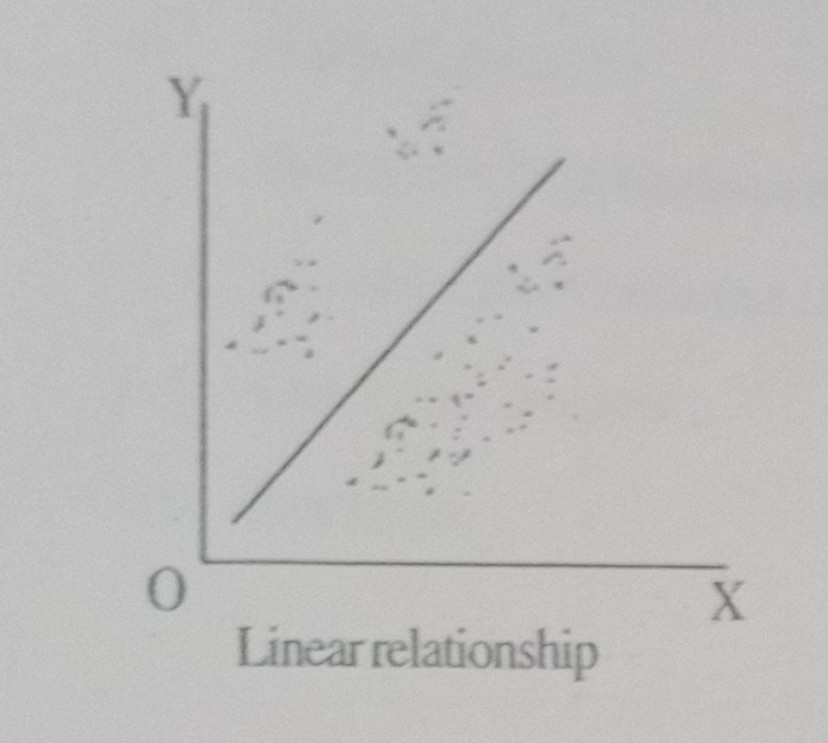
The above data can be presented in a scatter diagram. The purpose of the scatterplot is to visualize the relationship between the two variables represented by the horizontal and vertical axes. The demand has a linear inverse relationship between X and Y

The Supply price relation presents a case of linear direct relationship between X and Y.

Following diagram represents indefinite relation among X and Y variables

We can distinguish between linear and curve linear relationship of variable in correlation coefficient. If X and Y variables have been increasing or decreasing constantly then it is called as linear correlation. If X and Y variables are not increasing or decreasing proportionately it is called curvilinear relationship.

These are two types of relationship can be shown in a diagram



***Limitation:***

The graphical method is not reflecting the extent of relationship of variables. This is considered as crude method of assessing the relationship.

**Karl’s Pearson’s Coefficient of Correlation**

Karl Pearson was the one who found the method to estimate the linear relationship between variables. The coefficient is hence named Karl’s Pearson’s Coefficient of Correlation. Karl Pearson’s coefficient of correlation is also known as the product moment correlation coefficient.

r (small) is denote as the coefficient of correlation. The value of r lies between -1 and +1.

Formula to estimate the co-efficient of correlation is given by:

This can be written as:

The Correlation co-efficient is obtained by using any of the above formulas.

**Properties of Correlation Coefficient** Let r denote the correlation coefficient between two variables. r≥ is interpreted using the following properties: 1. The value of r ranges from – 1.0 to 0.0 or from 0.0 to 1.0

2. A value of r = 1.0 indicates that there exists perfect positive correlation between the two variables.

3. A value of r = - 1.0 indicates that there exists perfect negative correlation between the two variables.

4. A value r = 0.0 indicates zero correlation i.e., it shows that there is no correlation at all between the two variables.

5. A positive value of r shows a positive correlation between the two variables.

6. A negative value of r shows a negative correlation between the two variables.

7. A value of r = 0.9 and above indicates a very high degree of positive correlation between the two variables.

8. A value of - 0.9 ≥ r > - 1.0 shows a very high degree of negative correlation between the two variables. 9. For a reasonably high degree of positive correlation, we require r to be from 0.75 to 1.0.

10. A value of r from 0.6 to 0.75 may be taken as a moderate degree of positive correlation.

***Example***:

Calculate Karl Pearson’s correlation coefficient for the following data:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 84 | 88 | 93 | 103 | 78 | 80 | 110 | 92 |
| Y | 87 | 93 | 94 | 111 | 98 | 78 | 99 | 84 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 84  88  93  103  78  80  110  92 | 87  93  94  111  98  78  99  84 | -7  -3  2  12  -13  -11  19  1 | -6  0  1  18  5  -15  6  -9 | 49  9  4  144  169  121  361  1 | 36  0  1  324  25  225  36  81 | 42  0  2  216  -65  165  144  9 |
| **728** | **744** | **0** | **0** |  |  |  |
|  |  | **(Note: Minus values should be detected from plus value get )** | | | | |

**Probable Error of r**

If r is the co-efficient of correlation between a pair of values of X and Y or Y and X, then the probable error is given by:

Here the square of the correlation co-efficient is called the co-efficient of determination. i.e.,

If

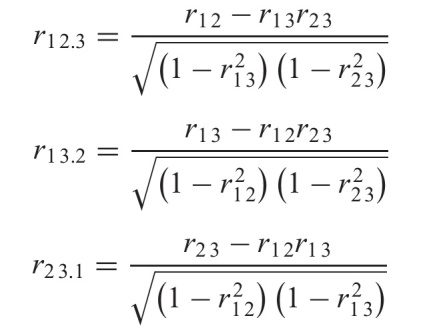
PE (r) is the limit such that probability of r lying between r and p is exactly ½ provided r is normally distributed. P is never negative.

**Multiple and Partial Correlations**

**Partial correlation** measures separately the relationship between two variables in such a way that the effects of other related variables are eliminated. In other words, in partial correlation analysis, we aim at measuring the relation between a dependent variable and a particular independent variable by holding all other variables constant. Thus, each partial coefficient of correlation measures the effect of its independent variable on the dependent variable. To obtain it, it is first necessary to compute the simple coefficients of correlation between each set of pairs of variables.

For the three variable regressions model three correlation coefficients, r12, r23 and r13 are computed [Y=1, X2=2, X3=3]. These correlation coefficients are called gross or simple correlation coefficients. The simple correlation coefficient cannot measure the “true” degree of linear association between Y and X2 when a third variable X3 is associated with both of them. Therefore, what is need is a correlation coefficient that is independent of the influence, if any, of X3 on X2 and Y.

Such a correlation coefficient can be obtained and is known appropriately as the partial correlation coefficient.



= Partial Correlation Coefficient (P.C.C) between Y and X2 keeping X3 constant

=P.C.C between Y and X3 holding X2 constant.

= P.C.C between X2 and X3 keeping Y as constant

These are also called as first order correlation coefficients. Order indicates the number of secondary subscripts. Thus, r12.34 is the correlation coefficient of order two, r12.345 is the correlation coefficient of order three etc.

