Research Methodology
BUS 610, MASTER OF BUSINESS ADMINISTRATION, M.B.A

4/20/2016
Kings University,
Honolulu, Hawaii

Learning Outcomes

Table of Contents

UNIT III : MEASUREMENT & DATA COLLECTION 49

UNIT IV : DATA PROCESSING AND ANALYSIS 65

UNIT 1: Research Methodology: An Introduction

Meaning of research

Research in common parlance refers to a search for knowledge. Once can also define research as a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation. The Advanced Learner's Dictionary of Current English lays down the meaning of research as "a careful investigation or inquiry especially through search for new facts in any branch of knowledge."[1] Redman and Mory define research as a "systematized effort to gain new knowledge."

Research is an academic activity and as such the term should be used in a technical sense. According to Clifford Woody research comprises defining and redefining problems, formulating hypothesis or suggested solutions; collecting, organising and evaluating data; making deductions and reaching conclusions; and at last carefully testing the conclusions to determine whether they fit the formulating hypothesis. D. Slesinger and M. Stephenson in the Encyclopaedia of Social Sciences define research as "the manipulation of things, concepts or symbols for the purpose of generalising to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art."[3] Research is, thus, an original contribution to the existing stock of knowledge making for its advancement. It is the persuit of truth with the help of study, observation, comparison and experiment. In short, the search for knowledge through objective and systematic method of finding solution to a problem is research. The systematic approach concerning generalisation and the formulation of a theory is also research. As such the term 'research' refers to the systematic method consisting of enunciating the problem, formulating a hypothesis, collecting the facts or data, analysing the facts and reaching certain conclusions either in the form of solutions(s) towards the concerned problem or in certain generalisations for some theoretical formulation.

Objectives of research

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet. Though each research study has its own specific purpose, we may think of research objectives as falling into a number of following broad groupings:
1. To gain familiarity with a phenomenon or to achieve new insights into it (studies with this object in view are termed as *exploratory* or *formulative* research studies);

2. To portray accurately the characteristics of a particular individual, situation or a group (studies with this object in view are known as *descriptive* research studies);
3. To determine the frequency with which something occurs or with which it is associated with something else (studies with this object in view are known as *diagnostic* research studies);
4. To test a hypothesis of a causal relationship between variables (such studies are known as *hypothesis-testing* research studies).

Research and Scientific Method

For a clear perception of the term research, one should know the meaning of scientific method. The two terms, research and scientific method, are closely related. Research, as we have already stated, can be termed as "an inquiry into the nature of, the reasons for, and the consequences of any particular set of circumstances, whether these circumstances are experimentally controlled or recorded just as they occur. Further, research implies the researcher is interested in more than particular results; he is interested in the repeatability of the results and in their extension to more complicated and general situations."7 On the other hand, the philosophy common to all research methods and techniques, although they may vary considerably from one science to another, is usually given the name of scientific method. In this context, Karl Pearson writes, "The scientific method is one and same in the branches (of science) and that method is the method of all logically trained minds … the unity of all sciences consists alone in its methods, not its material; the man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequences, is applying the Scientific Method and is a man of science." Scientific method is the pursuit of truth as determined by logical considerations. The ideal of science is to achieve a systematic interrelation of facts. Scientific method attempts to achieve "this ideal by experimentation, observation, logical arguments from accepted postulates and a combination of these three in varying proportions."9 In scientific method, logic aids in formulating propositions explicitly and accurately so that their possible alternatives become clear. Further, logic develops the consequences of such alternatives, and when these are compared with observable phenomena, it becomes possible for the researcher or the scientist to state which alternative is most in harmony with the observed facts. All this is done through experimentation and survey investigations which constitute the integral parts of scientific method.

Experimentation is done to test hypotheses and to discover new relationships,If any, among variables. But the conclusions drawn on the basis of experimental data are generally criticized for faulty assumptions, poorly designed experiments, badly executed experiments or faulty interpretations. As such the researcher must pay all possible attention while developing the experimental design and must state only

probable inferences. The purpose of survey investigations may also be to provide scientifically gathered information to work as a basis for the researchers for their conclusions. The scientific method is, thus, based on certain basic postulates which can be stated as under: 1. It relies on empirical evidence; 2. It utilizes relevant concepts; 3. It is committed to only objective considerations; 4. It presupposes ethical neutrality, i.e., it aims at nothing but making only adequate and correct statements about population objects; 5. It results into probabilistic predictions; 6. Its methodology is made known to all concerned for critical scrutiny are for use in testing the conclusions through replication; 7. It aims at formulating most general axioms or what can be termed as scientific theories.

Types of research

The basic types of research are as follows:

(i) *Descriptive vs. Analytical: Descriptive research* includes surveys and fact-finding enquiries of different kinds. The major purpose of descriptive research is description of the state of affairs as it exists at present. In social science and business research we quite often use the 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. Most *ex post facto research* projects are used for descriptive studies in which the researcher seeks to measure such items as, for example, frequency of shopping, preferences of people, or similar data. *Ex post facto studies* also include attempts by researchers to discover causes even when they cannot control the variables. The methods of research utilized in descriptive research are survey methods of all kinds, including comparative and correlational methods. In *analytical research*, on the other hand, the researcher has to use facts or information already available, and analyse these to make a critical evaluation of the material.

(ii) *Applied vs. Fundamental*: Research can either be applied (or action) research or fundamental (to basic or pure) research. *Applied research* aims at finding a solution for an immediate problem facing a society or an industrial/business organisation, whereas *fundamental research* is mainly concerned with generalisations and with the formulation of a theory. "Gathering knowledge for knowledge's sake is termed 'pure' or 'basic' research."[4] Research concerning some natural phenomenon or relating to pure mathematics are examples of fundamental research. Similarly, research studies, concerning human behaviour carried on with a view to make generalisations about human behaviour, are also examples of fundamental research, but research aimed at certain conclusions (say, a solution) facing a concrete social or business problem is an example of applied research. Research to identify social, economic or political trends that may affect a particular institution or the copy research (research to find out whether certain communications will be read and

understood) or the marketing research or evaluation research are examples of applied research. Thus, the central aim of applied research is to discover a solution for some pressing practical problem, whereas basic research is directed towards finding information that has a broad base of applications and thus, adds to the already existing organized body of scientific knowledge.

(iii) *Quantitative vs. Qualitative:* 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 especially 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.

(iv) *Conceptual vs. Empirical:* Conceptual research is that related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones. On the other hand, empirical research relies on experience or observation alone, often without due regard for system and theory. It is data-based research, coming up with conclusions which are capable of being verified by observation or experiment.

We can also call it as experimental type of research. In such a research it is necessary to get at facts first hand, at their source, and actively to go about doing certain things to stimulate the production of desired information. In such a research, the researcher must first provide himself with a working hypothesis or guess as to the probable results. He then works to get enough facts (data) to prove or disprove his hypothesis. He then sets up experimental designs which he thinks will manipulate the persons or the materials concerned so as to bring forth the desired information. Such research is thus characterised by the experimenter's control over the variables under study and his deliberate manipulation of one of them to study its effects. Empirical research is appropriate when proof is sought that certain variables affect

other variables in some way. Evidence gathered through experiments or empirical studies is today considered to be the most powerful support possible for a given hypothesis.

(v) *Some Other Types of Research:* All other types of research are variations of one or more

of the above stated approaches, based on either the purpose of research, or the time required to accomplish research, on the environment in which research is done, or on the basis of some other similar factor. Form the point of view of time, we can think of research either as *one-time research or longitudinal research*. In the former case the research is confined to a single time-period, whereas in the latter case the research is carried on over several time-periods. Research can be *field-setting research or laboratory research or simulation research*, depending upon the environment in which it is to be carried out. Research can as well be understood as *clinical or diagnostic research*. Such research follows case-study methods or in depth approaches to reach the basic causal relations. Such studies usually go deep into the causes of things or events that interest us, using very small samples and very deep probing data gathering devices. The research may be *exploratory* or it may be formalized. The objective of exploratory research is the development of hypotheses rather than their testing, whereas formalized research studies are those with substantial structure and with specific hypotheses to be tested. *Historical research* is that which utilizes historical sources like documents, remains, etc. to study events or ideas of the past, including the philosophy of persons and groups at any remote point of time. Research can also be classified as *conclusion-oriented* and decision-oriented. While doing conclusion oriented research, a researcher is free to pick up a problem, redesign the enquiry as he proceeds and is prepared to conceptualize as he wishes. Decision-oriented research is always for the need of a decision maker and the researcher in this case is not free to embark upon research according to his own inclination. Operations research is an example of decision oriented research since it is a scientific method of providing executive departments with a quantitative basis for decisions regarding operations under their control.

Research Process

Before embarking on the details of research methodology and techniques, it seems appropriate to present a brief overview of the research process. Research process consists of series of actions or steps necessary to effectively carry out research and the desired sequencing of these steps. The chart shown in Figure 1.1 well illustrates a research process

Review the literature

| Define research problem | Review concepts and theories | Formulate hypotheses | Design research (including sample design) | Collect data (Execution) | Analyse data (Test hypotheses if any) | Interpret and report |

I   II   III   IV   V   VI   VII

Where (F) = feed back (Helps in controlling the sub-system to which it is transmitted)
(FF) = feed forward (Serves the vital function of providing criteria for evaluation)

Fig. 1.1

The chart indicates that the research process consists of a number of closely related activities, as shown through I to VII. But such activities overlap continuously rather than following a strictly prescribed sequence. At times, the first step determines the nature of the last step to be undertaken. If subsequent procedures have not been taken into account in the early stages, serious difficulties may arise which may even prevent the completion of the study. One should remember that the various steps involved in a research process are not mutually exclusive; nor they are separate and distinct. They do not necessarily follow each other in any specific order and the researcher has to be constantly anticipating at each step in the research process the requirements of the subsequent steps. However, the following order concerning various steps provides a useful procedural guideline regarding the research process: (1) formulating the research problem; (2) extensive literature survey;(3) developing the hypothesis; (4) preparing the research design; (5) determining sample design;(6) collecting the data; (7) execution of the project; (8) analysis of data; (9) hypothesis testing;(10) generalisations and interpretation, and (11) preparation of the report or presentation of the results, i.e., formal write-up of conclusions reached.

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

1. Formulating the research problem: There are two types of research problems, viz., those which relate to states of nature and those which relate to relationships between variables. At the very outset the researcher must single out the problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject-matter that he would like to inquire into. Initially the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved. Then, the feasibility of a particular solution has to be considered before a

working formulation of the problem can be set up. The formulation of a general topic into a specific research problem, thus, constitutes the first step in a scientific enquiry. Essentially two steps are involved in formulating the research problem, viz., understanding the problem thoroughly, and rephrasing the same into meaningful terms from an analytical point of view.

The best way of understanding the problem is to discuss it with one's own colleagues or with those having some expertise in the matter. In an academic institution the researcher can seek the help from a guide who is usually an experienced man and has several research problems in mind. Often, the guide puts forth the problem in general terms and it is up to the researcher to narrow it down and phrase the problem in operational terms. In private business units or in governmental organisations, the problem is usually earmarked by the administrative agencies with who the researcher can discuss as to how the problem originally came about and what considerations are involved in its possible solutions.

The researcher must at the same time examine all available literature to get himself acquainted with the selected problem. He may review two types of literature—the conceptual literature concerning the concepts and theories, and the empirical literature consisting of studies made earlier which are similar to the one proposed. The basic outcome of this review will be the knowledge as to what data and other materials are available for operational purposes which will enable the researcher to specify his own research problem in a meaningful context. After this the researcher rephrases the problem into analytical or operational terms i.e., to put the problem in as specific terms as possible. This task of formulating, or defining, a research problem is a step of greatest importance in the entire research process. The problem to be investigated must be defined unambiguously for that will help discriminating relevant data from irrelevant ones. Care must however, be taken to verify the objectivity and validity of the background facts concerning the problem. Professor W.A. Neiswanger correctly states that the statement of the objective is of basic importance because it determines the data which are to be collected, the characteristics of the data which are relevant, relations which are to be explored, the choice of techniques to be used in these explorations and the form of the final report. If there are certain pertinent terms, the same should be clearly defined along with the task of formulating the problem. In fact, formulation of the problem often follows a sequential pattern where a number of formulations are set up, each formulation more specific than the preceding one, each one phrased in more analytical terms, and each more realistic in terms of the available data and resources.

2. Extensive literature survey: Once the problem is formulated, a brief summary of it should be written down. It is compulsory for a research worker writing a thesis for a Ph.D. degree to write a synopsis of the topic and submit it to the necessary Committee or the Research Board for approval. At this juncture the researcher should undertake extensive literature survey connected with the problem. For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals, conference proceedings, government reports, books etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another. The earlier studies, if any, which are similar to the study in hand, should be carefully studied. A good library will be a great help to the researcher at this stage.

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

How does one go about developing working hypotheses? The answer is by using the following approach:
(a) Discussions with colleagues and experts about the problem, its origin and the objectives in seeking a solution;
(b) Examination of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;
(c) Review of similar studies in the area or of the studies on similar problems; and
(d) Exploratory personal investigation which involves original field interviews on a limited scale with interested parties and individuals with a view to secure greater insight into the practical aspects of the problem.

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

4. Preparing the research design: The research problem having been formulated in clear cut terms, the researcher will be required to prepare a research design, i.e., he will have to state the conceptual structure within which research would be conducted. The preparation of such a design facilitates research to be as efficient as possible yielding maximal information. In other words, the function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money. But how all these can be achieved depends mainly on the research purpose. Research purposes may be grouped into four categories, viz., (i) Exploration, (ii) Description,(iii) Diagnosis, and (iv)experimentation. A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose of the research study is that of exploration. But when the purpose happens to be an accurate description of a situation or of an association between variables, the suitable design will be one that minimises bias and maximises the reliability of the data collected and analysed.

There are several research designs, such as, experimental and non-experimental hypothesis testing. Experimental designs can be either informal designs (such as before-and-after without control, after-only with control, before-and-after with control) or formal designs (such as completely randomized design, randomized block design, Latin square design, simple and complex factorial designs), out of which the researcher must select one for his own project.
The preparation of the research design, appropriate for a particular research problem, involves usually the consideration of the following:

(i)The means of obtaining the information;
(ii)The availability and skills of the researcher and his staff (if any);
(iii)Explanation of the way in which selected means of obtaining information will be organised and the reasoning leading to the selection;
(iv)The time available for research; and
(v)The cost factor relating to research, i.e., the finance available for the purpose.

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

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

6. Collecting the data: In dealing with any real life problem it is often found that data at hand are inadequate, and hence, it becomes necessary to collect data that are appropriate. There are several ways of collecting the appropriate data which differ considerably in context of money costs, time and other resources at the disposal of the researcher. Primary data can be collected either through experiment or through survey. If the researcher conducts an experiment, he observes some quantitative measurements, or the data, with the help of which he examines the truth contained in his hypothesis. But in the case of a survey, data can be collected by any one or more of the following ways:

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

(ii) *Through personal interview:* The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

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

(iv) *By mailing of questionnaires:* The researcher and the respondents do come in contact with each other if this method of survey is adopted. Questionnaires are mailed to the respondents with a request to return after completing the same. It is the most extensively used method in various economic and business surveys. Before applying this method, usually a Pilot Study for testing the questionnaire is conduced which reveals the weaknesses, if any, of the questionnaire. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information.

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

The researcher should select one of these methods of collecting the data taking into consideration the nature of investigation, objective and scope of the inquiry, financial resources, available time and the desired degree of accuracy.

7. Execution of the project: Execution of the project is a very important step in the research process. If the execution of the project proceeds on correct lines, the data to be collected would be adequate and dependable. The researcher should see that the project is executed in a systematic manner and in time. If the survey is to be conducted by means of structured questionnaires, data can be readily machine-processed. In such a situation, questions as well as the possible answers may be coded. If the data are to be collected through interviewers, arrangements should be made for proper selection and training of the interviewers. The training may be given with the help of instruction manuals which explain clearly the job of the interviewers at each step. Occasional field checks should be made to ensure that the interviewers are doing their assigned job sincerely and efficiently. A careful watch should be kept for unanticipated factors in order to keep the survey as much realistic as possible. This, in other words, means that steps should be taken to

ensure that the survey is under statistical control so that the collected information is in accordance with the pre-defined standard of accuracy. If some of the respondents do not cooperate, some suitable methods should be designed to tackle this problem. One method of dealing with the non-response problem is to make a list of the non-respondents and take a small sub-sample of them, and then with the help of experts vigorous efforts can be made for securing response.

8. Analysis of data: After the data have been collected, the researcher turns to the task of analysing them. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis. Thus, researcher should classify the raw data into some purposeful and usable categories. *Coding* operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted. *Editing* is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. *Tabulation* is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, especially in large inquiries, is tabulated by computers. Computers not only save time but also make it possible to study large number of variables affecting a problem simultaneously.

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

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

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

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

1. The layout of the report should be as follows: (*i*) the preliminary pages; (*ii*) the main text, and (*iii*) the end matter.
In its preliminary pages the report should carry title and date followed by acknowledgements and foreword. Then there should be a table of contents followed by a list of tables and list of graphs and charts, if any, given in the report.
The main text of the report should have the following parts:
(a) *Introduction:* It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.
(b) *Summary of findings:* After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarised.
(c) *Main report:* The main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.
(d) *Conclusion:* Towards the end of the main text, researcher should again put down the results of his research clearly and precisely. In fact, it is the final summing up.

