RESEARCH METHODS: A Brief Note for Students Use

CHAPTER ONE

CONCEPTS OF SCIENTIFIC RESEARCH: AN INTRODUCTION

Methods of acquiring knowledge

Authority: When knowledge is obtained from individuals who are qualified producers of knowledge, the sources of knowledge is said to be authority. E.g. knowledge about religion from priests and sheiks, students get knowledge from students, etc…

Tradition: When we accept without questioning the traditions of the society within which we are living, the source of knowledge is said to be tradition. E.g. it is not culturally accepted for males to wear females' dresses.

Experience: Our personal experience or our observation of the experiences of other people is another fundamental source of knowledge. We need no one to tell us about it. E.g. we know rainy months in a year from our experience.

Reasoning: Knowledge can be obtained based on some rules and forms of logic.

Deductive reasoning:- This is a thinking process in which one will go from general statements to specific statements. It is a means to evaluate the correctness of any given conclusion or idea by going from the known to the unknown. It consists

i. Major premise, a previously established fact

ii. Minor premise, a particular case to which the fact can be applied

iii. A conclusion. (i) and (ii) must be true for a conclusion to be true

e.g. All animals are mortal
Cat is an animal
Cat is mortal

Inductive reasoning:- In this case we need to go from specific to general ideas. Here conclusions will be given by observing

particular instances (situations) and generalizing from them to the whole phenomenon.

e.g. mammals are mortal
insects are mortal
birds are mortal
Fishes are mortal
Plants are mortal
Replies are mortal
All living things are mortal

Scientific Method: Here knowledge will be obtained through a step by step process of collecting data, analyzing it and reaching at a conclusion.

The philosophy common to all research methods and techniques, although they may vary considerably from one science to another, is usually given the name scientific method. Scientific method is the pursuit of truth as determined by logical considerations. **Scientific method is based on certain basic postulates** which can be stated as follows:

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 in to probabilistic predictions
6. Its methodology is made known to all concerned for critical scrutiny and for use in testing the conclusions through replication

It aims at formulating most general axioms or what can be termed as scientific theories.

Accordingly, **scientific method** implies an objective, logical and systematic method, i.e., a method free from personal bias or prejudice, a method to ascertain demonstrable qualities of a phenomenon capable of being verified, a method wherein the researcher is guided by the rules of logical reasoning, a method wherein the investigation proceeds in an orderly manner and a method that implies internal consistency.

Purpose of Studying Research Methods

Acquiring research skills may be helpful for the following.

1. We may often need more information before making certain decisions;

2. We may also be called up on to do a research study for a higher-level executives;
3. We may again need to buy research services from others, or at least evaluate research done by others; or
4. We may find a career position as a research specialist in various areas of our discipline.

What is Research?

There are several definitions given for research by different writers. Some of these definitions are the following:
Research *refers the application of the scientific method in the study of problems. It is a systematic attempt to obtain answers to questions about events through the application of scientific procedures.*
Research is a scientific and systematic search for pertinent information on a specific topic.
Research is a systematic inquiry aimed at providing information to solve problems.
Research is a systematic process of collecting and analyzing information for some purpose.

One can easily pick the terms *scientific* and *systematic* from the above definitions. Research is scientific because it is just a search for knowledge through recognized methods of data collection, analysis, and interpretation. research is systematic because it uses some steps or procedures in doing this.

Generally, **the purpose of research** is to discover answers to questions through the application of scientific procedures. It is an addition to the available knowledge, which contributes for further advancement.

Types of Research

Research can be classified based on

A.the goal of research
B.the spesific objective of research
C.the approaches used by the research
D.the designs used
E.the type of data used in research
F.the fields of study

1. **Descriptive Vs analytical:**

a) **Descriptive research**: the major purpose of descriptive research is description of the state of the state of affairs as it exists at present.

Ex post facto research: the main caracteristcs of this method is that a researcher has no control over the variables.

b)**analytical research:** the researcher has to use facts or information already available and analyze these to make a critical evaluation of the material.

I. Basic Vs Applied Research

a) Applied Research (also called action research): The purpose of applied research is finding a solution to solve an immediate, practical problem facing a society. It is oriented to a specific problem. It has a practical problem solving emphasis. It emerges out of a general problem which faces a society as whole.

b) Basic Research (also called pure or fundamental research): *is concerned with generalisations & formulation of a theory.* It has a more general orientation, adding to the existing body of knowledge in the discipline. It does not necessarily provide results of immediate, practical use, although such possibility is not ruled out. It may consist of a research to develop and/or improve on theories, techniques and measurements. It is aimed to solve perplexing questions of theoretical nature that have little direct impact on action, performance, or policy decisions.

Generally, basic and applied researches are differentiated not by their complexity or value, but by their goals or purposes. The primary purpose of basic research is the extension of knowledge; the purpose of applied research is the solution of an immediate practical problem.

II. Quantitative Vs Qualitative Research

a) Quantitative Research: Research based on the measurement of quantity or amount.

b) Qualitative Research: Research concerned with qualitative phenomenon.