Interpretation of Simple and Partial correlation coefficients

1. Even if r12=0,r12.3 will not be zero unless r13 or r23 or both are zero.
2. If r12=0 and r13 and r23 are non zero and are of the same sign, then r12.3 will be negative, whereas if they are of opposite signs, it will be positive.
3. The terms r12.3 and r12need not have same sign.
4. In the two variable model r2 lies between 0 and 1. The same property holds true of the squared partial correlation coefficient.

0 ≤r212+r213+r223-2r12r13r23 ≤ 1

1. Suppose r13=r23=0 this does not implies that r12=0. If Y and X3 and X2 and X3 are uncorrelated, it does not mean that Y and X2 are uncorrelated.

**Multiple Correlation Coefficients**

Multiple correlations involve the study of the joint effect of a group of variables upon a variable not included in that group. The multiple correlation is indicated as r1.23for the above example of three variables. If there are 1,2,3….n variables, then it is r1.23……n. r1.23 is calculated as below

r2.13 is calculated as

R3.12 is calculated as

**Spearman’s Rank Correlation Coefficient**

The Karl Pearson’s coefficient of correlation estimates the linear association between variables which are measurable. This can’t be used for all types of problems. Let us consider the case the athletic ability, debating capacity etc. These are non- measurable characteristics. Before estimating, we have to rank them. By adopting ranks, Charles Spearman developed a formula to estimate correlation. This is called as Spearman’s Rank Correlation coefficient.

Following is the procedure for estimating Spearman’s rank correlation coefficient.

* If there are two variables, say X and Y. Each has 10 observations.
* They have to be ranked from lowest to highest or vice-versa for both series.
* After that the difference of ranks has to be taken respectively.
* Then the differences have to be squared to get d2. After finding d2, we can estimate,

where d is the differences of ranking in two series, d2 is the square of d and n is the number of observations.

**It is a non-parametric measure of correlation.**

**This procedure makes use of the two sets of ranks that may be assigned to the sample values of X and Y.**

**Spearman Rank correlation coefficient could be computed in the following cases:**

**Both variables are quantitative.**

**Both variables are qualitative ordinal.**

**One variable is quantitative and the other is qualitative ordinal.**

***Example***: The following are profit (X) and sale (Y) of 10 Companies during 1991-92. Calculate the Spearman’s rank correlation coefficient.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X(Profit)** | | 1 | 2 | | 3 | | 4 | | 5 | | 6 | 7 | | 8 | 9 | 10 |
| **Y(Sale)** | | 10 | 7 | | 2 | | 6 | | 4 | | 8 | 3 | | 1 | 9 | 5 |
| **X** | **Y** | | | **Rx** | | **Ry** | | **d** | | **d2** | | |
| 1 | 10 | | | 1 | | 10 | | -9 | | 81 | | |
| 2 | 7 | | | 2 | | 7 | | -5 | | 25 | | |
| 3 | 2 | | | 3 | | 2 | | 1 | | 1 | | |
| 4 | 6 | | | 4 | | 6 | | -2 | | 4 | | |
| 5 | 4 | | | 5 | | 4 | | 1 | | 1 | | |
| 6 | 8 | | | 6 | | 8 | | -2 | | 4 | | |
| 7 | 3 | | | 7 | | 3 | | 4 | | 16 | | |
| 8 | 1 | | | 8 | | 1 | | 7 | | 49 | | |
| 9 | 9 | | | 9 | | 9 | | 0 | | 0 | | |
| 10 | 5 | | | 10 | | 5 | | 5 | | 25 | | |
|  | | | | | | | | | | **206** | | |

**Example** : The following are the marks obtained in Maths and Statistics. Calculate Spearman Rank Correlation efficient.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Marks in Maths X** | **Marks in Statistics Y** | **Rx** | **Ry** | **d** | **d2** |
| 45  45  60  38  50 | 60  61  58  48  48 | 3.5  3.5  1  5  2 | 2  1  3  4.5  4.5 | 1.5  2.5  2  0.5  2.5 | 2.25  6.25  4.00  2.25  6.25 |

***Note***: In case in the series there are same marks, we always make fraction as made in first two of Rx and last two Ry.

There is a tie between third and fourth rank in Rx. Hence a rank of 3+4/2 = 3.5 is assigned to both of them. Next observation is given the fifth rank.

Similarly in Ry, there is tie forfourth and fifth rank. Hence a rank of 4.5 is assigned to those two observations.

An elaborate formula which can be used in case of tied ranks is:

where is to be added to for each group of tied ranks. m stands for the number of items whose ranks are common. If there are more than one such group of individuals with common rank, the above value that is , is added as many times as the number of such groups.

**Regression Analysis**

The co-efficient of correlation indicates whether the variables are linearly related and if so, how strong the relationship is. Moreover in correlation analysis, one cannot distinguish between dependent and independent variables. In economic theories, we come across some variables which are causes and some variables which are affected from cause variables. The cause and effect relationship can be analysed with the help of regression.

The main objectives of regression analysis are the formulation and determination of the mathematical model or form of the relationship between variables and the use of the model for prediction purposes. Regression methods consist of determining the exact mathematical form of relationship.

The word regression, means reversion, was first used by Galton to whom regression analysis is due. His study was on heights of fathers and sons; he showed that the heights of son reverted towards the average rather than to the extreme values. In economics, we come across that the consumption has a positive relationship with disposable income. Disposable income is responsible for changes in consumption.

The statistical technique that is designed to obtain the linear straight line for the observed data is called regression. Regression is the estimation of unknown values or the prediction of one variable from the known values of other variable/variables. Simple linear regression involves a relationship between two variables only.

**Distinction between Correlation and Regression**

The correlation co-efficient merely tells of how strong is the correlation between two variables. The correlation co-efficient value is always between -1 and +1. Correlation does not explain the functional relationship. We cannot predict the value of variables from correlation.

The regression is a method which considers the functional relationship between variables.. It distinguishes between independent and dependent variables. It is used on conditionality. For the given values of independent variable, dependent variable values are estimated. Regression analysis attempts to measure the degree of relationship among the variables.

**Simple Regression Model**

If we assume linear regression between two variables; it means that if the set of observations on those variables are plotted on a graph, a straight line can be approximately drawn through a chart. For the purpose of estimation, we require a functional relationship between two variables, X and Y. The form of functions is Y=f(x) and X=f(y). Such a mathematical model of relationship is called regression. Y= f(x) is called the regression of Y on X and X = f(y) is called the regression of X and Y.

If Y= a1 + b1 X then

If X= a2 + b2Y then where and are arithmetic means of x and y. So by subtractions, we get and

b1 in byx and b2 in bxy are called linear regression co-efficient or simple regression co-efficient.

In the linear relationship between two variables Y = a + bx,

Y is dependent variable and X is independent variable; b is the slope or co-efficient of X; and a is the intercept. The b indicates the type of relationship and degree of relationship between two variables.