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

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

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

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

Criteria of Good Research

Whatever may be the types of research works and studies, one thing that is important is that they all meet on the common ground of scientific method employed by them. The qualities of a good research are:

1. *Good research is systematic:* It means that research is structured with specified steps to be taken in a specified sequence in accordance with the well-defined set of rules. Systematic characteristic of the research does not rule out creative thinking but it certainly does reject the use of guessing and intuition in arriving at conclusions.

2. *Good research is logical:* This implies that research is guided by the rules of logical reasoning and the logical process of induction and deduction are of great value in carrying out research. Induction is the process of reasoning from a part to the whole whereas deduction is the process of reasoning from some premise to a conclusion which follows from that very premise. In fact, logical reasoning makes research more meaningful in the context of decision making.

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

4. *Good research is replicable:* This characteristic allows research results to be verified by replicating the study and thereby building a sound basis for decisions.

Defining the research Problem

The definition of the problem stage in the research process is an extremely important one. Without a properly defined problem the research cannot progress. One cannot design a research plan without a very clear idea of what needs to be accomplished.

This stage can also be a very challenging one. The definition of the problem means more than just choosing a topic of interest.

Before we look at the train that will take us to our destination we need first to identify our final destination. Before we can understand how to develop a well-defined research problem we need to know what the goal or outcome of this process is.

The goal of this process of problem definition is to create research questions and hypotheses.

What are theories, research questions and hypotheses?

A theory is a set of propositions used to explain behaviours or phenomena to provide prediction and understanding.

A *research question* is a question that can be answered directly through the analysis of data. In other words, a research question is a question that is specific enough that we can create a survey question(s) to answer it.

A *hypothesis*, on the other hand, is the researcher's best guess to the answer of the research question. Hypotheses are generated from our theories. Research questions are often attempts to refute/validate theories through the testing of hypotheses. Though sometimes research questions are more information collecting, and thus a-theoretical.

The goal of this process of problem definition is to create research questions and hypotheses that:

- Are measurable--quantifiable/testable
- Are well-defined--no ambiguous language
- Are useful in decision-making or in answering the overall problem *(What is the goal of business research?)*

- Are directly connected to one another--hypothesis is not only a plausible answer to the research question but also directly answers the research question
- Encompass the full scope of the problem--Have all the important or relevant questions been asked?

In addition to these specific characteristics, research questions and hypotheses typically attempt to measure one or more of the following concepts:

- awareness: measure of knowledge
- behaviour: measure of events that are or have taken place
- motivation: measure of why people behave as they do
- opinion: measure of impression of worth or quality
- preference: measure of likes/dislikes
- desire: measure of wants
- interest: measure of concerns/curiosities
- intention: measure of anticipated behaviour
- demographic: measure of respondent's characteristics

Examples:

Broad Problem: Young adults binge drink beer too frequently.

Theory: The law of demand. Increases in price, reduces quantity demanded.

Research Question: Will a change in beer taxes alter consumption of beer?

Hypothesis: An increase in the beer tax will lower beer consumption.

Sometimes theory offers precise or sharp quantitative predictions while other times it merely suggests the direction of the effect as in this case. Still other times the prediction is ambiguous, depending on the relative size of competing effects.

Hypothesis Formulation

Hypotheses are the statements made on the population parameters such as the average desired speed of the car drivers is 75 km/h. Procedure adopted in testing the statistical validity of this statement is known as hypothesis testing. Sample data is used to prove or disprove this statement. In the hypothesis testing it is necessary

to make two statements about the parameter subjected to testing. These two are given below;

H_0: μ = 75 km/h

H_1: μ > 75 km/h

H_0 is called null hypothesis and the other one is known as the alternative hypothesis. The alternative hypothesis, shown above, is a one-sided alternative hypothesis. If H_1 is μ ≠ 75 km/h then it is called as two-sided alternative hypothesis since the average speed can be greater or lesser than 75 km/h.

This testing is specifically useful in knowing the effectiveness of some changes made in the model or experimental setup, improvement in the controlling measures etc. In case of concrete compressive strength, if some admixtures are used to improve the strength, or the changes in the proportion of the aggregates, it is necessary to know the after effects of these changes.

As discussed before, in hypothesis testing sample data is crucial and this testing is carried out on the basis of the statistics calculated from this data. There is a chance to commit error while testing the hypothesis since the testing is done on the basis of sample data.

Types of Errors

Two types of errors are normally committed while testing the hypothesis using the sample data. If the null hypothesis is rejected when it is true, then the error committed is known as type I error. When the analyst has failed to reject the null hypothesis even when it is not true then type II error is committed.

Probability of committing type I error is nothing but the significance level used for the hypothesis testing. Committing type I error can be controlled by the analyst by choosing the appropriate significance level for the hypothesis testing. Significance level used in the hypothesis testing controls the critical region for the test statistic. Since committing type I error is in control of the analyst, *rejecting the null hypothesis would be always a strong conclusion.*
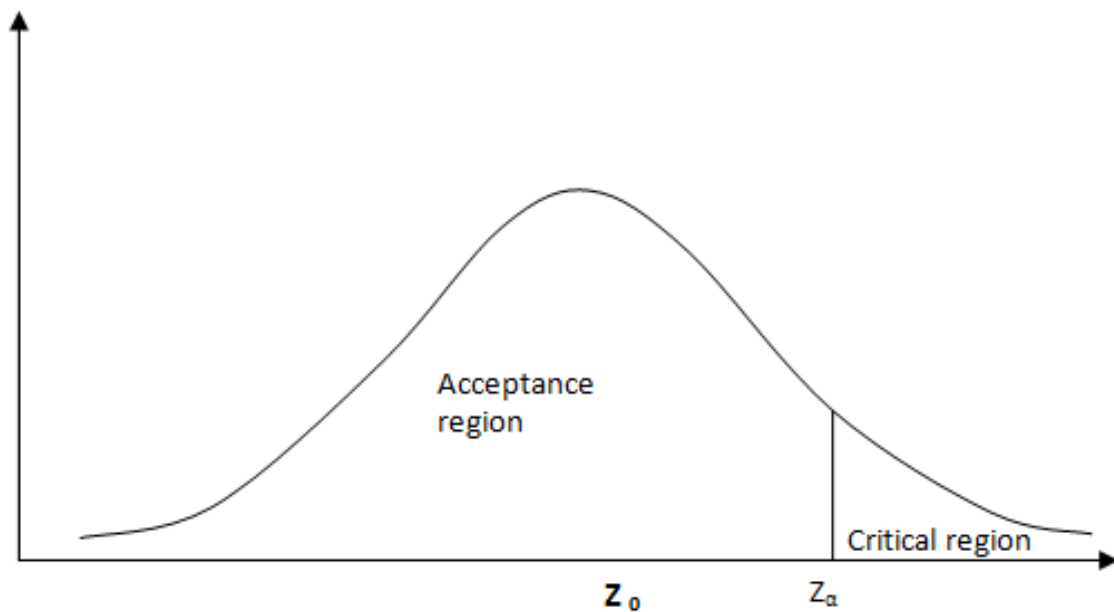
Figure 8.1: Acceptance and rejection regions for one-tailed (upper) testin

Hypotheses formulation requires understanding about the population data that is under consideration for testing. Hypotheses formulation related to the average speed of a particular vehicle type moving on a road section is discussed in the remainder of this section.

A road section has been upgraded and the analyst is optimistic about the increase in the average speed of the vehicles moving on that road section. In other words, the analyst is interested in proving that there is an increase in the average speed, compared to the speeds observed on the same road before up-gradation. As discussed earlier, rejecting the null hypothesis is a strong conclusion in hypotheses testing. On this basis, it can be said that whatever the hypothesis the analyst wants to reject must be taken as the null hypothesis. In the above example, null hypothesis may be "there is no significant change in the average speed of the vehicles even after upgrading the road section". Ideal outcome from this test should be 'reject null hypothesis', i.e. there is a significant change in the speeds of vehicles after up-gradation of the road section.

Some other examples are – If the problem is "when concrete properties are changed considerably when some admixtures are added" the null hypothesis may be no significant change in the concrete characteristics, - "In the trip generation modeling a particular variable is significantly contributing in explaining the variability of the trips generated", the null hypothesis may be that particular variable's contribution is nil, - "In a particular city modal share of public transport exceeds certain value", the

null hypothesis may be the proportion of public transport users are less than that value.

Hypothesis Testing

Another crucial aspect of the hypotheses testing is deciding the significance level of the test, α. As seen in the previous lectures the significance of test is nothing but 'what is the probability of committing type I error (rejecting null hypothesis even when it is true)', given the sample data. Decision on its value alters the test result since the significance value decides the critical regions of the test. If the analyst goes for smaller α value, the critical regions for rejecting the null hypothesis would be smaller and rejection of null hypothesis may not be possible.

The procedure for hypotheses testing is given below;

1. Formulate the null hypothesis, H_0. In the hypotheses testing the decision is made on this statement. As discussed earlier this statement should be the one the analyst is interested in rejecting. The acceptance and the rejection regions shown in Figure 3 are related to this statement.
2. Formulate the alternate hypothesis, H_1. This is what the analyst is interested in proving with the given sample data.
3. Identify and compute the test statistics from the sample data.
4. Identify the corresponding statistical distribution. This was discussed in detail in the previous lectures.
5. Assume significance level at which the test is to be carried out. Compute the confidence interval for the population parameter which is being tested. This interval is nothing but the acceptance region for the null hypothesis. Alternatively, the critical regions can also be identified based on the assumed significance level and the sampling distribution of the test statistic.
6. Test whether the statistic is falling in the acceptance region or not. If the test statistic is not falling in the acceptance region reject the null hypothesis. Rejecting null hypothesis implies the analyst is able to prove his statement on population parameter based on the given sample data.

Hypothesis testing on population mean

Two situations may arise out of this, first one is when the population variance is known and the second situation is if the population variance is unknown.

Population variance known

The hypothesis on population mean is necessary when it is unknown. Hypothesis is nothing but assuming certain mean for population data. In hypothesis testing, using the sample data collected from the same population the truth in the assumption would be tested. The hypotheses are formulated as follows;

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$ or $H_1: \mu > \mu_0$ or $H_1: \mu < \mu_0$

As discussed above another crucial aspect in hypothesis testing is the nature of the sampling distribution. If the Population is normal then the sampling statistic, i.e. $\bar{X}$, will also be normally distributed. Since the population variance σ^2 is known the test statistic for this hypothesis testing would be as follows;

$$Z_0 = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

If the significance of the test is α then from the normal tables the values $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ (for two tailed testing or Z_α for one tailed testing) can be obtained. If the value of Z_0 is falling in the critical region, reject the null hypothesis, if not falling in the critical region it can be said that the analyst has failed to reject the null hypothesis. Rejecting null hypothesis means the assumed population mean is different from the actual population mean at the significance level α. If the test is conducted at a different significance level the result of the hypothesis testing would be different.

Using the same sample data, the least significance level at which the null hypothesis can be rejected is known as p-value. If the test statistic follows standard normal distribution the p-values can be found as follows;

$$P = \begin{cases} 2[1 - \Phi(|Z_0|)] \, for \, a \, two \, tailed \, test \\ 1 - \Phi(Z_0) \, for \, an \, upper \, tail \, test \\ \Phi(Z_0) \, for \, a \, lower \, tail \, test \end{cases}$$

$\Phi(Z_0)$ represents the area under the normal curve on the left side of Z_0.

Another method for testing the hypotheses

Another way of testing the hypotheses is to find the confidence interval for the population parameter as discussed in the previous chapter. If the assumed population mean is not falling in the confidence interval then the null hypothesis is rejected otherwise the conclusion is that the analyst has failed to reject the null hypothesis.

Hypotheses testing on the means of two population data

When two populations of interest are to be compared using the unknown means and known variances the following procedure is to be adopted. If the populations are normally distributed and have the parameters μ_1, σ_1^2 and μ_2, σ_2^2, respectively, the comparison between μ_1 and μ_2 would be made based on two samples collected from the respective populations. Hypotheses on the two means could be as follows;

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Or

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

Here the test statistic would be the difference in the means of both the samples, that is $\bar{X}_1 - \bar{X}_2$. If both the populations are normally distributed, this statistic is normally distributed with parameters $\mu_1 - \mu_2$ and $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$. The transformed test statistic would be as follows;

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$ With mean 0 and variance 1.

When the Variances are unknown

As with the case of tests on the population mean when the variances are unknown the test statistic follows t-distribution. The following are the two sub cases that may arise out of this problem.

When the variances of the two populations are not equal

Since the variances of both the populations are unknown it can be assumed that either both of them are equal or not equal. In this case it is assumed that both the variances are not equal. The sampling distribution of $\bar{X}_1 - \bar{X}_2$ would have mean, μ_1 - μ_2 and variance, $\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}$.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

For the above hypotheses,

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$ is t-distributed with the degrees of freedom as given below;

$$\text{Degrees of freedom} = \frac{\left(\frac{S_1^2}{n_1} - \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 / n_1 + 1 + \left(\frac{S_1^2}{n_1}\right)^2 / n_1 + 1} - 2$$

For the above hypotheses formulation except the above variations the testing procedure would be similar to the one discussed in previous sections.

When the variances of the two populations are equal

Important aspect of this would the common variance for both the populations since it is assumed that the variances are equal. Sample data taken from both the samples would be used in getting the variance of the pooled data.

Variance of the pooled data, $$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

For the above hypotheses the test statistic is

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

which is t-distributed with $n_1 + n_2 - 2$ degrees of freedom

Hypotheses testing for goodness of fit

When the population distribution function is known to follow certain probability distribution this test is used for knowing the goodness-of-fit. Here also a sample data is to be collected from the population data and the frequency analysis has to be carried out on the data. The total data are divided into k classes, mostly depending upon the range of the observed data and any particular class must not have less than 8 observations. It is also necessary to calculate the parameters of the assumed probability distribution from the sample data. For example, if the time headways observed at a particular road section are assumed to follow negative exponential distribution, the mean headway is to be calculated from the observed data.

With the help of the assumed distribution and the parameters calculated from the sample data theoretical frequencies are to be generated. If O_i is the observed frequency and the E_i is the expected frequency then the following is the test statistic;

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

The test statistic follows chi-square distribution with k-p-1 degrees of freedom. p is the number of parameters in the assumed distribution and the 1 is used since the total frequency has been used in getting the theoretical frequency. From the assumed significance level and the degrees of freedom the critical regions are determined.

In this case the null hypothesis is 'the assumed distribution fits to the population data at the given significance level'. If the test statistic falls in the critical region reject the null hypothesis otherwise it may be said that the analyst has failed to see any difference between the theoretical and observed frequencies hence failed to reject the null hypothesis.

As can be seen from the formula for the chi-square observed value, it always takes positive values and no information is available on the location of the differences between the observed and the theoretical frequencies

Other Tests for goodness-of-fit

Most of the times, it is assumed that the population data follow certain distribution (in majority of the cases it is assumed to follow normal distribution). There are also problems where the distributional assumption cannot be made on the population data. In this scenario, some non-parametric tests are available for testing the goodness-of-fit of the observed and the theoretical densities.

Regression analysis:

Understanding of the probability concepts are compulsory when dealing with the regression models since the sample data are used in building the regression model (essentially in estimating the parameters of the regression model). Regression models are useful in predicting/estimating the average value of the dependent variable based on the value of the independent variable. A simple linear regression model capturing the variability of a dependent variable Y using the variability of an independent variable X is explained here. Here the term linear refers to the linearity in terms of the parameters and not in terms of the explanatory variables.

Y = m*X + c + U

Where, U is the variable that takes the value of the error or the difference between the actual/observed value of Y and the value of Y resulting from the model (m*X + c). This error might be resulting due to the other explanatory variables that are influencing the dependent variable. The same model can be written as;

$$Y_i = E\left(Y/X_i\right) + u_i$$

$$E\left(Y/X_i\right) = m * X_i + c$$

This is nothing but the conditional mean of Y denoted as a function of the explanatory variable X. This is also known as the population regression function. Where, the LHS in the first equation denotes the average value of the dependent variable at a given value of the independent variable. The u_i, the value taken by the error term U, is a random variable, and the expectation of this random variable for a given X_i (the conditional expectation) is zero.

This can be proved easily by taking the expectation on $$Y_i = E\left(Y/X_i\right) + u_i$$

$$E\left(\frac{Y_i}{X_i}\right) = E\left(E\left(Y/X_i\right)\right) + E\left(\frac{u_i}{X_i}\right)$$

$$E\left(\frac{Y_i}{X_i}\right) = E\left(Y/X_i\right) + E\left(\frac{u_i}{X_i}\right)$$

$E\left(\frac{Y_i}{X_i}\right)$ and $E\left(Y/X_i\right)$ are same and it implies that the conditional expectation of the random variable U is zero.

In reality regression models are always developed based on the sample data. Similar to the population regression function, based on the sampled data, a sample regression function may be formulated as shown below;

$$\hat{Y}_i = \hat{m} * X_i + \hat{c} + \hat{u}_i$$

Left hand side of the above regression function denotes the estimated conditional expectation of the dependent variable. $\hat{m}, \hat{c}$ are the estimators for the parameters of the population regression function or the statistics. $\hat{u}_i$ is the estimator for u_i. Values taken by the estimators are different for different samples. Using the sample data it is required to find the estimates for the population parameters and it is also necessary to explain how close the estimates are to the population parameters.