* **A regression model** can be **Deterministic Model** which is an equation or set of equations that allow us to ***fully determine*** the value of the dependent variable from the values of the independent variables. Eg: **Area of a circle: A =** π**\*r2**
* **Probabilistic Regression Model:** a method used to capture the ***randomness*** that is part of a real-life process.

Eg: y = 25,000 + 75x + ε here ԑ is the error term representing randomness.

**Regression Line: Line of Best Fit**

What is meant by the line of Best Fit?

We should determine regression coefficients in such a way that the points lies as close to the line as possible. This is called as the best fit. In other words, regression line minimizes the sum of the squared vertical deviations (ε) of each point from the regression line. By using the least squares method (a procedure that minimizes the vertical deviations of plotted points surrounding a straight line) one can construct a best fitting straight line to the scatter diagram points and then formulate a regression equation.

**Ordinary Least Squares (OLS) Method**

In the regression model,

****





The objective of the OLS method is to determine the slope and intercept that minimize the sum of the squared errors.



The OLS estimators are given as follows:





***Example:***

The regression of X (price) on Y (demand) has been determined for the data given below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Price(x)** | **Demand(y)** | **X2** | **Y2** | **XY** |
| 4  6  5  6  8  7 | 10  8  5  4  2  1 | 16  36  25  36  64  81 | 100  64  25  16  4  1 | 40  48  25  28  16  7 |
| **36** | **30** | **258** | **210** | **164** |

n = 6

The regression equation is: Y = 15.77 – 1.65 X

**Multiple Regression Analysis**

In multiple regression analysis the effect of more than one independent variable is studied on a dependent variable. In simple regression analysis we examine the relationship of one dependent variable and one independent variable.

**Three Variable Regression Model**

The general form a multiple regression model can be expressed as

Yi = ß1 + ß2X2i + ... + ßkXki + ui

This model has k unknown parameters (k-1 slope coefficients and one intercept term). Thus we may refer to this as the k-variate regression model.

This simplest multiple regression model incorporates all of the concepts underlying the more general k-variate model.

In multiple regression we distinguish between joint and partial (individual) effects of the explanatory variables. Each partial or individual regression coefficients in a multiple regression model represents the effect of a one unit change in the corresponding X variable on the dependent variable, holding all other independent variables constant.

These are called partial regression coefficients because they are similar to partial derivatives. For example, in the tri-variate model,

ß2 = ∂Y/∂X2 and ß3 = ∂Y/∂X3

Thus ß2 measures change in Y per unit change in X2, holding X3 constant. In other words, ß2 measures the change in the mean values of Y, per unit change in X2, holding X3 constant or the ‘direct’ or ‘net’ effect of a unit change in X2 on the mean value of Y

ß3 measures change in Y per unit change in X3, holding X2 constant. It means ß3 measures the change in the mean values of Y, per unit change in X3, holding X2 constant or the ‘direct’ or ‘net’ effect of a unit change in X2 on the mean value of Y.

Holding constant indicates the assessment of the true contribution of X2 to the change in Y, we control the influence of X3 and vice versa.

The OLS estimators of the multiple regression model

Yi=β1+β2X2i+β3X3i+ui are given by

=

**The Assumptions Underlying the Method of Least Squares**

In regression analysis our objective is not only to obtain βˆ1 and βˆ2 but also to draw inferences about the true β1 and β2. For example, we would like to know how close βˆ1 and βˆ2 are to their counterparts in the population

The Gaussian, standard, or classical linear regression model (CLRM), makes 10 assumptions.

ASSUMPTION 1

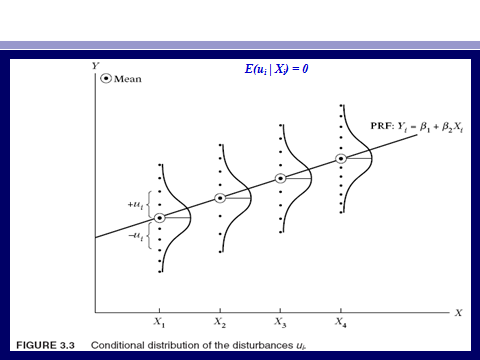
Linear Regression Model: The regression model is linear in the parameters, though it may or may not be linear in the variables.

That is the regression model is Yi = β1 + β2 Xi + ui

ASSUMPTION 2 : Fixed X Values or X Values Independent of the Error Term:

Values taken by the regressor X may be considered fixed in repeated samples. This means that our regression analysis is *conditional regression analysis*, that is, conditional on the given values of the regressor (s) *X.* Keeping the value of the independent variable *X fixed,* we draw at random a value for Y and *we can repeat this* process for all the *X values.*

ASSUMPTION 3 Zero Mean Value of Disturbance ui:

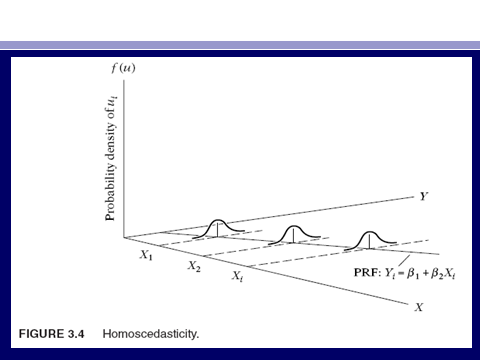
Given the value of Xi, the mean, or expected, value of the random disturbance term ui is zero. Symbolically, we have E(ui|Xi) = 0. Each Y population corresponding to a given X is distributed around its mean value with some Y values above the mean and some below it. The mean value of these deviations corresponding to any given X should be zero which can be observed for the following figure.

ASSUMPTION 4 Homoscedasticity or Constant Variance of ui:

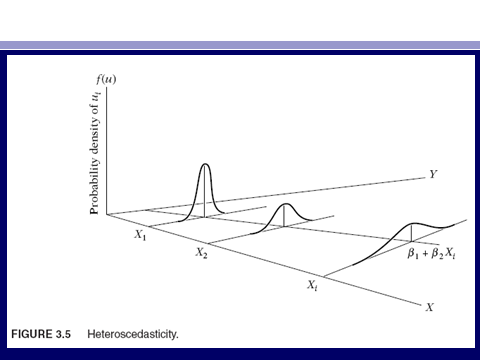
The variance of the error, or disturbance, term is the same regardless of the value of X. Symbolically, var (ui) = E[ui − E(ui|Xi)]2 = E(u2 i |Xi) = σ2

where var stands for variance homo means equal and scedasticity means spread, that is equal spread. Stated differently, it means that the Y populations corresponding to various X values have the same variance.

Put simply, the variation around the regression line (which is the line of average relationship between Y and X) is the same across the X values; it neither increases or decreases as X varies. Which can be observed from the following figure:



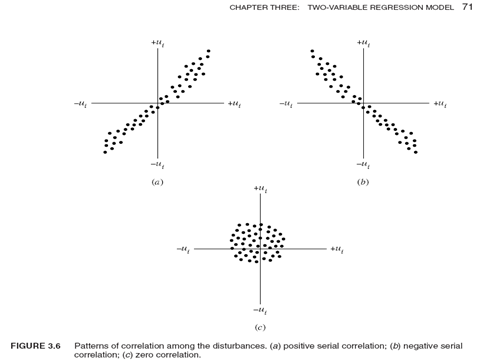
where the conditional variance of the Y population varies with X. This situation is known as heteroscedasticity, or unequal spread, or variance. Symbolically, this situation can be written as var (ui | Xi) = σ2i and shown in the figure given below:

****

ASSUMPTION 5 No Autocorrelation between the Disturbances:

Given any two X values, Xi and Xj (i ≠ j), the correlation between any two ui and uj(i ≠ j) is zero. In short, the observations are sampled independently. Symbolically, cov(ui, uj|Xi, Xj) = 0 cov(ui, uj) = 0, where i and j are two different observations and where cov means covariance.

The disturbances ui and uj are uncorrelated, i.e., no serial correlation or no autocorrelation. This means that, given Xi , the deviations of any two Y values from their mean value do not exhibit patterns. In the following figure, (a) and (b) represent a systematic pattern with positive and negative autocorrelation respectively whereas (c) shows no autocorrelation.\



ASSUMPTION 6 : Zero correlation between Xi and Ui.

The disturbance u and explanatory variable X are uncorrelated. The PRF assumes that X and u (which may represent the influence of all the omitted variables) have separate (and additive) influence on Y. But if X and u are correlated, it is not possible to assess their individual effects on Y. Thus, if X and u are positively correlated, X increases when u increases and it decreases when u decreases. Similarly, if X and u are negatively correlated, X increases when u decreases and it decreases when u increases. In either case, it is difficult to isolate the influence of X and u on Y.

ASSUMPTION 7 The Number of observations n must be greater than the number of parameters to be estimated: Alternatively, the number of observations must be greater than the number of explanatory variables. For example I f we have only single observation of Y and X, there is no way to estimate the two unknowns, β1 and β2. We need at least two pairs of observations to estimate the two unknowns.

ASSUMPTION 8: Variability in X values. The X values in a given sample must not all be the same and technically the variance of X should be finite positive value. If all the X values are identical, then Xi = X¯ and the denominator of that equation will be zero, making it impossible to estimate β2 and therefore β1.

ASSUMPTION 9 Regression model is correctly Specified

An econometric investigation begins with the specification of the econometric model underlying the phenomenon of interest. Some important questions that arise in the specification of the model include the following:

(1) What variables should be included in the model?

(2) What is the functional form of the model? Is it linear in the parameters, the variables, or both?

(3) What are the probabilistic assumptions made about the Yi , the Xi, and the ui entering the model?

A model should include all relevant variable, correct functional form and appropriate assumption related to Y, X and u.

ASSUMPTION 10: There is no perfect multicollinearity.

Perfect multicollinearity means existence of an exact linear relationship between two or more independent variables. When there is perfect multicollinearity, regression coefficients cannot be estimated as they will be of the indeterminate form, 0/0. Further, standard errors of the estimated coefficients tends to infinity as they will equal σ2/0, implying complete lack of precision. Hence regression analysis assumes no perfect multicollinearity among explanatory variables.

ASSUMPTION 11:The population error term, ui, follows the normal distribution

Any regression model that satisfies assumptions 1-10 is known as a Classical Linear Regression Model (CLRM), which is suitable for estimation alone. Any regression model that satisfies assumptions 1-11 is known as the Classical Normal Linear Regression Model (CNLR) , suitable not only for estimation but also for testing hypotheses.

**Module 6**

**INTERPRETATION AND REPORT WRITING**

After collecting and analyzing the data, the researcher has to accomplish the task of drawing inferences followed by report writing. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. It is only through interpretation that the researcher can expose relations and processes that underlie his findings. In case of hypotheses testing studies, if hypotheses are tested and upheld several times, the researcher may arrive at generalizations. But in case the researcher had no hypothesis to start with, he would try to explain his findings on the basis of some theory. This may at times result in new questions, leading to further researches. All this analytical information and consequential inference(s) may well be communicated, preferably through research report, to the consumers of research results who may be either an individual or a group of individuals or some public/private organisation.

**Meaning of Interpretation**

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz., (i) the effort to establish continuity in research through linking the results of a given study with those of another, and (ii) the establishment of some explanatory concepts. “In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses.”1 Thus, interpretation is the device through which the factors that seem to explain what has been observed by researcher in the course of the study can be better understood and it also provides a theoretical conception which can serve as a guide for further researches.

**Technique of Interpretation**

The task of interpretation is not an easy job, rather it requires a great skill and dexterity on the part of researcher. Interpretation is an art that one learns through practice and experience. The researcher may, at times, seek the guidance from experts for accomplishing the task of interpretation.

The technique of interpretation often involves the following steps:

(i) Researcher must give reasonable explanations of the relations which he has found and he must interpret the lines of relationship in terms of the underlying processes and must try to find out the thread of uniformity that lies under the surface layer of his diversified research findings. In fact, this is the technique of how generalization should be done and concepts be formulated.

(ii) Extraneous information, if collected during the study, must be considered while interpreting the final results of research study, for it may prove to be a key factor in understanding the problem under consideration.

(iii) It is advisable, before embarking upon final interpretation, to consult someone having insight into the study and who is frank and honest and will not hesitate to point out omissions and errors in logical argumentation. Such a consultation will result in correct interpretation and, thus, will enhance the utility of research results.

(iv)Researcher must accomplish the task of interpretation only after considering all relevant factors affecting the problem to avoid false generalization. He must be in no hurry while interpreting results, for quite often the conclusions, which appear to be all right at the beginning, may not at all be accurate.

**Importance of Report Writing**

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others.

The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process. But the general opinion is in favour of treating the presentation of research results or the writing of report as part and parcel of the research project. Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research. This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

**Types of Reports**

Broadly speaking reporting can be done in two ways :

a) **Oral or Verbal Report :** reporting verbally in person, for example; Presenting the findings in a conference or seminar or reporting orally to the superiors.

b) **Written Report :** Written reports are more formal, authentic and popular. Written reports can be presented in different ways as follows.

i) Sentence form reports : Communicating in sentence form

ii) Tabular reports: Communicating through figures in tables

iii) Graphic reports: Communicating through graphs and diagrams

iv) Combined reports: Communicating using all the three of the above. Generally, this is the most popular Research reports vary greatly in length and type. In each individual case, both the length and the form are largely dictated by the purpose of the study and problems at hand. For example, business organizations generally prefer reports in letter form, that too short in length. Banks, insurance and other financial institutions generally prefer figure form in tables. The reports prepared by government bureaus, enquiry commissions etc., are generally very comprehensive on the issues involved. Similarly research theses/dissertations usually prepared by students for Ph.D. degree are also elaborate and methodical.

It is, thus, clear that the results of a research enquiry can be presented in a number of ways. They may be termed as a technical report, a popular report, an article, or a monograph.