Regression Analysis II

But, in the OLS nothing can be inferred about the point estimators of the two parameters used in the above regression model. To have some idea on the properties of these estimators in the classical linear regression model certain assumptions are made. They are,

1. The regression model is linear in parameters.
2. Values taken by the explanatory variable, X, are fixed under repeated sampling. The error term and the explanatory variable are independent.
3. Mean value of the error term is zero.
4. Variance of the error term is same and it doesn't depend on the value of the explanatory variable. Normally it is denoted with the symbol σ^2. This is also known as homoscedasticity.
5. There is no autocorrelation between the explanatory variable and the error term.

6. The sample size used in estimating the parameters must be greater than the number of parameters to be estimated.
7. There must be some finite variability in the explanatory variable.

But to test the significance of these estimators and to draw the inferences about the population data from which the sample has been taken it is necessary to make certain assumptions about the probability density function of the error term. In general it is assumed that the error term follows normal distribution with mean zero and the variance σ^2. From this assumption and the central limit theorem it can be said that the dependent variable Y is normally distributed with mean (m*X + c) and the variance σ^2.

Regression Analysis III

It is necessary to know whether the estimates obtained using the OLS method are close to the population parameters. In other words, it is necessary to know the precision of these estimates. It is also necessary to know the overall suitability of the sample regression function in representing the population data that are being studied. Since the $\widehat{m}$ and the $\hat{c}$ are estimated from the sample data it can be said that these two are the statistics. Two know the precision of these estimates it is necessary to know their sampling distributions and the corresponding parameters. Based on the assumptions on the error term, the parameters of the sampling distribution are as shown below;

$$E[\widehat{m}] = m$$

$$Var(\widehat{m}) = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}$$

$$Standard\ error\ of\ \widehat{m} = \frac{\sigma}{\sqrt{\Sigma(x_i - \bar{x})^2}}$$

$$E[\hat{c}] = c$$

$$Var(\hat{c}) = \frac{\Sigma x_i^2}{n\,\Sigma(x_i - \bar{x})^2} * \sigma^2$$

$$Standard\ error\ of\ \hat{c} = \sqrt{\frac{\Sigma x_i^2}{n\,\Sigma(x_i - \bar{x})^2}} * \sigma$$

Since the value of σ is not known an estimate of the same can be obtained as given below;

$$\hat{\sigma}^2 = \frac{\sum(\hat{u}_i)^2}{n-2}$$

Where, $\hat{u}_i$ is the estimated error equals to $y_i - \hat{y}_i$; n-2 is the degrees of freedom. $\hat{\sigma}$ is known as the standard error of the regression. Once the sampling distributions are known, it is easy to get the confidence intervals for the estimated parameters. It is also possible to formulate the hypotheses on the parameters to further know the suitability of the corresponding variables.

UNIT II: Research Design

Meaning of research design

The formidable problem that follows the task of defining the research problem is the preparation of the design of the research project, popularly known as the "research design". Decisions regarding what, where, when, how much, by what means concerning an inquiry or
a research study constitute a research design. "A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure. "In fact, the research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to the final analysis of data.

More explicitly, the design decisions happen to be in respect of:
(i) What is the study about?
(ii) Why is the study being made?
(iii) Where will the study be carried out?
(iv) What type of data is required?
(v) Where can the required data be found?
(vi) What periods of time will the study include?
(vii) What will be the sample design?
(viii) What techniques of data collection will be used?
(ix) How will the data be analysed?
(x) In what style will the report be prepared?

Keeping in view the above stated design decisions; one may split the overall research design into the following parts:
(a) *The sampling design* which deals with the method of selecting items to be observed for the given study;
(b) T*he observational design* which relates to the conditions under which the observations are to be made;
(c) *The statistical design* which concerns with the question of how many items are to be observed and how the information and data gathered are to be analysed; and
(d) *The operational design* which deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

Thus, the important features of a research design are as below:

(i) It is a plan that specifies the sources and types of information relevant to the research problem.
(ii) It is a strategy specifying which approach will be used for gathering and analysing the data.
(iii) It also includes the time and cost budgets since most studies are done under these two constraints.

In brief, research design must, at least, contain—(a) a clear statement of the research problem;(b) procedures and techniques to be used for gathering information; (c) the population to be studied; and (d) methods to be used in processing and analysing data.

Features of a good design

A good design is often characterized by adjectives like flexible, appropriate, efficient, economical and so on. Generally, the design which minimizes bias and maximizes the reliability of the data collected and analyzed is considered a good design. The design which gives the smallest experimental error is supposed to be the best design in many investigations. Similarly, a design which yields maximal information and provides an opportunity for considering many different aspects of a problem is considered most appropriate and efficient design in respect of many research problems. Thus, the question of good design is related to the purpose or objective of the research problem and also with the nature of the problem to be studied. A design may be quite suitable in one case, but may be found wanting in one respect or the other in the context of some other research problem. One single design cannot serve the purpose of all types of research problems.

A research design appropriate for a particular research problem, usually involves the consideration of the following factors:
(i) the means of obtaining information;
(ii) the availability and skills of the researcher and his staff, if any;
(iii) the objective of the problem to be studied;
(iv) the nature of the problem to be studied; and
(v) the availability of time and money for the research work.
If the research study happens to be an exploratory or a formulative one, wherein the major emphasis is on discovery of ideas and insights, the research design most appropriate must be flexible enough to permit the consideration of many different aspects of a phenomenon. But when the purpose of a study is accurate description of a situation or of an association between variables (or in what are called the descriptive studies), accuracy becomes a major consideration and a research design

which minimizes bias and maximizes the reliability of the evidence collected is considered a good design.

Studies involving the testing of a hypothesis of a causal relationship between variables require a design which will permit inferences about causality in addition to the minimization of bias and maximization of reliability. But in practice it is the most difficult task to put a particular study in a particular group, for a given research may have in it elements of two or more of the functions of different studies. It is only on the basis of its primary function that a study can be categorized either as an exploratory or descriptive or hypothesis-testing study and accordingly the choice of a research design may be made in case of a particular study. Besides, the availability of time, money, skills of the research staff and the means of obtaining the information must be given due weightage while working out the relevant details of the research design such as experimental design, survey design, sample design and the like.

Types of Research Design

There are different types of research designs and be broadly categorised into 3 types:

- Exploratory Research Design
- Descriptive and Diagnostic Research Design
- Hypothesis-Testing Research Design.

Exploratory Research Design: The Exploratory Research Design is known as formulative research design. The main objective of using such a research design is to formulate a research problem for an in-depth or more precise investigation, or for developing a working hypothesis from an operational aspect. The major purpose of such studies is the discovery of ideas and insights. Therefore, such a research design suitable for such a study should be flexible enough to provide opportunity for considering different dimensions of the problem under study. The in-built flexibility in research design is required as the 20 initial research problem would be transformed into a more precise one in the exploratory study, which in turn may necessitate changes in the research procedure for collecting relevant data. Usually, the following three methods are considered in the context of a research design for such studies. They are (a) a survey of related literature; (b) experience survey; and (c) analysis of 'insight-stimulating' instances.

Descriptive And Diagnostic Research Design: A Descriptive Research Design is concerned with describing the characteristics of a particular individual or a group. Meanwhile, a diagnostic research design determines the frequency with which a variable occurs or its relationship with another variable. In other words, the study analysing whether a certain variable is associated with another comprises a diagnostic research study. On the other hand, a study that is concerned with specific predictions or with the narration of facts and characteristics related to an individual, group or situation, are instances of descriptive research studies. Generally, most of the social research design falls under this category. As a research design, both the descriptive and diagnostic studies share common requirements, hence they are grouped together. However, the procedure to be used and the research design needed is to be planned carefully. The research design must also make appropriate provision for protection against bias and thus maximize reliability, with due regard to the completion of the research study in an economical manner. The research design in such studies should be rigid and not flexible. Besides, it must also focus attention on the following: a) Formulation of the objectives of the study, b) Proper designing of the methods of data collection, c) Sample selection, d) Data collection, e) Processing and analysis of the collected data, and f) Reporting the findings.

Hypothesis-Testing Research Design: Hypothesis-Testing Research Designs are those in which the researcher tests the hypothesis of causal relationship between two or more variables. These studies require procedures that would not only decrease bias and enhance reliability, but also facilitate deriving inferences about the causality. Generally, experiments satisfy such requirements. Hence, when research design is discussed in such studies, it often refers to the design of experiments.

Sampling Design

All items in any field of inquiry constitute a 'Universe' or 'Population.' A complete enumeration of all items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry, when all items are covered, no element of chance is left and highest accuracy is obtained. But in practice this may not be true. Even the slightest element of bias in such an inquiry will get larger and larger as the number of observation increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy.

Therefore, when the field of inquiry is large, this method becomes difficult to adopt because of the resources involved. At times, this method is practically beyond the reach of ordinary researchers. Perhaps, government is the only institution which can get the complete enumeration carried out. Even the government adopts this in very rare cases such as population census conducted once in a decade. Further, many a time it is not possible to examine every item in the population, and sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population. In such cases there is no utility of census surveys.

However, it needs to be emphasized that when the universe is a small one, it is no use resorting to a sample survey. When field studies are undertaken in practical life, considerations of time and cost almost invariably lead to a selection of respondents i.e., selection of only a few items. The respondents selected should be as representative of the total population as possible in order to produce a miniature cross-section. The selected respondents constitute what is technically called a 'sample' and the selection process is called 'sampling technique.' The survey so conducted is known as 'sample survey'. Algebraically, let the population size be N and if a part of size n (which is < N) of this population is selected according to some rule for studying some characteristic of the population, the group consisting of these n units is known as 'sample'. Researcher must prepare a sample design for his study i.e., he must plan how a sample should be selected and of what size such a sample would be.

Implications of a sample design

A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample. Sample design may as well lay down the number of items to be included in the sample i.e., the size of the sample. Sample design is determined before data are collected. There are many sample designs from which a researcher can choose. Some designs are relatively more precise and easier to apply than others. Researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

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 program, 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 is 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 dictate the size of sample that we can draw as well. 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. Systematic bias results 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.

Sampling errors are the random variations in the sample estimates around the true population parameters. Since they occur randomly and are equally likely to be in either direction, their nature happens to be of compensatory type and the expected value of such errors happens to be equal to zero. Sampling error decreases with the increase in the size of the sample, and it happens to be of a smaller magnitude in case of homogeneous population.

Sampling error can be measured for a given sample design and size. The measurement of sampling error is usually called the 'precision of the sampling plan'. If we increase the sample size, the precision can be improved. But increasing the size of the sample has its own limitations viz., a large sized sample increases the cost of collecting data and also enhances the systematic bias. Thus the effective way to increase precision is usually to select a better sampling design which has a smaller sampling error for a given sample size at a given cost. In practice, however, people prefer a less precise design because it is easier to adopt the same and also because of the fact that systematic bias can be controlled in a better way in such a design.

In brief, while selecting a sampling procedure, researcher must ensure that the procedure causes a relatively small sampling error and helps to control the systematic bias in a better way.

Characteristics of a good sample design

From what has been stated above, we can list down the characteristics of a good sample design as under:

- Sample design must result in a truly representative sample.
- Sample design must be such which results in a small sampling error.
- Sample design must be viable in the context of funds available for the research study.
- Sample design must be such so that systematic bias can be controlled in a better way.
- Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

Different types of sample designs

There are different types of sample designs based on two factors viz., the representation basis and the element selection technique. On the representation basis, the sample may be probability sampling or it may be non-probability sampling. Probability sampling is based on the concept of random selection, whereas non-probability sampling is 'non-random' sampling. On element selection basis, the sample may be either unrestricted or restricted. When each sample element is drawn individually from the population at large, then the sample so drawn is known as 'unrestricted sample', whereas all other forms of sampling are covered under the term 'restricted sampling'.

Thus, sample designs are basically of two types' viz., non-probability sampling and probability sampling. We take up these two designs separately.

Non-probability sampling

Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by different names such as deliberate sampling, purposive sampling and judgment sampling. In this type of sampling, items for the sample are selected deliberately by the researcher; his choice concerning the items remains supreme. In other words, under non-probability sampling the organizers of the inquiry purposively choose the particular units of the universe for constituting a sample on the basis that the small mass that they so select out of a huge one will be typical or representative of the whole. For

instance, if the economic conditions of people living in a state are to be studied, a few towns and villages may be selected on purpose for intensive study, on the principle that they can represent the entire state. Thus, the judgment of the organizers of the study plays an important part in this sampling design.

In such a design, personal element has a great chance of entering into the selection of the sample. The investigator may select a sample which shall yield results favorable to his point of view and if that happens, the entire inquiry may get vitiated. Thus, there is always the danger of bias entering into this type of sampling technique. But in the investigators are impartial, work without bias and have the necessary experience so as to take sound judgment , the results obtained from an analysis of deliberately selected sample may be tolerably reliable. However, in such a sampling, there is no assurance that every element has some specifiable chance of being included. Sampling error in this type of sampling cannot be estimated and the element of bias, great or small, is always there. As such this sampling design in rarely adopted in large inquires of importance. However, in small inquiries and researches by individuals, this design may be adopted because of the relative advantage of time and money inherent in this method of sampling. *Quota sampling* is also an example of non-probability sampling. Under quota sampling the interviewers are simply given quotas to be filled from the different strata, with some restrictions on how they are to be filled. In other words, the actual selection of the items for the sample is left to the interviewer's discretion. This type of sampling is very convenient and is relatively inexpensive. But the samples so selected certainly do not possess the characteristic of random samples. Quota samples are essentially judgment samples and inferences drawn on their basis are not amenable to statistical treatment in a formal way.

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. It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process. Here it is blind chance alone that determines whether one item or the other is selected. The results obtained from probability or random sampling can be assured in terms of probability i.e., we can measure the errors of estimation or the significance of results obtained from a random sample, and this fact brings out the superiority of random sampling design over the deliberate sampling design. Random sampling ensures the law of Statistical Regularity which states that if on an average the

sample chosen is a random one, the sample will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the best technique of selecting a representative sample.

Random sampling from a finite population refers to that method of sample selection which gives each possible sample combination an equal probability of being picked up and each item in the entire population to have an equal chance of being included in the sample. This applies to sampling without replacement i.e., once an item is selected for the sample, it cannot appear in the sample again (Sampling with replacement is used less frequently in which procedure the element selected for the sample is returned to the population before the next element is selected. In such a situation the same element could appear twice in the same sample before the second element is chosen). In brief, the implications of random sampling (or simple random sampling) are:

(a) It gives each element in the population an equal probability of getting into the sample; and all choices are independent of one another.

(b) It gives each possible sample combination an equal probability of being chosen. Keeping this in view we can define a simple random sample (or simply a random sample) from a finite population as a sample which is chosen in such a way that each of the $_NC_n$ possible samples has the same probability, $1/_NC_n$, of being selected. To make it more clear we take a certain finite population consisting of six elements (say a, b, c, d, e, f) i.e., N = 6. Suppose that we want to take a sample of size n = 3 from it. Then there are $_6C_3$ = 20 possible distinct samples of the required size, and they consist of the elements *abc*, *abd*, *abe*, *abf*, *acd*, *ace*, *acf*, *ade*, *adf*, *aef*, *bcd*, *bce*, *bcf*, *bde*,*bdf*, *bef*, *cde*, *cdf*, *cef*, and *def*. If we choose one of these samples in such a way that each has the probability 1/20 of being chosen, we will then call this a random sample.

Other sampling designs are as follows :

Deliberate sampling*:*
Deliberate sampling is also known as purposive or non-probability sampling. This sampling method involves purposive or deliberate selection of particular units of the universe for constituting a sample which represents the universe. When population elements are selected for inclusion in the sample based on the ease of access, it can be called *convenience sampling*. If a researcher wishes to secure data from, say, gasoline buyers, he may select a fixed number of petrol stations and may

conduct interviews at these stations. This would be an example of convenience sample of gasoline buyers. At times such a procedure may give very biased results particularly when the population is not homogeneous. On the other hand, in *judgment sampling* the researcher's judgment is used for selecting items which he considers as representative of the population. For example, a judgment sample of college students might be taken to secure reactions to a new method of teaching. Judgment sampling is used quite frequently in qualitative research where the desire happens to be to develop hypotheses rather than to generalize to larger populations.

Simple random sampling:

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

Systematic sampling:

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

Stratified sampling

If the population from which a sample is to be drawn does not constitute a homogeneous group, then stratified sampling technique is applied so as to obtain a representative sample. In this technique, the population is stratified into a number of non-overlapping subpopulations or strata and sample items are selected from each stratum. If the items selected from each stratum is based on simple random sampling the entire procedure, first stratification and then simple random sampling, is known as *stratified random sampling*.

Quota sampling

In stratified sampling the cost of taking random samples from individual strata is often so expensive that interviewers are simply given quota to be filled from different strata, the actual selection of items for sample being left to the interviewer's judgment. This is called quota sampling. The size of the quota for each stratum is generally proportionate to the size of that stratum in the population. Quota sampling is thus an important form of non-probability sampling. Quota samples generally happen to be judgment samples rather than random samples.

Cluster sampling and area sampling

Cluster sampling involves grouping the population and then selecting the groups or the clusters rather than individual elements for inclusion in the sample. Suppose some departmental store wishes to sample its credit card holders. It has issued its cards to 15,000 customers. The sample size is to be kept say 450. For cluster sampling this list of 15,000 card holders could be formed into 100 clusters of 150 card holders each. Three clusters might then be selected for the sample randomly. The sample size must often be larger than the simple random sample to ensure the same level of accuracy because is cluster sampling procedural potential for order bias and other sources of error are usually accentuated. The clustering approach can, however, make the sampling procedure relatively easier and increase the efficiency of field work, especially in the case of personal interviews. *Area sampling* is quite close to cluster sampling and is often talked about when the total geographical area of interest happens to be big one. Under area sampling we first divide the total area into a number of smaller non-overlapping areas, generally called geographical clusters, then a number of these smaller areas are randomly selected, and all units in these small areas are included in the sample. Area sampling is especially helpful where we do not have the list of the population concerned. It also makes the field interviewing more efficient since interviewer can do many interviews at each location.