1) **Technical Report:**

A technical report is used whenever a full written report (ex: Ph.D. thesis) of the study is required either for evaluation or for record keeping or for public dissemination. The main emphasis in a technical report is on :

a) The methodology employed.

b) The objectives of the study.

c) The assumptions made / hypotheses formulated in the course of the study.

d) How and from what sources the data are collected and how have the data been analyzed.

e) The detailed presentation of the findings with evidence, and their limitations.

2) **Popular Report:**

A popular report is one which gives emphasis on simplicity and attractiveness. Its aim is to make the general public understand the findings and implications. Generally, it is simple. Simplicity is sought to be achieved through clear language and minimization of technical details. Attention of the readers is sought to be achieved through attractive layout, liberal use of graphs, charts, diagrams and pictures. In a popular report emphasis is given on practical aspects and policy implications.

3) **Research Article:**

Sometimes the findings of a research study can be published in the form of a short paper called an article. This is one form of dissemination. The research papers are generally prepared either to present in seminars and conferences or to publish in research journals. Since one of the objectives of doing research is to make a positive contribution to knowledge, in the field, publication (publicity) of the work serves the purpose.

4) **Monograph:** A monograph is a treatise or a long essay on a single subject. For the sake of convenience, reports may also be classified either on the basis of approach or on the basis of the nature of presentation such as:

i) Journalistic Report

ii) Business Report

iii) Project Report

iv) Dissertation

v) Enquiry Report (Commission Report), and

Reports prepared by journalists for publication in the media may be **journalistic reports**. These reports have news and information value. A **business report** may be defined as report for business communication from one departmentalhead to another, one functional area to another, or even from top to bottom inthe organizational structure on any specific aspect of business activity. Theseare observational reports which facilitate business decisions.

A **project report** is the report on a project undertaken by an individual or a group of individuals relating to any functional area or any segment of a functional area or any aspect of business, industry or society.

A **dissertation**, on the other hand, is a detailed discourse or report on the subject of study. Dissertations are generally used as documents to be submitted for the acquisition of higher research degrees from a university or an academic institution. The thesis is an example in point. An **enquiry report** or a commission of enquiry report is a detailed report prepared by a commission appointed for the specific purpose of conducting a detailed study of any matter of dispute or of a subject requiring greater insight. These reports facilitate action, since they contain expert opinions.

**Stages in Preparation of A Report**

Research reports are the product of slow and painstaking and accurate work. Therefore, the preparation of the report may be viewed in the following major stages.

1) The logical understanding and analysis of the subject matter.

2) Planning/designing the final outline of the report.

3) Write up/preparation of rough draft.

4) Polishing/finalization of the Report.

**1)** **Logical Understanding of the Subject Matter:** It is the first stage which is primarily concerned with the development of a subject. There are two ways to develop a subject viz. a. logically and b. chronologically. The logical development is done on the basis of mental connections and associations between one aspect and another by means of logical analysis. Logical treatment often consists of developing material from the simple to the most complex. Chronological development is based on a connection or sequence in time or happening of the events. The directions for doing something usually follow the chronological order.

**2)** **Designing the Final Outline of the Report:** It is the second stage in writing the report. Having understood the subject matter, the next stage is structuring the report and ordering the parts and sketching them. This stage can also be called as planning and organization stage. Ideas may pass through the author’s mind. Unless he first makes his plan/sketch/design he will be unable to achieve a harmonious succession and will not even know where to begin and how to end. Better communication of research results is partly a matter of language but mostly a matter of planning and organizing the report.

**3) Preparation of the Rough Draft:** The third stage is the write up/drafting of the report. This is the most crucial stage to the researcher, as he/she now sits to write down what he/she has done in his/her research study and what and how he/she wants to communicate the same. Here the clarity in communicating/reporting is influenced by some factors such as who the readers are, how technical the problem is, the researcher’s hold over the facts and techniques, the researcher’s command over language (his communication skills), the data and completeness of his notes and documentation and the availability of analyzed results. Depending on the above factors some authors may be able to write the report with one or two drafts. Some people who have less command over language, no clarity about the problem and subject matter may take more time for drafting the report and have to prepare more drafts (first draft, second draft, third draft, fourth draft etc.,)

**4) Finalization of the Report:** This is the last stage, perhaps the most difficult stage of all formal writing. It is easy to build the structure, but it takes more time for polishing and giving finishing touches. Take for example the construction of a house. Up to roofing (structure) stage the work is very quick but by the time the building is ready, it takes up a lot of time. The rough draft (whether it is second draft or ‘n’ th draft ) has to be rewritten, polished in terms of requirements. The careful revision of the rough draft makes the difference between a mediocre and a good piece of writing. While polishing and finalizing one should check the report for its weaknesses in logical development of the subject and presentation cohesion. He/she should also check the mechanics of writing — language, usage, grammar, spelling and punctuation.

**Structure of the Research Report**

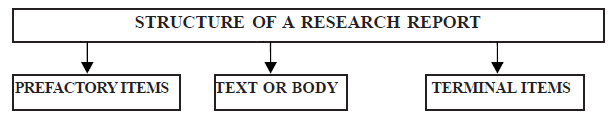
Under this head, the format/outline/sketch of a comprehensive technical report or research report is discussed below. A technical report has a number of clearly defined sections. The headings of the sections and their order may differ from one situation to another. The contents of a report can broadly be divided into three parts as :

1) The front matter or prefactory items.

2) The body or text of the report.

3) The back matter or terminal items.

The following chart summarizes the broad sequence of the contents of a research report.



1. Blank sheet Chapter 1: Introduction 1. Appendix, if any

2. Title page Chapter 2 to n: Presentation & 2. Glossary, if any

3. Approval sheet (if any) Description of Evidence 3. Bibliography

4. Researcher’s declaration 4. Index

5. Dedication (if any) Chapter n+1: Summary, 5. Blank sheet

6. Preface and/or Conclusions and

Acknowledgements Recommendations

7. Table of contents

8. List of tables

9. List of graphs/charts/

figures

10. List of cases, if any

11. Abstract or high lights (optional)

Let us discuss these items one by one in detail.

**Prefactory Items**

The various preliminaries to be included in the front pages of the report are briefly narrated hereunder:

1) **Title Page:** The first page of the report is the title page. The title page should carry a concise and adequately descriptive title of the research study, the name of the author, the name of the institution to whom it is submitted, the date of presentation.

2) **Approval Sheet:** If a certificate of approval is required either from the research supervisor or from the institution which provided the research facilities, it must be given.

3) **Researcher’s Declaration:** Generally the researcher has to declare/certify that it is his/her bonafide and original work done by him/her.

4) **Dedication:** If the author wants to dedicate the work to whom so ever he/she likes, he/she may do so.

5) **Preface or Acknowledgements:** A preface includes the background and reasons for the study. This is an appropriate place for him/her to make acknowledgements also. But if the researcher has opted to discuss the significance, reasons of the study elsewhere in the report he/she may not write ‘preface’. But he/she may use the page for only acknowledgements. In acknowledgements the researcher acknowledges the assistance and support received from individuals and organizations in conducting the research. It is intended to express his/her gratitude to them.