Multi-stage sampling

This is a further development of the idea of cluster sampling. This technique is meant for big inquiries extending to a considerably large geographical area like an entire country. Under multi-stage sampling the first stage may be to select large primary sampling units such as states, then districts, then towns and finally certain families within towns. If the technique of random-sampling is applied at all stages, the sampling procedure is described as multi-stage random sampling.

Sequential sampling

This is somewhat a complex sample design where the ultimate size of the sample is not fixed in advance but is determined according to mathematical decisions on the basis of information yielded as survey progresses. This design is usually adopted under acceptance sampling plan in the context of statistical quality control. In practice, several of the methods of sampling described above may well be used in the same study in which case it can be called mixed sampling. It may be pointed out here that normally one should resort to random sampling so that bias can be eliminated and sampling error can be estimated. But purposive sampling is considered desirable when the universe happens to be small and a known characteristic of it is to be studied intensively. Also, there are conditions under which sample designs other than random sampling may be considered better for reasons like convenience and low costs.

The sample design to be used must be decided by the researcher taking into consideration the nature of the inquiry and other related factors.

UNIT III : MEASUREMENT & DATA COLLECTION

Measurement in research

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, especially 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.

It is easy to assign numbers in respect of properties of some objects, but it is relatively difficult in respect of others. For instance, measuring such things as social conformity, intelligence, or marital adjustment is much less obvious and requires much closer attention than measuring physical weight, biological age or a person's financial assets. In other words, properties like weight, height, etc., can be measured directly with some standard unit of measurement, but it is not that easy to measure properties like motivation to succeed, ability to stand stress and the like. We can expect high accuracy in measuring the length of pipe with a yard stick, but if the concept is abstract and the measurement tools are not standardized, we are less confident about the accuracy of the results of measurement.

Technically speaking, measurement is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence. In measuring, we devise some form of scale in the range (in terms of set theory, range may refer to some set) and then transform or map the properties of objects from the domain (in terms of set theory, domain may refer to some other set) onto this scale. For example, in case we are to find the male to female attendance ratio while conducting a study of persons who attend some show, then we may tabulate those who come to the show according to sex. In terms of set theory, this process is one of mapping the observed physical properties of those coming to the show (the domain) on to a sex classification (the range). The rule of correspondence is: If the object in the domain appears to be male, assign to "0" and if female assign to "1". Similarly, we can record a person's marital status as 1, 2, 3 or 4, depending on whether the

person is single, married, widowed or divorced. We can as well record "Yes or No" answers to a question as "0" and "1" (or as 1 and 2 or perhaps as 59 and 60). In this artificial or nominal way, categorical data (qualitative or descriptive) can be made into numerical data and if we thus code the various categories, we refer to the numbers we record as nominal data. *Nominal data* are numerical in name only, because they do not share any of the properties of the numbers we deal in ordinary arithmetic. For instance if we record marital status as 1, 2, 3, or 4 as stated above, we cannot write 4 > 2 or 3 < 4 and we cannot write 3 − 1 = 4 − 2, 1 + 3 = 4 or 4 ⬚⬚2 = 2. In those situations when we cannot do anything except set up inequalities, we refer to the data as *ordinal data.* For instance, if one mineral can scratch another, it receives a higher hardness number and on Mohs' scale the numbers from 1 to 10 are assigned respectively to talc, gypsum, calcite, fluorite, apatite, feldspar, quartz, topaz, sapphire and diamond. With these numbers we can write 5 > 2 or 6 < 9 as apatite is harder than gypsum and feldspar is softer than sapphire, but we cannot write for example 10 − 9 = 5 − 4, because the difference in hardness between diamond and sapphire is actually much greater than that between apatite and fluorite. It would also be meaningless to say that topaz is twice as hard as fluorite simply because their respective hardness numbers on Mohs' scale are 8 and 4. The greater than symbol (i.e., >) in connection with ordinal data may be used to designate "happier than" "preferred to" and so on.

When in addition to setting up inequalities we can also form differences, we refer to the data as *interval data.* Suppose we are given the following temperature readings (in degrees Fahrenheit): 58°, 63°, 70°, 95°, 110°, 126° and 135°. In this case, we can write 100° > 70° or 95° < 135° which simply means that 110° is warmer than 70° and that 95° is cooler than 135°. We can also write for example 95° − 70° = 135° − 110°, since equal temperature differences are equal in the sense that the same amount of heat is required to raise the temperature of an object from 70° to 95° or from 110° to 135°. On the other hand, it would not mean much if we said that 126° is twice as hot as 63°, even though 126° ⬚⬚63° = 2. To show the reason, we have only to change to the centigrade scale, where the first temperature becomes 5/9 (126 − 32) = 52°, the second temperature becomes 5/9 (63 − 32) = 17° and the first figure is now more than three times the second. This difficulty arises from the fact that Fahrenheit and Centigrade scales both have artificial origins (zeros) i.e., the number 0 of neither scale is indicative of the absence of whatever quantity we are trying to measure. When in addition to setting up inequalities and forming differences we can also form

quotients (i.e., when we can perform all the customary operations of mathematics), we refer to such data as *ratio data.* In this sense, ratio data includes all the usual measurement (or determinations) of length, height, money amounts, weight, volume, area, pressures etc.

The above stated distinction between nominal, ordinal, interval and ratio data is important for the nature of a set of data may suggest the use of particular statistical techniques*. A researcher has to be quite alert about this aspect while measuring properties of objects or of abstract concepts.

* When data can be measured in units which are interchangeable e.g., weights (by ratio scales), temperatures (by interval scales), that data is said to be parametric and can be subjected to most kinds of statistical and mathematical processes. But when data is measured in units which are not interchangeable, e.g., product preferences (by ordinal scales), the data is said to be non-parametric and is susceptible only to a limited extent to mathematical and statistical treatment.

Measurement scales

From what has been stated above, we can write that scales of measurement can be considered in terms of their mathematical properties. The most widely used classification of measurement scales are: (a) nominal scale; (b) ordinal scale; (c) interval scale; and (d) ratio scale.

Nominal scale: Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of 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.

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.

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.

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 behavioral 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.

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.

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

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

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

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

Meaning of Scaling

In research we quite often face measurement problem (since we want a valid measurement but may not obtain it), especially when the concepts to be measured are complex and abstract and we do not possess the standardized measurement tools. Alternatively, we can say that while measuring attitudes and opinions, we face the problem of their valid measurement. Similar problem may be faced by a researcher, of course in a lesser degree, while measuring physical or institutional concepts. As such we should study some procedures which may enable us to measure abstract concepts more accurately. This brings us to the study of scaling techniques.

Scaling describes the procedures of assigning numbers to various degrees of opinion, attitude and other concepts. This can be done in two ways viz - (i) making a judgment about some characteristic of an individual and then placing him directly on a scale that has been defined in terms of that characteristic and (ii) constructing questionnaires in such a way that the score of individual's responses assigns him a place on a scale. It may be stated here that a scale is a continuum, consisting of the highest point (in terms of some characteristic e.g., preference, favorable factors, etc.) and the lowest point along with several intermediate points between these two extreme points. These scale-point positions are so related to each other that when the first point happens to be the highest point, the second point indicates a higher degree in terms of a given characteristic as compared to the third point and the third point indicates a higher degree as compared to the fourth and so on. Numbers for measuring the distinctions of degree in the attitudes/opinions are, thus, assigned to individuals corresponding to their scale-positions. All this is better understood when we talk about scaling technique(s). Hence the term 'scaling' is applied to the procedures for attempting to determine quantitative measures of subjective abstract concepts. Scaling has been defined as a "procedure for the assignment of numbers (or other symbols) to a property of objects in order to impart some of the characteristics of numbers to the properties in question.

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.

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.

There are several ways of collecting primary data. The most used methods are as below:

Observation Method:

In observation method, the information is sought by way of investigator's own direct observation without asking from the respondent. The main advantage of this method is that it is free from subjective biasness, as it is free from respondent's willingness. It is, however, an expensive and time consuming method. Moreover, the information provided by this method is very limited and some of the more busy people like executives may not be accessible to direct observation.

Interview Method:

Primary data may be collected either through personal interviews or through telephonic interviews:

(a) In the personal interviews the interviewer asks questions generally in a face to face contact. Through interview method more and reliable information may be obtained. Personal information can be obtained easily under this method. It is, however, a very expensive and time consuming method, especially when large and widely spread geographical sample is taken. Certain types of respondents, such as officials, executives or people of high income groups, may not be easily accessible. In this method, the respondent may give wrong and imaginary information. For effective interview there should be a good rapport with respondents which is often very difficult to develop. For a good result the interviewer's approach should be friendly, courteous, conversational and unbiased for which a proper training is required.

(b) In telephonic interviews contact is made with the respondents through telephone.

The main merits of telephonic interviews are:

(i) It is more flexible and faster than other methods

(ii) It is cheaper and less time consuming.

(iii) Recall is easy and replies can be recorded without causing embarrassment to respondents.

(iv) At times, access can be made to respondents who otherwise cannot be contacted for one reason or the other.

(v) No staff is required and wider representation of sample is possible.

This method also has several drawbacks. For example, the surveys are restricted to respondents who have telephonic facilities and little time is given to respondents for considered answers. It is not suitable for intensive surveys where comprehensive answers are required for various questions.

Questionnaire Method:

In this method a questionnaire is mailed to the person concerned with a request to answer the questions and return the questionnaire. This method is most extensively applied in various researches of human and economic geography.

The main merits of this method are given below:

(i) There is low cost even when the universe is large and is widespread geographically.

(ii) It is free from bias of interviewer as answers are respondent's own words.

(iii) Respondents, who are not easily approachable, can also be reached conveniently.

Moreover, respondents are given enough time to give well thought answers.

The main demerits of this method are:

(i) Low rate of return of the duly filled questionnaires.

(ii) It can be used only when the respondents are educated and cooperative.

(iii) The control of the questionnaire may be lost once it is sent.

(iv) It is difficult to know whether willing respondents are truly representative.

(v) This method is likely to be the slowest of all.

Schedule Method:

This method of data collection is very much like the collection of data through questionnaires, with little difference that lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for this purpose. 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. This method 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 organizations. Population census all over the world is conducted through this method. For the selection of an appropriate method, the objective, nature and scope of study, the availability of time and funds, and the level of precision required are to be taken into consideration.

Collection of secondary data

Secondary data means data that are already available i.e., they refer to the data which have already been collected and analyzed by someone else. When the researcher utilizes secondary data, then he has to look into various sources from where he can obtain them. In this case he is certainly not confronted with the problems that are usually associated with the collection of original data. Secondary data may either be published data or unpublished data. Usually published data are available in: (a) various publications of the central, state are local governments; (b) various publications of foreign governments or of international bodies and their

subsidiary organizations; (c) technical and trade journals; (d) books, magazines and newspapers; (e) reports and publications of various associations connected with business and industry, banks, stock exchanges, etc.; (f) reports prepared by research scholars, universities, economists, etc. in different fields; and (g) public records and statistics, historical documents, and other sources of published information. The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labor bureaus and other public/ private individuals and organizations.

Researcher must be very careful in using secondary data. He must make a minute scrutiny because it is just possible that the secondary data may be unsuitable or may be inadequate in the context of the problem which the researcher wants to study.

By way of caution, the researcher, before using secondary data, must see that they possess following characteristics:

- Reliability of data
- Suitability of data
- Adequacy of data

Selection of appropriate method for data collection

There are various methods of data collection. As such the researcher must judiciously select the method/methods for his own study, keeping in view the following factors:

1. Nature, scope and object of enquiry: This constitutes the most important factor affecting the choice of a particular method. The method selected should be such that it suits the type of enquiry that is to be conducted by the researcher. This factor is also important in deciding whether the data already available (secondary data) are to be used or the data not yet available (primary data) are to be collected.

2. Availability of funds: Availability of funds for the research project determines to a large extent the method to be used for the collection of data. When funds at the disposal of the researcher are very limited, he will have to select a comparatively cheaper method which may not be as efficient and effective as some other costly method. Finance, in fact, is a big constraint in practice and the researcher has to act within this limitation.

3. Time factor: Availability of time has also to be taken into account in deciding a particular method of data collection. Some methods take relatively more time, whereas with others the data can be collected in a comparatively shorter duration. The time at the disposal of the researcher, thus, affects the selection of the method by which the data are to be collected.

4. Precision required: Precision required is yet another important factor to be considered at the time of selecting the method of collection of data.

But one must always remember that each method of data collection has its uses and none is
superior in all situations. Also over and above all this, much depends upon the ability and experience of the researcher. Dr. A.L. Bowley's remark in this context is very appropriate when he says that "in collection of statistical data common sense is the chief requisite and experience the chief teacher."

Case Study Method

The case study method is a very popular form of qualitative analysis and involves a careful and complete observation of a social unit, be that unit a person, a family, an institution, a cultural group or even the entire community. It is a method of study in depth rather than breadth. The case study 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 behavior-patterns of the given unit as an integrated totality. According to H. Odum, "The case study method is a technique by which individual factor whether it be an institution or just an episode in the life of an individual or a group is analyzed in its relationship to any other in the group."[5] Thus, a fairly exhaustive study of a person (as to what he does and has done, what he thinks he does and had done and what he expects to do and says he ought to do) or group is called a life or case history. Burgess has used the words "the social microscope" for the case study method."

Pauline V. Young describes case study as "a comprehensive study of a social unit be that unit a person, a group, a social institution, a district or a community."[7] In brief, we can say
that case study method is a form of qualitative analysis where in careful and complete observation of an individual or a situation or an institution is done; efforts are made to study each and every aspect of the concerning unit in minute details and then from case data generalizations and inferences are drawn.

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 behavior 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 generalized knowledge to get richer and richer. In its absence, generalized social science may get handicapped.

Evolution and scope

The case study method is a widely used systematic field research technique in sociology these days. The credit for introducing this method to the field of social investigation goes to Frederic Le Play who used it as a hand-maiden to statistics in his studies of family budgets. Herbert Spencer was the first to use case material in his comparative study of different cultures. Dr. William Healy resorted to this method in his study of juvenile delinquency, and considered it as a better method over and above the mere use of statistical data. Similarly, anthropologists, historians, novelists and dramatists have used this method concerning problems pertaining to their areas of interests. Even management experts use case study methods for getting clues to several management problems. In brief, case study method is being used in several disciplines. Not only this, its use is increasing day by day.

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 behavior 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 program 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.

- Being an exhaustive study of a social unit, the case study method enables us to understand fully the behavior 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 behavior directly and not by an indirect and abstract approach."
- 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 behavior.
- 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.
- It helps in formulating relevant hypotheses along with the data which may be helpful in testing them. Case studies, thus, enable the generalized knowledge to get richer and richer.
- 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.

- 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.
- 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.
- 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 organizing data".
- 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.
- 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.
- Case study method enhances the experience of the researcher and this in turn increases his analyzing ability and skill.
- 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.
- 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.
- 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.

- 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."[8] Real information is often not collected because the subjectivity of the researcher does enter in the collection of information in a case study.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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

UNIT IV : DATA PROCESSING AND 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".1 But there are persons (Selltiz, Jahoda and others) who do not like to make difference between processing and analysis. They opine that analysis of data in a general way involves a number of closely related operations which are performed with the purpose of summarising the collected data and organising these in such a manner that they answer the research question(s). We, however, shall prefer to observe the difference between the two terms as stated here in order to understand their implications more clearly.

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

Editing

Editing of data is a process of examining the collected raw data (especially 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. A t 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: 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. 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. They must make entries (if any) on the form in some distinctive colour and that too in a standardised form. They should initial all answers which they change or supply. Editor's initials and the date of editing should be placed on each completed form or schedule.

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.

Classification

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

Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

Classification according to attributes

As stated above, data are classified on the basis of common 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 2n, 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.

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. Classification according to class intervals usually involves the following three main problems:

How may classes should be there? What should be their magnitudes?
There can be no specific answer with regard to the number of classes. The decision about this calls for skill and experience of the researcher. However, the objective should be to display the data in such a way as to make it meaningful for the analyst. Typically, we may have 5 to 15 classes. With regard to the second part of the question, we can say that, to the extent possible, class-intervals should be of equal magnitudes, but in some cases unequal magnitudes may result in better classification. Hence researcher's objective judgment plays an important part in this connection. Multiples of 2, 5 and 10 are generally preferred while determining class magnitudes.

Some statisticians adopt the following formula, suggested by H.A. Sturges, determining the size of class interval:

$$i = R/(1 + 3.3 \log N)$$

where
i = size of class interval;
R = Range (i.e., difference between the values of the largest item and smallest item among the given items);
N = Number of items to be grouped.

It should also be kept in mind that in case one or two or very few items have very high or very low values, one may use what are known as open-ended intervals in the overall frequency distribution. Such intervals may be expressed like under Rs500 or Rs10001 and over. Such intervals are generally not desirable, but often cannot be avoided. The researcher must always remain conscious of this fact while deciding the issue of the total number of class intervals in which the data are to be classified.

How to choose class limits?
While choosing class limits, the researcher must take into consideration the criterion that the mid-point (generally worked out first by taking the sum of the upper limit and lower limit of a class and then divide this sum by 2) of a class-interval and the actual average of items of that class interval should remain as close to each other as possible. Consistent with this, the class limits should be located at multiples of 2, 5, 10, 20, 100 and such other figures. Class limits may generally be stated in any of the following forms:
Exclusive type class intervals:

They are usually stated as follows:
10–20
20–30
30–40
40–50
The above intervals should be read as under:
10 and under 20
20 and under 30
30 and under 40
40 and under 50

Thus, under the exclusive type class intervals, the items whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly 30 would be put in 30–40 class interval and not in 20–30 class interval. In simple words, we can say that under exclusive type class intervals, the upper limit of a class interval is excluded and items with values less than the upper limit (but not less than the lower limit) are put in the given class interval.

Inclusive type class intervals:

They are usually stated as follows:

11–20
21–30
31–40
41–50

In inclusive type class intervals the upper limit of a class interval is also included in the concerning class interval. Thus, an item whose value is 20 will be put in 11–20 class interval. The stated upper limit of the class interval 11–20 is 20 but the real limit is 20.99999 and as such 11–20 class interval really means 11 and under 21. When the phenomenon under consideration happens to be a discrete one (i.e., can be measured and stated only in integers), then we should adopt inclusive type classification. But when the phenomenon happens to be a continuous one capable of being measured in fractions as well, we can use exclusive type class intervals.

How to determine the frequency of each class?

This can be done either by tally sheets or by mechanical aids. Under the technique of tally sheet, the class-groups are written on a sheet of paper (commonly known as the tally sheet) and for each item a stroke (usually a small vertical line) is marked against the class group in which it falls. The general practice is that after every four small vertical lines in a class group, the fifth line for the item falling in the same group, is indicated as horizontal line through the said four lines and the resulting flower (IIII) represents five items. All this facilitates the counting of items in each one of the class groups. An illustrative tally sheet can be shown as under:

An Illustrative Tally Sheet for Determining the Number of 70 Families in Different Income Groups

Income groups (Rupees)	Tally mark	Number of families or (Class frequency)
Below 400	卌 卌 III	13
401–800	卌 卌 卌 卌	20
801–1200	卌 卌 II	12
1201–1600	卌 卌 卌 III	18
1601 and above	卌 II	7
Total		70

Alternatively, class frequencies can be determined, especially in case of large inquires and surveys, by mechanical aids i.e., with the help of machines viz., sorting machines that are available for the purpose. Some machines are hand operated, whereas other work with electricity. There are machines which can sort out cards at a speed of something like 25000 cards per hour. This method is fast but expensive.

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 availability of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation if other factors are favourable and necessary facilities are available. Hand tabulation is usually preferred in case of small inquiries where the number of questionnaires is small and they are of relatively short length. Hand tabulation may be done using the 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.

Generally accepted principles of tabulation: Such principles of tabulation, particularly of constructing statistical tables, can be briefly states as follows:*

1. Every table should have a clear, concise and adequate title so as to make the table intelligible without reference to the text and this title should always be placed just above the body of the table.
2. Every table should be given a distinct number to facilitate easy reference.
3. The column headings (captions) and the row headings (stubs) of the table should be clear and brief.
4. The units of measurement under each heading or sub-heading must always be indicated.
5. Explanatory footnotes, if any, concerning the table should be placed directly beneath the table, along with the reference symbols used in the table.
6. Source or sources from where the data in the table have been obtained must be indicated just below the table.
7. Usually the columns are separated from one another by lines which make the table more readable and attractive. Lines are always drawn at the top and bottom of the table and below the captions.
8. There should be thick lines to separate the data under one class from the data under another class and the lines separating the sub-divisions of the classes should be comparatively thin lines.
9. The columns may be numbered to facilitate reference.
10. Those columns whose data are to be compared should be kept side by side. Similarly, percentages and/or averages must also be kept close to the data.
11. It is generally considered better to approximate figures before tabulation as the same would reduce unnecessary details in the table itself.
12. In order to emphasise the relative significance of certain categories, different kinds of type, spacing and indentations may be used.
13. It is important that all column figures be properly aligned. Decimal points and (+) or (−) signs should be in perfect alignment.
14. Abbreviations should be avoided to the extent possible and ditto marks should not be used in the table.
15. Miscellaneous and exceptional items, if any, should be usually placed in the last row of the table.
16. Table should be made as logical, clear, accurate and simple as possible. If the data happen to be very large, they should not be crowded in a single table for that would make the table unwieldy and inconvenient.
17. Total of rows should normally be placed in the extreme right column and that of columns should be placed at the bottom.
18. The arrangement of the categories in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate

comparison. Above all, the table must suit the needs and requirements of an investigation.

Some problems in processing

We can take up the following two problems of processing the data for analytical purposes:

The problem concerning "Don't know" (or DK) responses: While processing the data, the researcher often comes across some responses that are difficult to handle. One category of such responses may be 'Don't Know Response' or simply DK response. When the DK response group is small, it is of little significance. But when it is relatively big, it becomes a matter of major concern in which case the question arises: Is the question which elicited DK response useless? The answer depends on two points viz., the respondent actually may not know the answer or the researcher may fail in obtaining the appropriate information. In the first case the concerned question is said to be alright and DK response is taken as legitimate DK response. But in the second case, DK response is more likely to be a failure of the questioning process.

How DK responses are to be dealt with by researchers? The best way is to design better type of questions. Good rapport of interviewers with respondents will result in minimising DK responses. But what about the DK responses that have already taken place? One way to tackle this issue is to estimate the allocation of DK answers from other data in the questionnaire. The other way is to keep DK responses as a separate category in tabulation where we can consider it as a separate reply category if DK responses happen to be legitimate, otherwise we should let the reader make his own decision. Yet another way is to assume that DK responses occur more or less randomly and as such we may distribute them among the other answers in the ratio in which the latter have occurred. Similar results will be achieved if all DK replies are excluded from tabulation and that too without inflating the actual number of other responses.

Use or percentages: Percentages are often used in data presentation for they simplify numbers, reducing all of them to a 0 to 100 range. Through the use of percentages, the data are reduced in the standard form with base equal to 100

which fact facilitates relative comparisons. While using percentages, the following rules should be kept in view by researchers:

1. Two or more percentages must not be averaged unless each is weighted by the group size from which it has been derived.
2. Use of too large percentages should be avoided, since a large percentage is difficult to understand and tends to confuse, defeating the very purpose for which percentages are used.
3. Percentages hide the base from which they have been computed. If this is not kept in view, the real differences may not be correctly read.
4. Percentage decreases can never exceed 100 per cent and as such for calculating the percentage of decrease, the higher figure should invariably be taken as the base.
5. Percentages should generally be worked out in the direction of the causal-factor in case of two-dimension tables and for this purpose we must select the more significant factor out of the two given factors as the causal factor.

Elements/types of analysis

As stated earlier, by analysis we mean the computation of certain indices or measures along with searching for patterns of relationship that exist among the data groups. Analysis, particularly in case of survey or experimental data, involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences. Analysis may, therefore, be categorised as descriptive analysis and inferential analysis (Inferential analysis is often known as statistical analysis). "Descriptive analysis is largely the study of distributions of one variable. This study provides us with profiles of companies, work groups, persons and other subjects on any of a multiple of characteristics such as size. Composition, efficiency, preferences, etc."2. this sort of analysis may be in respect of one variable (described as unidimensional analysis), or in respect of two variables (described as bivariate analysis) or in respect of more than two variables (described as multivariate analysis). In this context we work out various measures that show the size and shape of a distribution(s) along with the study of measuring relationships between two or more variables.

We may as well talk of correlation analysis and causal analysis. Correlation analysis studies the joint variation of two or more variables for determining the amount of

correlation between two or more variables. Causal analysis is concerned with the study of how one or more variables affect changes in another variable. It is thus a study of functional relationships existing between two or more variables. This analysis can be termed as regression analysis. Causal analysis is considered relatively more important in experimental researches, whereas in most social and business researches our interest lies in understanding and controlling relationships between variables then with determining causes per se and as such we consider correlation analysis as relatively more important.

In modern times, with the availability of computer facilities, there has been a rapid development of multivariate analysis which may be defined as "all statistical methods which simultaneously analyse more than two variables on a sample of observations". Usually the following analyses are involved when we make a reference of multivariate analysis:

- Multiple regression analysis: This analysis is adopted when the researcher has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about the dependent variable based on its covariance with all the concerned independent variables.
- Multiple discriminant analysis: This analysis is appropriate when the researcher has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute. The object of this analysis happens to be to predict an entity's possibility of belonging to a particular group based on several predictor variables.
- Multivariate analysis of variance (or multi-ANOVA): This analysis is an extension of two way ANOVA, wherein the ratio of among group variance to within group variance is worked out on a set of variables.
- Canonical analysis: This analysis can be used in case of both measurable and non-measurable variables for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.
- Inferential analysis: is concerned with the various tests of significance for testing hypotheses in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

Statistics in research

The role of statistics in research is to function as a tool in designing research, analysing its data and drawing conclusions from them. Most research studies result in a large volume of raw data which must be suitably reduced so that the same can be read easily and can be used for further analysis. Clearly the science of statistics cannot be ignored by any research worker, even though he may not have occasion to use statistical methods in all their details and ramifications. Classification and tabulation, as stated earlier, achieve this objective to some extent, but we have to go a step further and develop certain indices or measures to summarise the collected/classified data. Only after this we can adopt the process of generalisation from small groups (i.e., samples) to population. If fact, there are two major areas of statistics viz., descriptive statistics and inferential statistics. Descriptive statistics concern the development of certain indices from the raw data, whereas inferential statistics concern with the process of generalisation. Inferential statistics are also known as sampling statistics and are mainly concerned with two major type of problems:

- the estimation of population parameters, and
- the testing of statistical hypotheses.

The important statistical measures* that are used to summarise the survey/research data are:

- ➢ measures of central tendency or statistical averages;
- ➢ measures of dispersion;
- ➢ measures of asymmetry (skewness);
- ➢ measures of relationship; and
- ➢ other measures.

Amongst the measures of central tendency, the three most important ones are the arithmetic average or mean, median and mode. Geometric mean and harmonic mean are also sometimes used.

From among the measures of dispersion, variance, and its square root—the standard deviation are the most often used measures. Other measures such as mean deviation, range, etc. are also used. For comparison purpose, we use mostly the coefficient of standard deviation or the coefficient of variation.

In respect of the measures of skewness and kurtosis, we mostly use the first measure of skewness based on mean and mode or on mean and median. Other measures of skewness, based on quartiles or on the methods of moments, are also used sometimes. Kurtosis is also used to measure the peakedness of the curve of the frequency distribution.

Amongst the measures of relationship, Karl Pearson's coefficient of correlation is the frequently used measure in case of statistics of variables, whereas Yule's coefficient of association is used in case of statistics of attributes. Multiple correlation coefficient, partial correlation coefficient, regression analysis, etc., are other important measures often used by a researcher.

Index numbers, analysis of time series, coefficient of contingency, etc., are other measures that may as well be used by a researcher, depending upon the nature of the problem under study.

We give below a brief outline of some important measures (out of the above listed measures) often used in the context of research studies.

Measures of central tendency

Measures of central tendency (or statistical averages) tell us the point about which items have a tendency to cluster. Such a measure is considered as the most representative figure for the entire mass of data. Measure of central tendency is also known as statistical average. Mean, median and mode are the most popular averages. Mean, also known as arithmetic average, is the most common measure of central tendency and may be defined as the value which we get by dividing the total of the values of various given items in a series by the total number of items. we can work it out as under:

$$\text{Mean (or } \overline{X})^* = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

where $\overline{X}$ = The symbol we use for mean (pronounced as X bar)

$\sum$ = Symbol for summation

X_i = Value of the ith item X, $i = 1, 2, \ldots, n$

n = total number of items

In case of a frequency distribution, we can work out mean in this way:

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{f_1 X_1 + f_2 X_2 + \ldots + f_n X_n}{f_1 + f_2 + \ldots + f_n = n}$$

Sometimes, instead of calculating the simple mean, as stated above, we may workout the weighted mean for a realistic average. The weighted mean can be worked out as follows:

$$\overline{X}_w = \frac{\sum w_i X_i}{\sum w_i}$$

where $\overline{X}_w$ = Weighted item

w_i = weight of ith item X

X_i = value of the ith item X

Mean is the simplest measurement of central tendency and is a widely used measure. Its chief use consists in summarising the essential features of a series and in enabling data to be compared. It is amenable to algebraic treatment and is used in further statistical calculations. It is a relatively stable measure of central tendency. But it suffers from some limitations viz., it is unduly affected by extreme items; it may not coincide with the actual value of an item in a series, and it may lead to wrong impressions, particularly when the item values are not given with the average. However, mean is better than other averages, especially in economic and social studies where direct quantitative measurements are possible.

Median is the value of the middle item of series when it is arranged in ascending or descending order of magnitude. It divides the series into two halves; in one half all items are less than median, whereas in the other half all items have values higher than median. If the values of the items arranged in the ascending order are: 60, 74, 80, 90, 95, 100, then the value of the 4th item viz., 88 is the value of median. We can also write thus:

$$\text{Median } (M) = \text{Value of } \left(\frac{n+1}{2}\right)\text{th item}$$

Median is a positional average and is used only in the context of qualitative phenomena, for example, in estimating intelligence, etc., which are often encountered in sociological fields. Median is not useful where items need to be

assigned relative importance and weights. It is not frequently used in sampling statistics.

Mode is the most commonly or frequently occurring value in a series. The mode in a distribution is that item around which there is maximum concentration. In general, mode is the size of the item which has the maximum frequency, but at items such an item may not be mode on account of the effect of the frequencies of the neighbouring items. Like median, mode is a positional average and is not affected by the values of extreme items. it is, therefore, useful in all situations where we want to eliminate the effect of extreme variations. Mode is particularly useful in the study of popular sizes. For example, a manufacturer of shoes is usually interested in finding out the size most in demand so that he may manufacture a larger quantity of that size. In other words, he wants a modal size to be determined for median or mean size would not serve his purpose. But there are certain limitations of mode as well. For example, it is not amenable to algebraic treatment and sometimes remains indeterminate when we have two or more model values in a series. It is considered unsuitable in cases where we want to give relative importance to items under consideration.

Geometric mean is also useful under certain conditions. It is defined as the nth root of the product of the values of n times in a given series. Symbolically, we can put it thus:

$$\text{Geometric mean (or G.M.)} = \sqrt[n]{\pi X_i}$$

$$= \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \dots X_n}$$

where

        G.M. = geometric mean,

        n = number of items.

        X_i = ith value of the variable X

        π = conventional product notation

For instance, the geometric mean of the numbers, 4, 6, and 9 is worked out as

$$\text{G.M.} = \sqrt[3]{4.6.9}$$

$$= 6$$

Harmonic mean is of limited application, particularly in cases where time and rate are involved. The harmonic mean gives largest weight to the smallest item and smallest weight to the largest item. As such it is used in cases like time and motion

study where time is variable and distance constant. From what has been stated above, we can say that there are several types of statistical averages. Researcher has to make a choice for some average. There are no hard and fast rules for the selection of a particular average in statistical analysis for the selection of an average mostly depends on the nature, type of objectives of the research study. One particular type of average cannot be taken as appropriate for all types of studies. The chief characteristics and the limitations of the various averages must be kept in view; discriminate use of average is very essential for sound statistical analysis.

Measures of dispersion

Averages can represent a series only as best as a single figure can, but it certainly cannot reveal the entire story of any phenomenon under study. Especially it fails to give any idea about the scatter of the values of items of a variable in the series around the true value of average. In order to measure this scatter, statistical devices called measures of dispersion are calculated. Important measures of dispersion are

> ➢ range,
> ➢ mean deviation, and
> ➢ Standard deviation.

Range is the simplest possible measure of dispersion and is defined as the difference between the values of the extreme items of a series. Thus,

Range=Highest value of an item in a series - Lowest value of an item in a series

The utility of range is that it gives an idea of the variability very quickly, but the drawback is that range is affected very greatly by fluctuations of sampling. Its value is never stable, being based on only two values of the variable. As such, range is mostly used as a rough measure of variability and is not considered as an appropriate measure in serious research studies.

Mean deviation is the average of difference of the values of items from some average of the series. Such a difference is technically described as deviation. In calculating mean deviation we ignore the minus sign of deviations while taking their total for obtaining the mean deviation. Mean deviation is, thus, obtained as under:

$$\text{Mean deviation from mean } (\delta_{\bar{x}}) = \frac{\Sigma |X_i - \bar{X}|}{n}, \quad \text{if deviations, } |X_i - \bar{X}|, \text{ are obtained from}$$

or   arithmetic average.

$$\text{Mean deviation from median } (\delta_m) = \frac{\Sigma |X_i - M|}{n}, \quad \text{if deviations, } |X_i - M|, \text{ are obtained}$$

or   from median

$$\text{Mean deviation from mode } (\delta_z) = \frac{\Sigma |X_i - Z|}{n}, \quad \text{if deviations, } |X_i - Z|, \text{ are obtained from}$$

mode.

where δ = Symbol for mean deviation (pronounced as delta);

X_i = ith values of the variable X;

n = number of items;

$\bar{X}$ = Arithmetic average;

M = Median;

Z = Mode.