6) **Table of Contents:** A table of contents gives an outline of the contents of the report. It contains a list of the chapters and their titles with page numbers. It facilitates easy location of topics in the report. The chapter headings may be typed with capital letters.

7) **List of Tables:** The researcher must have collected lot of data and analyzed the same and presented in the form of tables. These tables may be listed chapter wise and the list be presented with page numbers for easy location and reference.

8) **List of Graphs/Charts/Figures:** If there are many graphs and charts they should also be listed with page numbers, after the list of tables separately.

9) **List of Cases/Exhibits:** If there are many cases/exhibits they should also be listed.

10) **Abstract:** An abstract is a synopsis. It should be as brief as possible and run about one or two pages. It is placed at the prefactory part of the report so that a reader can get a quick over view of the report. It contains a brief and precise statement of the purpose and a bare summary of the findings or the results of the study.

**The Text/Body of the Report**

After the preliminary items, the body of the report is presented. It is the major and main part of the report. It consists of the text and context chapters of the study. Normally the body may be divided into 3 (three) parts.

i) The introduction

ii) The description and discussion of evidence and findings

iii) The summary, conclusions and recommendations

i) **Introduction**

Generally this is the first chapter in the body of the report. It is devoted introducing the theoretical background of the problem and the methodology adopted for attacking the problem.

It may consist of the following aspects:

– Significance and justification of the topic.

– Theoretical background of the topic.

– Statement of the problem.

– Review of literature.

– Objectives of the study.

– Hypotheses to be tested.

– Definition of special terms, concepts and units of study.

– Scope of the study – geographical scope i.e. area/places to be covered, content, scope i.e., aspects to be included/excluded.

– Period of study i.e., reference period.

– Sources of data i.e., primary or secondary or both.

– Methods of data collection i.e., sample or census.

– Sampling design.

– Data collection instruments.

– Field work.

– Data processing and analysis plan.

– Limitations of the study, if any.

– An over view of the report i.e., chapter plan.

ii) **Description and Discussion of Evidence**

This is the major and main part of the report. It is divided into several chapters depending upon the number of objectives of the study, each being devoted to presenting the results pertaining to some aspect. The chapters should be well balanced, mutually related and arranged in logical sequence. The results should be reported as accurately and completely as possible explaining as to their bearing on the research questions and hypotheses. Each chapter should be given an appropriate heading. Depending upon the need, a chapter may also be divided into sections. The entire verbal presentation should run in an independent stream and must be written according to best composition rules. Each chapter should end with a summary and lead into the next chapter with a smooth transition sentence.

While dealing with the subject matter of text the following aspects should be taken care of.

They are :

1) Headings 3) Foot notes

2) Quotations 4) Exhibits

1) **Headings.** The following types of headings are commonly used.

– CENTRE HEAD ( All capitals, without underlining)

– Centre Subhead (Capital and lower case, with underlining)

– SIDE HEAD ( All capitals without underlining)

– Side Sub Head (Capital and lower case letters with underlining)

– Paragraph Head followed by a colon (Capital and lower case underline)

Which combinations of headings to use depends on the number of classifications or divisions that a chapter has. The headings are illustrated below:

**Centre Head.** A Centre head is typed in all capital letters. If the title is long, the inverted pyramid style (i.e., the second line shorter than the first, the third line shorter than the second) is used. All caps headings are not underlined. Underlining is unnecessary because capital letters are enough to attract the reader’s attention.

**Example**

CHALKING OUT A PROGRAMME FOR IMPORT SUBSTITUTION AND EXPORT PROMOTION

**Centre Subhead.** The first letter of the first and the last word and all nouns, adjectives, verbs and adverbs in the title are capitalized. Articles, prepositions and conjunctions are not capitalized.

**Example**

Chalking out a Programme for Import Substitution and Export Promotion

**Side Heads.** Words in the side head are either written in all capitals or capitalized as in the centre sub head and underlined.

**Example:** Import Substitution and Export Promotion

**Paragraph Head.** Words in a paragraph head are capitalized as in the centre sub head and underlined. At the end, a colon appears, and then the paragraph starts.

**Example:** Import Substitution and Export Promotion: The Seventh Five-Year

Plan of India has attempted ……

2) **Quotations**

**Quotation Marks:** Double quotation marks (“ ”) are used. A quotation within a quotation is put in single quotation marks (‘ ’). Example: He said, “To the selfish, ‘freedom’ is synonymous with license”.

**When to Use Quotation Marks:** Quotation marks are used for

1) A directly quoted passage or word.

2) A word or phrase to be emphasized, and

3) Titles of articles, chapters, sections of a book, reports, and unpublished works.

**How to Quote:** a) All quotations should correspond exactly to the original in *wording, spelling, and punctuation.*

b) Quotations up to three typewritten lines are run into the text.

c) Direct quotations over three typewritten lines are set in indented paragraphs.

d) Quotation marks are not used for indented paragraphs.

**Five ways of introducing a Quotation:** These are given below.

a) ***Introduction:*** He/she said, “The primary test of success in a negotiation is the presence of goodwill on both sides”.

b) ***Interpolation:*** “The primary test of success in a negotiation”, he/she said, “is the presence of goodwill on both sides”.

c) ***End Reference:*** “ The primary test of success in a negotiation is the presence of goodwill on both sides”, he/she said.

d) ***Indented Paragraph:*** He/she said: For the workers no real advance in their standard of living is possible without steady increase in productivity because any increase in wages generally, beyond certain narrow units, would otherwise be nullified by a rise in prices.

e) ***Running into a Sentence:*** He/she recommended that “joint management councils be set up in all establishments in the public as well as private sector in which conditions favourable to the success of the scheme exist”.

3) **Foot Notes**

A foot note either indicates the source of the reference or provides an explanation which is not important enough to include in the text.

In the traditional system, both kinds of footnotes are treated in the same form and are included either at the bottom of the page or at the end of the chapter or book.

In the modern system, explanatory footnotes are put at the bottom of the page and are linked with the text with a footnote number. But source references are incorporated within the text and are supplemented by a bibliographical note at the end of the chapter or book.

**Types of Footnotes:**

**Rationale of Footnotes:** Footnotes help the readers to check the accuracy of the interpretation of the source by going to the source if they want to. They are also an acknowledgement of the author’s indebtedness to the sources. They lend authority to the work and help the readers to distinguish between the author’s own contribution and that of others.

**Where to put the Footnote:** Footnotes appear at the bottom of the page or at the end of the chapter (before the appendices section).

**Numbering of Footnotes:** a) For any editorial comment on the chapter or title, an asterisk is used.

b) In the text Arabic numerals are used for footnoting. Each new chapter begins with number 1.

c) The number is typed half a space above the line or within parentheses. No space is given between the number and the word. No punctuation mark is used after the number.

d) The number is placed at the end of a sentence or, if necessary to clarify the meaning, at the end of the relevant word or phrase. Commonly, the number appears after the last quotation mark. In an indented paragraph, the number appears at the end of the last sentence in the quotation.