When mean deviation is divided by the average used in finding out the mean deviation itself, the resulting quantity is described as the coefficient of mean deviation. Coefficient of mean deviation is a relative measure of dispersion and is comparable to similar measure of other series. Mean deviation and its coefficient are used in statistical studies for judging the variability, and thereby render the study of central tendency of a series more precise by throwing light on the typical average. It is a better measure of variability than range as it takes into consideration the values of all items of a series. Even then it is not a frequently used measure as it is not amenable to algebraic process.

Standard deviation is most widely used measure of dispersion of a series and is commonly denoted by the symbol ' s ' (pronounced as sigma). Standard deviation is defined as the square-root of the average of squares of deviations, when such deviations for the values of individual items in a series are obtained from the arithmetic average. It is worked out as under:

$$\text{Standard deviation}^* (\sigma) = \sqrt{\frac{\Sigma (X_i - \bar{X})^2}{n}}$$

When we divide the standard deviation by the arithmetic average of the series, the resulting quantity is known as coefficient of standard deviation which happens to be a relative measure and is often used for comparing with similar measure of other series. When this coefficient of standard deviation is multiplied by 100, the resulting

figure is known as coefficient of variation. Sometimes, we work out the square of standard deviation, known as variance, which is frequently used in the context of analysis of variation.

The standard deviation (along with several related measures like variance, coefficient of variation, etc.) is used mostly in research studies and is regarded as a very satisfactory measure of dispersion in a series. It is amenable to mathematical manipulation because the algebraic signs are not ignored in its calculation (as we ignore in case of mean deviation). It is less affected by fluctuations of sampling. These advantages make standard deviation and its coefficient a very popular measure of the scatteredness of a series. It is popularly used in the context of estimation and testing of hypotheses.
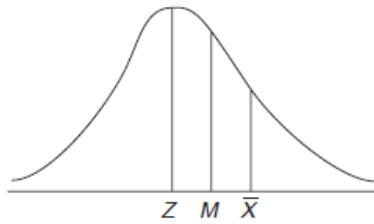
Measures of asymmetry (skewness)

When the distribution of item in a series happens to be perfectly symmetrical, we then have the following type of curve for the distribution:

$(\bar{X} = M = Z)$
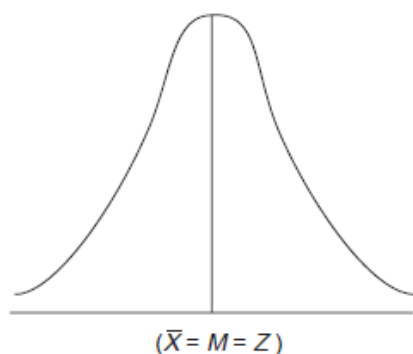
Curve showing no skewness in which case we have $\bar{X} = M = Z$

Such a curve is technically described as a normal curve and the relating distribution as normal distribution. Such a curve is perfectly bell shaped curve in which case the value of X or M or Z is just the same and skewness is altogether absent. But if the curve is distorted (whether on the right side or on the left side), we have asymmetrical distribution which indicates that there is skewness. If the curve is distorted on the right side, we have positive skewness but when the curve is distorted towards left, we have negative skewness as shown here under:

Curve showing positive skewness
In case of positive skewness we have:
$Z < M < \overline{X}$
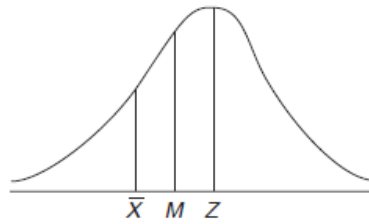
Curve showing negative skewness
In case of negative skewness we have:
$\overline{X} < M < Z$

Skewness is, thus, a measure of asymmetry and shows the manner in which the items are clustered around the average. In a symmetrical distribution, the items show a perfect balance on either side of the mode, but in a skew distribution the balance is thrown to one side. The amount by which the balance exceeds on one side measures the skewness of the series. The difference between the mean, median or the mode provides an easy way of expressing skewness in a series. In case of positive skewness, we have Z < M < X and in case of negative skewness we have X < M < Z. Usually we measure skewness in this way:

Skewness = X – Z and its coefficient (j) is worked

In case Z is not well defined, then we work out skewness as under:

Skewness = 3(X – M) and its coefficient (j) is worked

Skewness = $\overline{X} - Z$ and its coefficient (j) is worked

out as $j = \dfrac{\overline{X} - Z}{\sigma}$

In case Z is not well defined, then we work out skewness as under:

Skewness = $3(\overline{X} - M)$ and its coefficient (j) is worked

out as $j = \dfrac{3(\overline{X} - M)}{\sigma}$

The significance of skewness lies in the fact that through it one can study the formation of series and can have the idea about the shape of the curve, whether normal or otherwise, when the items of a given series are plotted on a graph. Kurtosis is the measure of flat-top of a curve. A bell shaped curve or the normal curve is Mesokurtic because it is kurtic in the centre; but if the curve is relatively more peaked than the normal curve, it is called Leptokurtic whereas a curve is more flat than the normal curve, it is called Platykurtic. In brief, Kurtosis is the humped

nature of the curve and points to the nature of distribution of items in the middle of a series.

It may be pointed out here that knowing the shape of the distribution curve is crucial to the use of statistical methods in research analysis since most methods make specific assumptions about the nature of the distribution curve.

Measures of relationship

So far we have dealt with those statistical measures that we use in context of univariate population i.e., the population consisting of measurement of only one variable. But if we have the data on two variables, we are said to have a bivariate population and if the data happen to be on more than two variables, the population is known as multivariate population. If for every measurement of a variable, X, we have corresponding value of a second variable, Y, the resulting pairs of values are called a bivariate population. In addition, we may also have a corresponding value of the third variable, Z, or the forth variable, W, and so on; the resulting pairs of values are called a multivariate population. In case of bivariate or multivariate populations, we often wish to know the relation of the two and/or more variables in the data to one another. We may like to know, for example, whether the number of hours students devote for studies is somehow related to their family income, to age, to sex or to similar other factor. There are several methods of determining the relationship between variables, but no method can tell us for certain that a correlation is indicative of causal relationship. Thus we have to answer two types of questions in bivariate or multivariate populations viz.,

1. Does there exist association or correlation between the two (or more) variables? If yes, of what degree?
2. Is there any cause and effect relationship between the two variables in case of the bivariate population or between one variable on one side and two or more variables on the other side in case of multivariate population? If yes, of what degree and in which direction?

The first question is answered by the use of correlation technique and the second question by the technique of regression. There are several methods of applying the

two techniques, but the important ones are as under:

In case of bivariate population: Correlation can be studied through cross tabulation, Charles Spearman's coefficient of correlation, Karl Pearson's coefficient of correlation; whereas cause and effect relationship can be studied through simple regression equations.

In case of multivariate population: Correlation can be studied through coefficient of multiple correlation, coefficient of partial correlation; whereas cause and effect relationship can be studied through multiple regression equations.

We can now briefly take up the above methods one by one.

Cross tabulation approach is especially useful when the data are in nominal form. Under it we classify each variable into two or more categories and then cross classify the variables in these subcategories. Then we look for interactions between them which may be symmetrical, reciprocal or asymmetrical. A symmetrical relationship is one in which the two variables vary together, but we assume that neither variable is due to the other. A reciprocal relationship exists when the two variables mutually influence or reinforce each other. Asymmetrical relationship is said to exist if one variable (the independent variable) is responsible for another variable (the dependent variable). The cross classification procedure begins with a two-way table which indicates whether there is or there is not an interrelationship between the variables. This sort of analysis can be further elaborated in which case a third factor is introduced into the association through cross-classifying the three variables. By doing so we find conditional relationship in which factor X appears to affect factor Y only when factor Z is held constant. The correlation, if any, found through this approach is not considered a very powerful form of statistical correlation and accordingly we use some other methods when data happen to be either ordinal or interval or ratio data.

Charles Spearman's coefficient of correlation (or rank correlation) is the technique of determining the degree of correlation between two variables in case of ordinal data where ranks are given to the different values of the variables. The main objective of this coefficient is to determine the extent to which the two sets of ranking are similar or dissimilar. This coefficient is determined as under:

$$\text{Spearman's coefficient of correlation (or } r_s) = 1 - \left[\frac{6\sum d_i^2}{n(n^2 - 1)}\right]$$

Where di = difference between ranks of ith pair of the two variables;

n = number of pairs of observations.

As rank correlation is a non-parametric technique for measuring relationship between paired observations of two variables when data are in the ranked form, we have dealt with this technique in greater details later on in the book in chapter entitled 'Hypotheses Testing II (Non-parametric tests)'.

Karl Pearson's coefficient of correlation (or simple correlation) is the most widely used method of measuring the degree of relationship between two variables. This coefficient assumes the following:

1. that there is linear relationship between the two variables;
2. that the two variables are casually related which means that one of the variables is independent and the other one is dependent; and
3. a large number of independent causes are operating in both variables so as to produce a normal distribution.

Karl Pearson's coefficient of correlation can be worked out thus.

$$\text{Karl Pearson's coefficient of correlation (or } r)^* = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{n \cdot \sigma_X \cdot \sigma_Y}$$

This is the short cut approach for finding 'r' in case of ungrouped data. If the data happen to be grouped data (i.e., the case of bivariate frequency distribution), we shall have to write Karl Pearson's coefficient of correlation as under:

where fij is the frequency of a particular cell in the correlation table and all other values are defined as earlier.

where X_i = *i*th value of X variable

$\overline{X}$ = mean of X

Y_i = *i*th value of Y variable

$\overline{Y}$ = Mean of Y

n = number of pairs of observations of X and Y

σ_X = Standard deviation of X

σ_Y = Standard deviation of Y

In case we use assumed means (A_x and A_y for variables X and Y respectively) in place of true means, then Karl Person's formula is reduced to:

$$\frac{\dfrac{\sum dx_i \cdot dy_i}{n} - \left(\dfrac{\sum dx_i \cdot \sum dy_i}{n}\right)}{\sqrt{\dfrac{\sum dx_i^2}{n} - \left(\dfrac{\sum dx_i}{n}\right)^2} \sqrt{\dfrac{\sum dy_i^2}{n} - \left(\dfrac{\sum dy_i}{n}\right)^2}}$$

$$\frac{\dfrac{\sum dx_i \cdot dy_i}{n} - \left(\dfrac{\sum dx_i \cdot \sum dy_i}{n}\right)}{\sqrt{\dfrac{\sum dx_i^2}{n} - \left(\dfrac{\sum dx_i}{n}\right)^2} \sqrt{\dfrac{\sum dy_i^2}{n} - \left(\dfrac{\sum dy_i}{n}\right)^2}}$$

where $\sum dx_i = \sum(X_i - A_x)$

$\sum dy_i = \sum(Y_i - A_y)$

$\sum dx_i^2 = \sum(X_i - A_x)^2$

$\sum dy_i^2 = \sum(Y_i - A_y)^2$

$\sum dx_i \cdot dy_i = \sum(X_i - A_x)(Y_i - A_y)$

n = number of pairs of observations of X and Y.

Karl Pearson's coefficient of correlation is also known as the product moment correlation coefficient. The value of 'r' lies between ± 1. Positive values of r indicate positive correlation between the two variables (i.e., changes in both variables take place in the statement direction), whereas negative values of 'r' indicate negative correlation i.e., changes in the two variables taking place in the opposite directions. A zero value of 'r' indicates that there is no association between the two variables. When r = (+) 1, it indicates perfect positive correlation and when it is (−)1, it indicates perfect negative correlation, meaning thereby that variations in independent variable (X) explain 100% of the variations in the dependent variable (Y). We can also say that for a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction, then correlation will be termed as perfect positive. But if such change occurs in the opposite direction, the correlation will be termed as perfect negative. The value of 'r' nearer to +1 or −1 indicates high degree of correlation between the two variables.

Simple regression analysis

Regression is the determination of a statistical relationship between two or more variables. In simple regression, we have only two variables, one variable (defined as independent) is the cause of the behaviour of another one (defined as dependent variable). Regression can only interpret what exists physically i.e., there must be a physical way in which independent variable X can affect dependent variable Y. The basic relationship between X and Y is given by

Y = a + bX

where the symbol Y denotes the estimated value of Y for a given value of X. This equation is known as the regression equation of Y on X (also represents the regression line of Y on X when drawn on a graph) which means that each unit change in X produces a change of b in Y, which is positive for direct and negative for inverse relationships.

Then generally used method to find the 'best' fit that a straight line of this kind can give is the least-square method. To use it efficiently, we first determine

$$\sum x_i^2 = \sum X_i^2 - n\overline{X}^2$$

$$\sum y_i^2 = \sum Y_i^2 - n\overline{Y}^2$$

$$\sum x_i y_i = \sum X_i Y_i - n\overline{X} \cdot \overline{Y}$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2}, \quad a = \overline{Y} - b\overline{X}$$

Thus, the regression analysis is a statistical method to deal with the formulation of mathematical model depicting relationship amongst variables which can be used for the purpose of prediction of the values of dependent variable, given the values of the independent variable.

[Alternatively, for fitting a regression equation of the type $ Y = a + bX to the given values of X and Y variables, we can find the values of the two constants viz., a and b by using the following two normal equations:

$$\sum Y_i = na + b\sum X_i$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$$

and then solving these equations for finding a and b values. Once these values are obtained and have been put in the equation $ Y = a + bX, we say that we have fitted the regression equation of Y on X to the given data. In a similar fashion, we can develop the regression equation of X and Y viz

Multiple correlation and regression

When there are two or more than two independent variables, the analysis concerning relationship is known as multiple correlations and the equation describing such relationship as the multiple regression equation. We here explain multiple correlation and regression taking only two independent variables and one dependent variable (Convenient computer programs exist for dealing with a great number of variables). In this situation the results are interpreted as shown below: Multiple regression equation assumes the form

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

where X1 and X2 are two independent variables and Y being the dependent variable, and the constants a, b1 and b2 can be solved by solving the following three normal equations:

$$\sum Y_i = na + b_1 \sum X_{1i} + b_2 \sum X_{2i}$$
$$\sum X_{1i} Y_i = a \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i}$$
$$\sum X_{2i} Y_i = a \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2$$

(It may be noted that the number of normal equations would depend upon the number of independent variables. If there are 2 independent variables, then 3 equations, if there are 3 independent variables then 4 equations and so on, are used.)

In multiple regression analysis, the regression coefficients (viz., b1 b2) become less reliable as the degree of correlation between the independent variables (viz., X1, X2) increases. If there is a high degree of correlation between independent variables, we have a problem of what is commonly described as the problem of multicollinearity. In such a situation we should use only one set of the independent variable to make our estimate. In fact, adding a second variable, say X2, that is correlated with the first variable, say X1, distorts the values of the regression coefficients. Nevertheless,

the prediction for the dependent variable can be made even when multicollinearity is present, but in such a situation enough care should be taken in selecting the independent variables to estimate a dependent variable so as to ensure that multi-collinearity is reduced to the minimum. With more than one independent variable, we may make a difference between the collective effect of the two independent variables and the individual effect of each of them taken separately. The collective effect is given by the coefficient of multiple correlation,

$R_{y \cdot x_1 x_2}$ defined as under:

$$R_{y \cdot x_1 x_2} = \sqrt{\frac{b_1 \sum Y_i X_{1i} - n\bar{Y}\bar{X}_1 + b_2 \sum Y_i X_{2i} - n\bar{Y}\bar{X}_2}{\sum Y_i^2 - n\bar{Y}^2}}$$

Alternatively, we can write

$$R_{y \cdot x_1 x_2} = \sqrt{\frac{b_1 \sum x_{1i} y_i + b_2 \sum x_{2i} y_i}{\sum Y_i^2}}$$

where

$$x_{1i} = (X_{1i} - \bar{X}_1)$$
$$x_{2i} = (X_{2i} - \bar{X}_2)$$
$$y_i = (Y_i - \bar{Y})$$

and b1 and b2 are the regression coefficients.

Partial correlation

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 as stated earlier. In the case of two independent variables, we shall have two partial correlation coefficients denoted ryx1 ×x2 and ryx x 2 × 1 which are worked out as under:

$$r_{yx_1 \cdot x_2} = \frac{R^2_{y \cdot x_1 x_2} - r^2_{yx_2}}{1 - r^2_{yx_2}}$$

This measures the effort of X1 on Y, more precisely, that proportion of the variation of Y not explained by X2 which is explained by X1. Also,

$$r_{yx_2 \cdot x_1} = \frac{R^2_{y \cdot x_1 x_2} - r^2_{yx_1}}{1 - r^2_{yx_1}}$$

These formulae of the alternative approach are based on simple coefficients of correlation (also known as zero order coefficients since no variable is held constant when simple correlation coefficients are worked out). The partial correlation coefficients are called first order coefficients when one variable is held constant as shown above; they are known as second order coefficients when two variables are held constant and so on.

Association in case of attributes

When data is collected on the basis of some attribute or attributes, we have statistics commonly termed as statistics of attributes. It is not necessary that the objects may process only one attribute; rather it would be found that the objects possess more than one attribute. In such a situation our interest may remain in knowing whether the attributes are associated with each other or not. For example, among a group of people we may find that some of them are inoculated against small-pox and among the inoculated we may observe that some of them suffered from small-pox after inoculation.

The important question which may arise for the observation is regarding the efficiency of inoculation for its popularity will depend upon the immunity which it provides against small-pox. In other words, we may be interested in knowing whether inoculation and immunity from small-pox are associated.

Technically, we say that the two attributes are associated if they appear together in a greater number of cases than is to be expected if they are independent and not simply on the basis that they are appearing together in a number of cases as is done in ordinary life. The association may be positive or negative (negative association is also known as disassociation). If class frequency of AB, symbolically written as (AB), is greater than the expectation of AB being together if they are independent, then we say the two attributes are positively associated; but if the class frequency of AB is

less than this expectation, the two attributes are said to be negatively associated. In case the class frequency of AB is equal to expectation, the two attributes are considered as independent i.e., are said to have no association. It can be put symbolically as shown hereunder:

Where (AB) = frequency of class AB and

$\frac{(A)}{N} \times \frac{(B)}{N} \times N$ = Expectation of AB, if A and B are independent, and N being the number of items

In order to find out the degree or intensity of association between two or more sets of attributes, we should work out the coefficient of association. Professor Yule's coefficient of association is most popular and is often used for the purpose. It can be mentioned as under:

$$Q_{AB} = \frac{(AB)(ab) - (Ab)(aB)}{(AB)(ab) + (Ab)(aB)}$$

where,

Q_{AB} = Yule's coefficient of association between attributes A and B.
(AB) = Frequency of class AB in which A and B are present.
(Ab) = Frequency of class Ab in which A is present but B is absent.
(aB) = Frequency of class aB in which A is absent but B is present.
(ab) = Frequency of class ab in which both A and B are absent.

The value of this coefficient will be somewhere between +1 and –1. If the attributes are completely associated (perfect positive association) with each other, the coefficient will be +1, and if they are completely disassociated (perfect negative association), the coefficient will be –1. If the attributes are completely independent of each other, the coefficient of association will be 0. The varying degrees of the coefficients of association are to be read and understood according to their positive and negative nature between +1 and –1.

Sometimes the association between two attributes, A and B, may be regarded as unwarranted when we find that the observed association between A and B is due to the association of both A and B with another attribute C. For example, we may observe positive association between inoculation and exemption for small-pox, but such association may be the result of the fact that there is positive association between inoculation and richer section of society and also that there is positive association between exemption from small-pox and richer section of society.

The sort of association between A and B in the population of C is described as partial association as distinguished from total association between A and B in the overall universe. We can work out the coefficient of partial association between A

and B in the population of C by just modifying the above stated formula for finding association between A and B as shown below:

$$Q_{AB.C} = \frac{(ABC)(abC) - (AbC)(aBC)}{(ABC)(abC) + (AbC)(aBC)}$$

where,

$Q_{AB.C}$ = Coefficient of partial association between A and B in the population of C; and all other values are the class frequencies of the respective classes (A, B, C denotes the presence of concerning attributes and a, b, c denotes the absence of concerning attributes).

At times, we may come across cases of illusory association, wherein association between two attributes does not correspond to any real relationship. This sort of association may be the result of some attribute, say C with which attributes A and B are associated (but in reality there is no association between A and B). Such association may also be the result of the fact that the at=tributes A and B might not have been properly defined or might not have been correctly recorded. Researcher must remain alert and must not conclude association between A and B when in fact there is no such association in reality.

In order to judge the significance of association between two attributes, we make use of Chi square test by finding the value of Chi-square (c2) and using Chi-square distribution the value of c2 can be worked out as under:

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad\qquad i = 1, 2, 3 \ldots$$

$$j = 1, 2, 3 \ldots$$

where

O_{ij} = observed frequencies

E_{ij} = expected frequencies.

Association between two attributes in case of manifold classification and the resulting contingency table can be studied as explained below:

We can have manifold classification of the two attributes in which case each of the two attributes are first observed and then each one is classified into two or more subclasses, resulting into what is called as contingency table. The following is an

example of 4 × 4 contingency table with two attributes A and B, each one of which has been further classified into four sub-categories.

		Attribute A				
		A_1	A_2	A_3	A_4	Total
	B_1	$(A_1 B_1)$	$(A_2 B_1)$	$(A_3 B_1)$	$(A_4 B_1)$	(B_1)
Attribute B	B_2	$(A_1 B_2)$	$(A_2 B_2)$	$(A_3 B_2)$	$(A_4 B_2)$	(B_2)
	B_3	$(A_1 B_3)$	$(A_2 B_3)$	$(A_3 B_3)$	$(A_4 B_3)$	(B_3)
	B_4	$(A_1 B_4)$	$(A_2 B_4)$	$(A_3 B_4)$	$(A_4 B_4)$	(B_4)
	Total	(A_1)	(A_2)	(A_3)	(A_4)	N

Association can be studied in a contingency table through Yule's coefficient of association as stated above, but for this purpose we have to reduce the contingency table into 2 × 2 table by combining some classes. For instance, if we combine (A1) + (A2) to form (A) and (A3) + (A4) to form (a) and similarly if we combine (B1) + (B2) to form (B) and (B3) + (B4) to form (b) in the above contingency table, then we can write the table in the form of a 2 × 2 table as shown in Table below

		Attribute		
		A	a	Total
Attribute	B b	(AB) (Ab)	(aB) (ab)	(B) (b)
	Total	(A)	(a)	N

After reducing a contingency table in a two-by-two table through the process of combining some classes, we can work out the association as explained above. But the practice of combining classes is not considered very correct and at times it is inconvenient also, Karl Pearson has suggested a measure known as Coefficient of mean square contingency for studying association in contingency tables. This can be obtained as under:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

where

C = Coefficient of contingency

χ^2 = Chi-square value which is $= \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$

N = number of items.

This is considered a satisfactory measure of studying association in contingency tables.

Other measures

1. **Index numbers:** When series are expressed in same units, we can use averages for the purpose of comparison, but when the units in which two or more series are expressed happen to be different, statistical averages cannot be used to compare them. In such situations we have to rely upon some relative measurement which consists in reducing the figures to a common base. Once such method is to convert the series into a series of index numbers. This is done when we express the given figures as percentages of some specific figure on a certain data. We can, thus, define an index number as a number which is used to measure the level of a given phenomenon as compared to the level of the same phenomenon at some standard date. The use of index number weights more as a special type of average, meant to study the changes in the effect of such factors which are incapable of being measured directly. But one must always remember that index numbers measure only the relative changes.

    Changes in various economic and social phenomena can be measured and compared through index numbers. Different indices serve different purposes. Specific commodity indices are to serve as a measure of changes in the phenomenon of that commodity only. Index numbers may measure cost of living of different classes of people. In economic sphere, index numbers are often termed as 'economic barometers measuring the economic phenomenon in all its aspects either directly by measuring the same phenomenon or indirectly by measuring something else which reflects upon the main phenomenon.

    But index numbers have their own limitations with which researcher must always keep himself aware. For instance, index numbers are only approximate indicators and as such give only a fair idea of changes but cannot give an accurate idea. Chances of error also remain at one point or the other while constructing an index number but this does not diminish the

utility of index numbers for they still can indicate the trend of the phenomenon being measured. However, to avoid fallacious conclusions, index numbers prepared for one purpose should not be used for other purposes or for the same purpose at other places.

2. **Time series analysis:** In the context of economic and business researches, we may obtain quite often data relating to some time period concerning a given phenomenon. Such data is labelled as 'Time Series'. More clearly it can be stated that series of successive observations of the given phenomenon over a period of time are referred to as time series. Such series are usually the result of the effects of one or more of the following factors:

   1. Secular trend or long term trend that shows the direction of the series in a long period of time. The effect of trend (whether it happens to be a growth factor or a decline factor) is gradual, but extends more or less consistently throughout the entire period of time under consideration. Sometimes, secular trend is simply stated as trend (or T).

   2. Short time oscillations i.e., changes taking place in the short period of time only and such changes can be the effect of the following factors:
      - Cyclical fluctuations (or C) are the fluctuations as a result of business cycles and are generally referred to as long term movements that represent consistently recurring rises and declines in an activity.
      - Seasonal fluctuations (or S) are of short duration occurring in a regular sequence at specific intervals of time. Such fluctuations are the result of changing seasons. Usually these fluctuations involve patterns of change within a year that end to be repeated from year to year. Cyclical fluctuations and seasonal fluctuations taken together constitute short-period regular fluctuations.
      - Irregular fluctuations (or I), also known as Random fluctuations, are variations which take place in a completely unpredictable fashion.

All these factors stated above are termed as components of time series and when we try to analyse time series, we try to isolate and measure the effects of various types of these factors on a series. To study the effect of one type of factor, the other type of factor is eliminated from the series. The given series is, thus, left with the effects of one type of factor only.

For analysing time series, we usually have two models; multiplicative model; and

additive model. Multiplicative model assumes that the various components interact in a multiplicative manner to produce the given values of the overall time series and can be stated as under:

$Y = T \times C \times S \times I$

where

Y = observed values of time series, T = Trend, C = Cyclical fluctuations, S = Seasonal fluctuations,

I = Irregular fluctuations.

Additive model considers the total of various components resulting in the given values of the overall time series and can be stated as:

$Y = T + C + S + I$

There are various methods of isolating trend from the given series viz., the free hand method, semiaverage method, method of moving averages, method of least squares and similarly there are methods of measuring cyclical and seasonal variations and whatever variations are left over are considered as random or irregular fluctuations.

The analysis of time series is done to understand the dynamic conditions for achieving the shortterm and long-term goals of business firm(s). The past trends can be used to evaluate the success or failure of management policy or policies practiced hitherto. On the basis of past trends, the future patterns can be predicted and policy or policies may accordingly be formulated. We can as well study properly the effects of factors causing changes in the short period of time only, once we have eliminated the effects of trend. By studying cyclical variations, we can keep in view the impact of cyclical changes while formulating various policies to make them as realistic as possible. The knowledge of seasonal variations will be of great help to us in taking decisions regarding inventory, production, purchases and sales policies so as to optimize working results. Thus, analysis of time series is important in context of long term as well as short term forecasting and is considered a very powerful tool in the hands of business analysts and researchers.

UNIT V: INTEPRETATION & 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:
1. 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.

2. 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.

3. 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.

4. 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.

Precautions in interpretation

One should always remember that even if the data are properly collected and analyzed, wrong interpretation would lead to inaccurate conclusions. It is, therefore, absolutely essential that the task of interpretation be accomplished with patience in an impartial manner and also in correct perspective.

Researcher must pay attention to the following points for correct interpretation:

o At the outset, researcher must invariably satisfy himself that (a) the data are appropriate, trustworthy and adequate for drawing inferences; (b) the data reflect good homogeneity; and that (c) proper analysis has been done through statistical methods.

o The researcher must remain cautious about the errors that can possibly arise in the process of interpreting results. Errors can arise due to false generalization and/or due to wrong interpretation of statistical measures, such as the application of findings beyond the range of observations, identification of correlation with causation and the like. Another major pitfall is the tendency to affirm that definite relationships exist on the basis of confirmation of particular hypotheses. In fact, the positive test results accepting the hypothesis must be interpreted as "being in accord" with the hypothesis, rather than as "confirming the validity of the hypothesis". The researcher must remain vigilant about all such things so that false generalization

may not take place. He should be well equipped with and must know the correct use of statistical measures for drawing inferences concerning his study.

- o He must always keep in view that the task of interpretation is very much intertwined with analysis and cannot be distinctly separated. As such he must take the task of interpretation as a special aspect of analysis and accordingly must take all those precautions that one usually observes while going through the process of analysis viz., precautions concerning the reliability of data, computational checks, validation and comparison of results.
- o He must never lose sight of the fact that his task is not only to make sensitive observations of relevant occurrences, but also to identify and disengage the factors that are initially hidden to the eye. This will enable him to do his job of interpretation on proper lines. Broad generalization should be avoided as most research is not amenable to it because the coverage may be restricted to a particular time, a particular area and particular conditions. Such restrictions, if any, must invariably be specified and the results must be framed within their limits.
- o The researcher must remember that "ideally in the course of a research study, there should be constant interaction between initial hypothesis, empirical observation and theoretical conceptions. It is exactly in this area of interaction between theoretical orientation and empirical observation that opportunities for originality and creativity lie."[2] He must pay special attention to this aspect while engaged in the task of interpretation.

Significance of report writing

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented 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 favor 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.

Different steps in writing report

Research reports are the product of slow, painstaking, accurate inductive work. The usual steps involved in writing report are: (a) logical analysis of the subject-matter; (b) preparation of the final outline; (c) preparation of the rough draft; (d) rewriting and polishing; (c) preparation of the final bibliography; and (f) writing the final draft. Though all these steps are self-explanatory, yet a brief mention of each one of these will be appropriate for better understanding.

Logical analysis of the subject matter: It is the first step which is primarily concerned with the development of a subject. There are two ways in which to develop a subject (a) logically and (b) chronologically. The logical development is made on the basis of mental connections and associations between the one thing and another by means of analysis. Logical treatment often consists in developing the material from the simple possible to the most complex structures. Chronological development is based on a connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order.

Preparation of the final outline: It is the next step in writing the research report "Outlines are the framework upon which long written works are constructed. They are an aid to the logical organization of the material and a reminder of the points to be stressed in the report."[3]

Preparation of the rough draft: This follows the logical analysis of the subject and the preparation of the final outline. Such a step is of utmost importance for the researcher now sits to write down what he has done in the context of his research study. He will write down the procedure adopted by him in collecting the material for his study along with various limitations faced by him, the technique of analysis adopted by him, the broad findings and generalizations and the various suggestions he wants to offer regarding the problem concerned.

Rewriting and polishing of the rough draft: This step happens to be most difficult part of all formal writing. Usually this step requires more time than the writing of the rough draft. The careful revision makes the difference between a mediocre and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation. The researcher should also "see whether or not the material, as it is presented, has unity and cohesion; does the report stand upright and firm and exhibit a definite pattern, like a marble arch? Or does it resemble an old wall of moldering cement and loose brick."[4] In addition the researcher should give due attention to the fact that in his rough draft he has been

consistent or not. He should check the mechanics of writing—grammar, spelling and usage.

Preparation of the final bibliography: Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended to the research report, is a list of books in some way pertinent to the research which has been done. It should contain all those works which the researcher has consulted. The bibliography should be arranged alphabetically and may be divided into two parts; the first part may contain the names of books and pamphlets, and the second part may contain the names of magazine and newspaper articles. Generally, this pattern of bibliography is considered convenient and satisfactory from the point of view of reader, though it is not the only way of presenting bibliography. The entries in bibliography should be made adopting the following order:

For books and pamphlets the order may be as under:

1. Name of author, last name first.
2. Title, underlined to indicate italics.
3. Place, publisher, and date of publication.
4. Number of volumes.

For magazines and newspapers the order may be as under:
1. Name of the author, last name first.
2. Title of article, in quotation marks.
3. Name of periodical, underlined to indicate italics.
4. The volume or volume and number.
5. The date of the issue.
6. The pagination.
The only thing important is that, whatever method one selects, it must remain consistent.

Writing the final draft: This constitutes the last step. The final draft should be written in a concise and objective style and in simple language, avoiding vague expressions such as "it seems", "there may be", and the like ones. While writing the final draft, the researcher must avoid abstract terminology and technical jargon. Illustrations and examples based on common experiences must be incorporated in the final draft as they happen to be most effective in communicating the research findings to others. A research report should not be dull, but must enthuse people and maintain interest and must show originality. It must be remembered that every report should be an attempt to solve some intellectual problem and must contribute

to the solution of a problem and must add to the knowledge of both the researcher and the reader.

Layout of the research report

Anybody, who is reading the research report, must necessarily be conveyed enough about the study so that he can place it in its general scientific context, judge the adequacy of its methods and thus form an opinion of how seriously the findings are to be taken. For this purpose there is the need of proper layout of the report. The layout of the report means as to what the research report should contain. A comprehensive layout of the research report should comprise (A) preliminary pages; (B) the main text; and (C) the end matter. Let us deal with them separately.

Preliminary Pages

In its preliminary pages the report should carry a *title and date,* followed by acknowledgements in the form of 'Preface' or 'Foreword'. Then there should be a *table of contents* followed by *list of tables and illustrations* so that the decision-maker or anybody interested in reading the report can easily locate the required information in the report.

Main Text

The main text provides the complete outline of the research report along with all details. Title of the research study is repeated at the top of the first page of the main text and then follows the other details on pages numbered consecutively, beginning with the second page. Each main section of the report should begin on a new page. The main text of the report should have the following sections:

- o **Introduction:** The purpose of introduction is to introduce the research project to the readers. It should contain a clear statement of the objectives of research i.e., enough background should be given to make clear to the reader why the problem was considered worth investigating. A brief summary of other relevant research may also be stated so that the present study can be seen in that context. The hypotheses of study, if any, and the definitions of the major concepts employed in the study should be explicitly stated in the introduction of the report. The methodology adopted in conducting the study must be fully explained. The scientific reader would like to know in detail about such thing: How was the study carried out? What was its basic design? If the study was an experimental one, then what were the experimental manipulations? If the data were collected by means of questionnaires or interviews, then exactly what questions were

asked (The questionnaire or interview schedule is usually given in an appendix)? If measurements were based on observation, then what instructions were given to the observers? Regarding the sample used in the study the reader should be told: Who were the subjects? How many were there? How were they selected? All these questions are crucial for estimating the probable limits of generalizability of the findings. The statistical analysis adopted must also be clearly stated. In addition to all this, the scope of the study should be stated and the boundary lines be demarcated. The various limitations, under which the research project was completed, must also be narrated.

- o **Statement of findings and recommendations:** After introduction, the research report must contain a statement of findings and recommendations in non-technical language so that it can be easily understood by all concerned. If the findings happen to be extensive, at this point they should be put in the summarized form.