4) **Exhibits**

**Tables:**

**Reference and Interpretation:** Before a table is introduced, it is referred to in the text (e.g., see Table 1.1; refer to Table 1.1; as in Table 1.1; Table 1.1 indicates). A table is meant only to expand, clarify, or give visual explanation rather than stand by itself. The text should highlight the table’s focus and conclusions.

**Identification:** a) Each table is given a number, title, and, if needed, a subtitle. All identifications are centred.

b) Arabic numerals, instead of Roman numerals or capital letters, are recommended for numbering the tables. Usually technical monographs and books contain many tables. As the number increases, Roman numerals become unfamiliar to the reader. Roman numerals also occupy more space than Arabic numerals. If there are more than 26 tables, capital letters will not be sufficient to identify them. Tables can be numbered consecutively throughout the chapter as 1.1, 1.2, 1.3,… wherein the first number refers to the chapter and the second number to the table.

b) For the title and sub title, all capital letter are used.

c) Abbreviations and symbols are not used in the title or sub title.

**Checklist:** Relevance, accuracy, and clarity are of utmost importance in tables.

When entering the table, check the following:

1) Have the explanation and reference to the table been given in the text?

2) Is it essential to have the table for clarity and extra information?

3) Is the representation of the data comprehensive and understandable?

4) Is the table number correct?

5) Are the title and subtitle clear and concise?

6) Are the column headings clearly classified?

7) Are the row captions clearly classified?

8) Are the data accurately entered and represented?

9) Are the totals and other computations correct?

10) Has the source been given?

11) Have all the uncommon abbreviations been spelt out?

12) Have all footnote entries been made?

13) If column rules are used, have all rules been properly drawn?

**Illustrations:** Illustrations cover charts, graphs, diagrams, and maps. Most of the instructions given for tables hold good for illustrations.

**Identification:** Illustrations are identified as FIGURE, CHART, MAP or DIAGRAM. The identification marks (i.e. number, title, and, if any, sub title) are put at the bottom, because an illustration, unlike a table, is studied from bottom upwards.

**Terminal Items**

This section follows the text. First comes the appendices section, then the bibliography and glossary. Each section is separated by a divider page on which only the words APPENDICES, BIBLIOGRAPHY, or GLOSSARY all in capital letters appear.

All reference section pages are numbered in Arabic numerals in continuation with the page numbers of the text.

**1) Appendices**

**What goes into an Appendix:** a) Supplementary or secondary references are put in the appendices section. But all primary reference material of immediate importance to the reader is incorporated in the text. The appendices help the author to authenticate the thesis and help the reader to check the data.

b) The material that is usually put in the appendices is indicated below:

1) Original data

2) Long tables

3) Long quotations

4) Supportive legal decisions, laws and documents

5) Illustrative material

6) Extensive computations

7) Questionnaires and letters

8) Schedules or forms used in collecting data

9) Case studies / histories

10) Transcripts of interviews

**Numbering of Appendices:** The appendices can be serialized with capital letters (Appendix A, Appendix B) to differentiate from the chapter or table numbers.

**References to Appendices:** a) In the text, the reader’s attention is drawn to the appendices as in the case of tables.

b) All appendices are listed in the table of contents.

**2) Bibliographies**

**Positioning of the Bibliography:** The bibliography comes after the appendices section and is separated from it by a division sheet written BIBLIOGRAPHY. It is listed as a major section in all capital letters in the table of contents. A bibliography contains the source of every reference cited in the footnote and any other relevant works that the author has consulted. It gives the reader an idea of the literature available on the subject that has influenced or aided the author.

**Bibliographical Information:** The following information must be given for each bibliographical reference.

**Books Magazines and Newspapers**

1) Author(s) 1) Author(s)

2) Title (underlined) 2) Title of the article (Within quotation marks)

3) Place of publication 3) Title of the magazine (underlined)

4) Publisher 4) Volume number (Roman numerals)

5) Date of publication 5) Serial number (Arabic numerals)

6) Date of issue

3) **Glossary**

**What is a Glossary:** A glossary is a short dictionary giving definitions and examples of terms and phrases which are technical, used in a special connotation by the author, unfamiliar to the reader, or foreign to the language in which the book is written. It is listed as a major section in capital letters in the table of contents.

**Positioning of a Glossary:** The glossary appears after the bibliography. It may also appear in the introductory pages of a book after the lists of tables and illustrations.

**Order of Listing:** Items are listed in alphabetical order.

**Example:**

**Centre Heading** is listed under C and not under H.

4) **Index**

Index may be either subject index or author index. Author index consists of important names of persons discussed in the report, arranged in alphabetical order. Subject index includes a detailed reference to all important matters discussed in the report such as places, events, definitions, concepts etc., and presented in alphabetical order. Index is not generally included in graduate / post graduate students research reports. However, if the report is prepared for publication or intended as a work of reference, an index is desirable.

**Plagiarism**

Plagiarismis the act of stealing someone else's work and attempting to "pass it off" as your own. This can apply to anything, from term papers to photographs to songs, even ideas!

Plagiarism is the use of others' published and unpublished ideas or words (or other intellectual property) without attribution or permission, and presenting them as new and original rather than derived from an existing source. The intent and effect of plagiarism is to mislead the reader as to the contributions of the plagiarizer. This applies whether the ideas or words are taken from abstracts, research grant applications, Institutional Review Board applications, or unpublished or published manuscripts in any publication format (print or electronic).

**Types of Plagiarism**

**Copying:**

The most well-known and, sadly, the most common type of plagiarism is the simplest: **copying**. If you copy someone else's work and put your name on it, you have plagiarized.

**Patchwork Plagiarism:**

The second kind of plagiarism is similar to copying and is perhaps the second most common type of plagiarism: patchwork plagiarism. This occurs when the plagiarizer borrows the "phrases and clauses from the original source and weaves them into his own writing" without putting the phrases in quotation marks or citing the author.

**Paraphrasing Plagiarism:**

The third type of plagiarism is called **paraphrasing plagiarism**. This occurs when the plagiarizer paraphrases or summarizes another's work without citing the source. Even changing the words a little or using synonyms but retaining the author's essential thoughts, sentence structure, and/or style without citing the source is still considered plagiarism.

**Unintentional:**

The fourth type of plagiarism is called unintentional plagiarism -- it occurs when the writer incorrectly quotes and/or incorrectly cites a source they are using.  How is this plagiarism, if the author didn't mean to do it? If a writer has incorrectly quoted or incorrectly cited a source, it could be misconstrued as dishonesty on the writer's part. The dishonest usage of another's work is most often considered plagiarism. Therefore, the incorrect usage of another's work, whether it's intentional or not, could be taken for "real" plagiarism.

**Avoiding Plagiarism**

Avoiding plagiarism is quite simple. The best method for avoiding it is to **simply be honest**; when you've used a source in your research, give credit where it's due. Acknowledge the author of the original work you've used.

* Another way to avoid plagiarism is to **use your own work as often as possible**. Quoting and citing sources is usually required and inevitable when doing research -- that's how you "back up" your own work. But using someone else's work excessively can be construed as plagiarism.
* Another way to it is to **quote and/or cite your sources properly**.In order to properly quote your sources, one should consult the style manual that would be appropriate for the research.
* Proper Citations In order to properly cite your sources, you should also consult the style manual that would be appropriate for the research. The following examples are formatted in MLA, APA, and Chicago (Turabian is similar to Chicago) formats.

**Research Ethics**

Ethics are the principles and guidelines that help us to uphold the things we value. Ethics and law are different aspects, although laws of the land are intended to be based on certain ethics. Almost all societies have legal rules to govern certain behaviour in a country or society, but ethical norms tend to be broader and more informal than laws. An action may be legal but unethical or illegal but ethical. Ethics aim to achieve two fundamental objectives, i.e., to tell us how we ought to act in a given situation and to provide us with strong reasons for doing so.

Ethics always emerge from conflict between values, and research ethics are not an exception. In research, these conflicts may take different forms, such as participant’s concern for privacy versus some justification for manipulation, openness and replication versus confidentiality, present loss versus future benefits and so on. Each decision made in research involves a potential compromise of one value for another. However, still researchers must try to minimize the risk to participants, colleagues and society while trying to optimize the quality of outcome. Research ethics help us to reconcile value conflicts.

The benefits of observing ethics in research studies are as follows:

**1.** It helps in promoting the aims of research, such as bringing out the truth and avoidance of errors.

**2.** It promotes the values that are essential to collaborative work, such as trust, accountability, mutual respect and fairness.

**3.** It holds the researcher accountable to the public and society.

**4.** It helps in building public support for research, which in return can help in getting participants who take part in the research willingly.

**Some desirable elements to ensure Ethics in research**

The following is a general summary of some research ethical guidelines and principals that various codes address.

**1.** Honesty in reporting data, results, methods and procedures and publication status.

**2.** Objectivity to avoid bias in experimental design, data analysis, interpretation and peer review.

**3.** Integrity, acting with sincerity, striving for consistency of thought and action.

**4.** Carefulness to avoid careless errors and negligence and proper documentation of all aspects.

**5.** Openness in sharing data, results, ideas, tools, resources and openness to criticisms and new ideas.

**6.** Respect for intellectual property rights, such as patents, copyrights and other forms of intellectual property.

**7.** Confidentiality in context of communications, personal records and privacy issues.

**8.** Responsible publication with an aim to serve the society. Avoiding wasteful and duplicative publication.

**9.** Responsible mentoring in terms of guiding research students.

**10.** Respect for colleagues translates to extending fair treatment to the colleagues.

**11.** Social responsibility means to serve the society and different stakeholders.

**12.** Non-discrimination against colleagues or students on the basis of sex, race or factors that are not related to their scientific competence and integrity.

**13.** Enhancing competence for own professional advancement or lifelong learning and taking steps to promote competence in science as a whole.

**14.** Ensuring legality of the whole process by obeying relevant laws, i.e., institutional and governmental policies.

**15.** Animal care through proper experimental designs.

**Key terms in Research Ethical Issues**

Some of the **key terms** used in the context of ethical issues concerning researchers are as follows:

**1. Fabricating behaviour:** Creation of spurious data by researcher, their recording and drawing inferences.

**2. Falsification:** It manipulates the research material, equipment and processes or changes or omits data or results such that the research is not accurately represented in the research records.

**3. Plagiarism:** It is the act of appropriating somebody else’s ideas, thoughts, pictures, theories, words or stories as your own. If a researcher plagiarizes the work of others, the integrity, ethics and trustworthiness of the sum total of his or her research becomes questionable. Plagiarism is both an illegal and punishable act and is considered to be on the same level as stealing from the author who originally created it. It can take the following forms.

**(a) Intra-corpal:** A case of plagiarism where one student has copied from another in the same submission is known as intra-corpal plagiarism.

**(b) Extra-corpal:** It is an instance of plagiarism where a student has copied the material from an external source (***Example:*** Books, journal article, world wide web, etc.).

**(c) Autoplagiarism:** It is citing one’s own work without acknowledgement.

**4. Multiple authorship:** There can be many improprieties in authorship. Improper assignment of credit, such as excluding other authors, inclusion of other as authors who have not made a definite contribution towards the work published or submission of multi-authored publication without the knowledge of all the authors.

**5. Peer review:** It is the process in which an author submits a written manuscript or an article to a journal for publication. The journal editor distributes the article to experts or reviewers. The peer review process seldom proceeds in a straight line. The entire process may involve several rounds of communication between the editor, the reviewers and the original author before an article is ready for publication. The two most important ethics in the process are maintaining confidentiality and protection of intellectual property. Reviewers and author should not know the names of each other. Only then, the peer review process can be genuinely open and beneficial. None in the process can publicly disclose the information in the article or use the information in a submitted article for personal gain.

**6. Duplicate and partial publication:** It is publishing the same data and same results in more than one publication or journal. This is unethical but may be acceptable in certain cases, such as publishing results in a journal to provide research participants with a summary of the results. Partial publication involves publishing parts of your results in different journals. It is specifically unethical for a small, focused study. However, in case of large studies with many variables, this may be acceptable as different publications involve different research questions and different data and it actually advances the interest of the study.

**Important measures to make research more ethical**

**1. Informed consent:** The provision of informed consent also includes the knowledge that the informed participation is voluntary and that participants can withdraw from the study at any time.

**2. Protective research design:** This involves estimating the probability of happening of harmful effects, their severity and the likely duration of these effects.

**3. Screening:** It is an attempt to select only those individuals for study who show a high tolerance for potential risks.

**4. Pilot studies:** When the potential harms are uncertain, a useful precaution involves a pilot study with follow-up diagnostic interviews to assess the effects and request advice from the participants.

**5. Outside proposal review:** Requesting others to review research proposals is a helpful precaution in minimizing risks.

**6. Professional codes:** Two features of professional codes are important for discussion. Firstly, professional codes have been developed inductively from the wide research experiences of professionals.

Secondly, professional codes place strong emphasis on researchers’ responsibility for their research.

**7. Government regulations:** Government regulations such as state and central laws are designed to protect or advance the interests of society and its individuals. Thus, the researchers are required to take certain precautions.

**References**

1. Anderson, Sweeney and Willams, (2012) *Statistics for Business and Economics*, Thomson South Western, US.
2. Hair, Anderson, Tatham and Black (2009) *Multivariate Data Analysis*, Pearson Education, Singapore
3. Bryman Alan, (2012) *Social Research Methods*, Oxford University Press, Oxford.
4. Kothari C.R., (2014) *Research Methodology,* New Age International Publication, New Delhi.
5. Krishnawamy O.R. and Ranghanathan, M., (2011) *Methodology of Research in Social Sciences,* Himalaya Publishing House, Bangalore.
6. Kurian C.T. (1984) *Research Methodology in Economics***,** Institute of Development Studies, Madras.
7. Majumdar P.K., (2011) *Research Methods in Social Science*, Viva Books Private Limited, New Delhi.
8. Robert, A. Day, (2011) *How to Write and Publish a Scientific Paper,* Cambridge University Press, Great Britain.