- o **Results:** A detailed presentation of the findings of the study, with supporting data in the form of tables and charts together with a validation of results, is the next step in writing the main text of the report. This generally comprises the main body of the report, extending over several chapters. The result section of the report should contain statistical summaries and reductions of the data rather than the raw data. All the results should be presented in logical sequence and split into readily identifiable sections. All relevant results must find a place in the report. But how one is to decide about what is relevant is the basic question. Quite often guidance comes primarily from the research problem and from the hypotheses, if any, with which the study was concerned. But ultimately the researcher must rely on his own judgment in deciding the outline of his report. "Nevertheless, it is still necessary that he states clearly the problem with which he was concerned, the procedure by which he worked on the problem, the conclusions at which he arrived, and the bases for his conclusions."

- o **Implications of the results:** Toward the end of the main text, the researcher should again put down the results of his research clearly and precisely. He should, state the implications that flow from the results of the study, for the general reader is interested in the implications for understanding the human behavior. Such implications may have three aspects as stated below :(a) A statement of the inferences drawn from the present study which may be expected to apply in similar circumstances. (b) The conditions of the present study which may limit the extent of legitimate

generalizations of the inferences drawn from the study.(c) The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them. It is considered a good practice to finish the report with a short conclusion which summarizes and recapitulates the main points of the study. The conclusion drawn from the study should be clearly related to the hypotheses that were stated in the introductory section. At the same time, a forecast of the probable future of the subject and an indication of the kind of research which needs to be done in that particular field is useful and desirable.

- **Summary:** It has become customary to conclude the research report with a very brief summary, resting in brief the research problem, the methodology, the major findings and the major conclusions drawn from the research results.

- **End Matter** At the end of the report, appendices should be enlisted in respect of all technical data such as questionnaires, sample information, mathematical derivations and the like ones. Bibliography of sources consulted should also be given. Index (an alphabetical listing of names, places and topics along with the numbers of the pages in a book or report on which they are mentioned or discussed) should invariably be given at the end of the report. The value of index lies in the fact that it works as a guide to the reader for the contents in the report.

Types of reports

Research reports vary greatly in length and type. In each individual case, both the length and the form are largely dictated by the problems at hand. For instance, business firms prefer reports in the letter form, just one or two pages in length. Banks, insurance organizations and financial institutions are generally fond of the short balance-sheet type of tabulation for their annual reports to their customers and shareholders. Mathematicians prefer to write the results of their investigations in the form of algebraic notations. Chemists report their results in symbols and formulae. Students of literature usually write long reports presenting the critical analysis of some writer or period or the like with a liberal use of quotations from the works of the author under discussion. In the field of education and psychology, the favorite form is the report on the results of experimentation accompanied by the detailed statistical tabulations. Clinical psychologists and social pathologists frequently find it necessary to make use of the case-history form.

News items in the daily papers are also forms of report writing. They represent firsthand on-the scene accounts of the events described or compilations of interviews with persons who were on the scene. In such reports the first paragraph usually contains the important information in detail and the succeeding paragraphs contain material which is progressively less and less important. Book-reviews which analyze the content of the book and report on the author's intentions, his success or failure in achieving his aims, his language, his style, scholarship, bias or his point of view.

Such reviews also happen to be a kind of short report. The reports prepared by governmental bureaus, special commissions, and similar other organisations are generally very comprehensive reports on the issues involved. Such reports are usually considered as important research products. Similarly, Ph.D. theses and dissertations are also a form of report-writing, usually completed by students in academic institutions.

The above narration throws light on the fact that the results of a research investigation can be presented in a number of ways viz., a technical report, a popular report, an article, a monograph or at times even in the form of oral presentation. The method(s) of presentation to be used in a particular study depends on the circumstances under which the study arose and the nature of the results. A *technical report* is used whenever a full written report of the study is required whether for recordkeeping or for public dissemination. A *popular report* is used if the research results have policy implications. We give below a few details about the said two types of reports:

Technical Report

In the technical report the main emphasis is on (i) the methods employed, (it) assumptions made in the course of the study, (iii) the detailed presentation of the findings including their limitations and supporting data.

A general outline of a technical report can be as follows:

1. Summary of results: A brief review of the main findings just in two or three pages.

2. Nature of the study: Description of the general objectives of study, formulation of the problem in operational terms, the working hypothesis, the type of analysis and data required, etc.

3. Methods employed: Specific methods used in the study and their limitations. For instance, in sampling studies we should give details of sample design viz., sample size, sample selection, etc.

4. Data: Discussion of data collected their sources, characteristics and limitations. If secondary data are used, their suitability to the problem at hand be fully assessed. In case of a survey, the manner in which data were collected should be fully described.

5. Analysis of data and presentation of findings: The analysis of data and presentation of the findings of the study with supporting data in the form of tables and charts be fully narrated. This, in fact, happens to be the main body of the report usually extending over several chapters.

6. Conclusions: A detailed summary of the findings and the policy implications drawn from the results be explained.

7. Bibliography: Bibliography of various sources consulted be prepared and attached.

8. Technical appendices: Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones.

9. Index: Index must be prepared and be given invariably in the report at the end. The order presented above only gives a general idea of the nature of a technical report; the order of presentation may not necessarily be the same in all the technical reports. This, in other words, means that the presentation may vary in different reports; even the different sections outlined above will not always be the same, nor will all these sections appear in any particular report. It should, however, be remembered that even in a technical report, simple presentation and ready availability of the findings remain an important consideration and as such the liberal use of charts and diagrams is considered desirable.

Popular Report

The popular report is one which gives emphasis on simplicity and attractiveness. The simplification should be sought through clear writing, minimization of technical, particularly mathematical, details and liberal use of charts and diagrams. Attractive layout along with large print, many subheadings, even an occasional cartoon now and then is another characteristic feature of the popular report.

Besides, in such a report emphasis is given on practical aspects and policy implications.

We give below a general outline of a popular report.
1. The findings and their implications: Emphasis in the report is given on the findings of most practical interest and on the implications of these findings.
2. Recommendations for action: Recommendations for action on the basis of the findings of the study is made in this section of the report.
3. Objective of the study: A general review of how the problem arises is presented along with the specific objectives of the project under study.
4. Methods employed: A brief and non-technical description of the methods and techniques used, including a short review of the data on which the study is based, is given in this part of the report.
5. Results: This section constitutes the main body of the report wherein the results of the study are presented in clear and non-technical terms with liberal use of all sorts of illustrations such as charts, diagrams and the like ones.
6. Technical appendices: More detailed information on methods used, forms, etc. is presented in the form of appendices. But the appendices are often not detailed if the report is entirely meant for general public.

There can be several variations of the form in which a popular report can be prepared. The only important thing about such a report is that it gives emphasis on simplicity and policy implications from the operational point of view, avoiding the technical details of all sorts to the extent possible.

Oral presentation

At times oral presentation of the results of the study is considered effective, particularly in cases where policy recommendations are indicated by project results. The merit of this approach lies in the fact that it provides an opportunity for give-and-take decisions which generally lead to a better understanding of the findings and their implications. But the main demerit of this sort of presentation is the lack of any permanent record concerning the research details and it may be just possible that the findings may fade away from people's memory even before an action is taken. In order to overcome this difficulty, a written report may be circulated before the oral presentation and referred to frequently during the discussion. Oral presentation is effective when supplemented by various visual devices. Use of slides, wall charts and blackboards is quite helpful in contributing to clarity and in reducing the boredom, if any. Distributing a board outline, with a few important tables and charts concerning the research results, makes the listeners attentive who have a ready outline on which to focus their thinking. This very often happens in academic

institutions where the researcher discusses this research findings and policy implications with others either in a seminar or in a group discussion.

Thus, research results can be reported in more than one ways, but the usual practice adopted, in academic institutions particularly, is that of writing the Technical Report and then preparing several research papers to be discussed at various forums in one form or the other. But in practical field and with problems having policy implications, the technique followed is that of writing a popular report.

Researches done on governmental account or on behalf of some major public or private organizations are usually presented in the form of technical reports.

Mechanics of writing a research report

There are very definite and set rules which should be followed in the actual preparation of the

research report or paper. Once the techniques are finally decided, they should be scrupulously adhered to, and no deviation permitted. The criteria of format should be decided as soon as the materials for the research paper have been assembled. The following points deserve mention so far as the mechanics of writing a report are concerned:

1. Size and physical design: The manuscript should be written on unruled paper 8 1 2⬜⬜× 11⬜⬜in size. If it is to be written by hand, then black or blue-black ink should be used. A margin of at least one and one-half inches should be allowed at the left hand and of at least half an inch at the right hand of the paper. There should also be one-inch margins, top and bottom. The paper should be neat and legible. If the manuscript is to be typed, then all typing should be double-spaced on one side of the page only except for the insertion of the long quotations.

2. Procedure: Various steps in writing the report should be strictly adhered (All such steps have already been explained earlier in this chapter).

3. Layout: Keeping in view the objective and nature of the problem, the layout of the report should be thought of and decided and accordingly adopted (The layout of the research report and various types of reports have been described in this chapter earlier which should be taken as a guide for report-writing in case of a particular problem).

4. Treatment of quotations: Quotations should be placed in quotation marks and double spaced, forming an immediate part of the text. But if a quotation is of a considerable length (more than four or five type written lines) then it should be single-spaced and indented at least half an inch to the right of the normal text margin.

5. The footnotes: Regarding footnotes one should keep in view the followings:

(a) The footnotes serve two purposes viz., the identification of materials used in quotations in the report and the notice of materials not immediately necessary to

the body of the research text but still of supplemental value. In other words, footnotes are meant for cross references, citation of authorities and sources, acknowledgement and elucidation or explanation of a point of view. It should always be kept in view that footnote is neither an end nor a means of the display of scholarship. The modern tendency is to make the minimum use of footnotes for scholarship does not need to be displayed.

(b) Footnotes are placed at the bottom of the page on which the reference or quotation which they identify or supplement ends. Footnotes are customarily separated from the textual material by a space of half an inch and a line about one and a half inches long.

(c) Footnotes should be numbered consecutively, usually beginning with 1 in each chapter separately. The number should be put slightly above the line, say at the end of a quotation. At the foot of the page, again, the footnote number should be indented and typed a little above the line. Thus, consecutive numbers must be used to correlate the reference in the text with its corresponding note at the bottom of the page, except in case of statistical tables and other numerical material, where symbols such as the asterisk (*) or the like one may be used to prevent confusion.

(d) Footnotes are always typed in single space though they are divided from one another by double space.

6. Documentation style: Regarding documentation, the first footnote reference to any given work should be complete in its documentation, giving all the essential facts about the edition used. Such documentary footnotes follow a general sequence. The common order may be described as under:

(i) Regarding the single-volume reference

1. Author's name in normal order (and not beginning with the last name as in a bibliography) followed by a comma;

2. Title of work, underlined to indicate italics;

3. Place and date of publication;

4. Pagination references (The page number).

(ii) Regarding multivolume reference

1. Author's name in the normal order;

2. Title of work, underlined to indicate italics;

3. Place and date of publication;

4. Number of volume;

5. Pagination references (The page number).

(iii) Regarding works arranged alphabetically

For works arranged alphabetically such as encyclopedias and dictionaries, no pagination reference is usually needed. But if there should be a detailed reference to

a long encyclopedia article, volume and pagination reference may be found necessary.

(iv)Regarding periodicals reference

1. Name of the author in normal order;

2. Title of article, in quotation marks;

3. Name of periodical, underlined to indicate italics;

4. Volume number;

5. Date of issuance;

6. Pagination.

(v)Regarding anthologies and collections reference

Quotations from anthologies or collections of literary works must be acknowledged not only by author, but also by the name of the collector.

(vi)Regarding second-hand quotations reference

In such cases the documentation should be handled as follows:

1. Original author and title;

2. "Quoted or cited in,";

3. Second author and work.

 (vii) Case of multiple authorship

If there are more than two authors or editors, then in the documentation the name of only the first is given and the multiple authorship is indicated by "et al." or "and others".

Subsequent references to the same work need not be so detailed as stated above. If the work is cited again without any other work intervening, it may be indicated as *ibid,* followed by a comma and the page number. A single page should be referred to as p., but more than one page be referred to as pp. If there are several pages referred to at a stretch, the practice is to use often the page number, for example, pp. 190ff, which means page number 190 and the following pages; but only for page 190 and the following page '190f'. Roman numerical is generally used to indicate the number of the volume of a book. Op. cit. (opera citato, in the work cited) or Loc. cit. (loco citato, in the place cited) are two of the very convenient abbreviations used in the footnotes. Op. cit. or Loc. cit. after the writer's name would suggest that the reference is to work by the writer which has been cited in detail in an earlier footnote but intervened by some other references.

7. Punctuation and abbreviations in footnotes: The first item after the number in the footnote is the author's name, given in the normal signature order. This is followed by a comma. After the comma, the title of the book is given: the article (such as "A", "An", "The" etc.) is omitted and only the first word and proper nouns and adjectives are capitalized. The title is followed by a comma. Information concerning the edition is given next. This entry is followed by a comma. The place of publication is then stated; it may be mentioned in an abbreviated form, if the place

happens to be a famous one such as Lond. for London, N.Y. for New York, N.D. for New Delhi and so on. This entry is followed by a comma. Then the name of the publisher is mentioned and this entry is closed by a comma. It is followed by the date of publication if the date is given on the title page. If the date appears in the copyright notice on the reverse side of the title page or elsewhere in the volume, the comma should be omitted and the date enclosed in square brackets [c 1978], [1978]. The entry is followed by a comma. Then follow the volume and page references and are separated by a comma if both are given. A period closes the complete documentary reference. But one should remember that the documentation regarding acknowledgements from magazine articles and periodical literature follow a different form as stated earlier while explaining the entries in the bibliography. Certain English and Latin abbreviations are quite often used in bibliographies and footnotes to eliminate tedious repetition.

8. Use of statistics, charts and graphs: A judicious use of statistics in research reports is often considered a virtue for it contributes a great deal towards the clarification and simplification of the material and research results. One may well remember that a good picture is often worth more than a thousand words. Statistics are usually presented in the form of tables, charts, bars and line-graphs and pictograms. Such presentation should be self-explanatory and complete in itself. It should be suitable and appropriate looking to the problem at hand. Finally, statistical presentation should be neat and attractive.

9. The final draft: Revising and rewriting the rough draft of the report should be done with great care before writing the final draft. For the purpose, the researcher should put to himself questions like: Are the sentences written in the report clear? Are they grammatically correct? Do they say what is meant'? Do the various points incorporated in the report fit together logically? "Having at least one colleague read the report just before the final revision is extremely helpful. Sentences that seem crystal-clear to the writer may prove quite confusing to other people; a connection that had seemed self-evident may strike others as a *non-sequitur.* A friendly critic, by pointing out passages that seem unclear or illogical, and perhaps suggesting ways of remedying the difficulties, can be an invaluable aid in achieving the goal of adequate communication."

10. Bibliography: Bibliography should be prepared and appended to the research report as discussed earlier.

11. Preparation of the index: At the end of the report, an index should invariably be given, the value of which lies in the fact that it acts as a good guide, to the reader. Index may be prepared both as subject index and as author index. The former gives the names of the subject-topics or concepts along with the number of pages on which they have appeared or discussed in the report, whereas the latter gives the similar information regarding the names of authors. The index should always be

arranged alphabetically. Some people prefer to prepare only one index common for names of authors, subject-topics, concepts and the like ones.

Precautions for writing research reports

Research report is a channel of communicating the research findings to the readers of the report. A good research report is one which does this task efficiently and effectively. As such it must be prepared keeping the following precautions in view:

1. While determining the length of the report (since research reports vary greatly in length), one should keep in view the fact that it should be long enough to cover the subject but short enough to maintain interest. In fact, report-writing should not be a means to learning more and more about less and less.

2. A research report should not, if this can be avoided, be dull; it should be such as to sustain reader's interest.

3. Abstract terminology and technical jargon should be avoided in a research report. The report should be able to convey the matter as simply as possible. This, in other words, means that report should be written in an objective style in simple language, avoiding expressions such as "it seems," "there may be" and the like.

4. Readers are often interested in acquiring a quick knowledge of the main findings and as such the report must provide a ready availability of the findings. For this purpose, charts, graphs and the statistical tables may be used for the various results in the main report in addition to the summary of important findings.

5. The layout of the report should be well thought out and must be appropriate and in accordance with the objective of the research problem.

6. The reports should be free from grammatical mistakes and must be prepared strictly in accordance with the techniques of composition of report-writing such as the use of quotations, footnotes, documentation, proper punctuation and use of abbreviations in footnotes and the like.

7. The report must present the logical analysis of the subject matter. It must reflect a structure where in the different pieces of analysis relating to the research problem fit well.

8. A research report should show originality and should necessarily be an attempt to solve some intellectual problem. It must contribute to the solution of a problem and must add to the store of knowledge.

9. Towards the end, the report must also state the policy implications relating to the problem under consideration. It is usually considered desirable if the report makes a forecast of the probable future of the subject concerned and indicates the kinds of research still needs to be done in that particular field.

10. Appendices should be enlisted in respect of all the technical data in the report.

11. Bibliography of sources consulted is a must for a good report and must necessarily be given.

12. Index is also considered an essential part of a good report and as such must be prepared and appended at the end.

13. Report must be attractive in appearance, neat and clean, whether typed or printed.

14. Calculated confidence limits must be mentioned and the various constraints experienced in conducting the research study may also be stated in the report.

15. Objective of the study, the nature of the problem, the methods employed and the analysis techniques adopted must all be clearly stated in the beginning of the report in the form of introduction.