III. Conceptual Vs Empirical Research

a) Conceptual Research: related to abstract ideas or theories. Commonly used by philosophers and thinkers.

b) Empirical Research: relies on experience or observation alone. It is data based research coming up with conclusions which are capable of being verified by observation or experiment.

The Systematic Process of Research

1. ***Formulating the Research Problem:*** The first step in the research process is the choice of a suitable problem for

investigation. Research starts with a felt difficulty. It takes place when there is a problematic situation and a need to solve the problem. **Sources of a research problem** are one's professional experience, theory, and professional literature. Initially the problem may be stated in a broad general way and then narrowed down to a specific research problem.

2. ***Review the Literature:*** Once the problem is formulated, a brief summary of it should be written down. At this juncture, the researcher should undertake extensive literature survey connected with the problem. The researcher may review two types of literature-conceptual and empirical. **Conceptual literature** is concerned with related concepts and theories. **Empirical literature** is about studies made earlier which are similar to the one proposed.
3. ***Formulation of Working Hypothesis:*** After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. Hypotheses provide focal point for research. The role of hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right truck. However, we may encounter a problem where we do not need working hypotheses.
4. ***Preparing the Research Design (Including Sample Design):*** The research problem having been formulated in clear cut terms, the researcher will be required to prepare a research design. This is stating the conceptual structures with in which research would be conducted. In addition, the researcher must design the way of selecting a sample or what is known as the sample design.
5. ***Collecting the Data (Execution):*** The researcher should select one of the methods of data collection taking in to consideration the nature of investigation, objective and scope of the inquiry, financial resources, available time, and the desired degree of accuracy. Then by using the selected method, adequate and dependable data have to be collected.
6. ***Analysis of Data (Including Test of Hypothesis):*** After the data have been collected, the researcher turns to the task of analyzing them. The analytical techniques have to be appropriate and adequate. After analyzing the data, the researcher is in apposition to test the hypothesis, if any, he had formulated earlier.
7. ***Drawing Conclusions:*** This is making generalizations after the analyses have been made. The conclusions must be based on the data and the analysis within the framework of the research study. Finally, the researcher has to prepare the report of what has been done by him.

Characteristics of a Good Research

1. The purpose of the research, or the problem involved, should be clearly defined and sharply delineated in terms as unambiguous as possible
2. The research procedures used should be described in sufficient detail to permit another researcher to repeat the research for further advancement.
3. The procedural design of the research should be carefully planned to yield results that are as objective as possible.
4. The researcher should report, with complete frankness, flaws in procedural design and estimate their effort up on the findings.
5. Analysis of the data should be sufficiently adequate to reveal its significance, and the methods of analysis used should be appropriate.
6. Conclusions should be confined to those justified by the data of the research and limited to those for which the data provide an adequate basis
7. Greater confidence in the research is warranted if the researcher is experienced, has good reputation in research, and is a person of integrity.

CHAPTER TWO

DEFINING THE RESEARCH PROBLEM

This is the beginning activity of the research process, and it is often the most difficult. The problem must be identified with adequate specificity. It is at this point in many studies that hypotheses are generated. Variables must be identified and defined adequately so that necessary data can be identified in preparation for data collection. A researcher must find the problem and formulate it so that it becomes susceptible to research. Like a medical doctor, a researcher must examine carefully all the symptoms concerning the problem before suggesting a solution. That is why quite often we all hear that a problem clearly stated is a problem half solved.

WHAT IS A RESEARCH PROBLEM?

A **research problem**, in general, refers to some difficulty which a researcher experiences in the context of either a theoretical or practical situation and wants to obtain a solution for that problem. The individual or the organization can be said to have the problem only if the individual or the organization does not know what course of action is best, i.e. the individual or the organization must be in doubt about the solution.

The components of the research problem:

1. There must be an individual or a group which has some difficulty or the problem
2. There must be some objective(s) to be attained at. If one wants nothing, one cannot have a problem
3. There must be alternative means (or the courses of action) for obtaining the objective(s) one wishes to attain. If a researcher has no choice of means, he can not have a problem.
4. There must remain some doubt in the mind of a researcher with regard to the selection of alternatives. This means the research must answer the question concerning the relative efficiency of the possible alternatives.
5. There must be some environment(s) to which the difficulty pertains.

Thus, a research problem is one which requires a researcher to find out the best solution for the given problem, i.e., to find out by which course of action the objective can be attained optimally in the context of a given environment.

SELECTION OF A RESEARCH PROBLEM/criteria

The first step in the problem identification involves selecting a research topic. It is then from a research topic that a specific statement of the problem is generated.

However, the following points may need to be observed by a researcher in selecting a research problem or a subject for research.

i) Subject which is overdone should not be normally chosen, for it will be a difficult task to throw any new light in such a case.
ii) Controversial subject should not become the choice of an average researcher.
iii) Too narrow or too vague problems should be avoided
iv) The subject selected for research should be familiar and feasible so that the related research material or sources of research are with in one's reach. A researcher should contact an expert who is already engaged in research. He may as well read articles published in current literature available on the subject. Generally, the researcher should make all possible efforts in selecting a problem.
v) The importance of the subject, the qualifications and the training of a researcher, the costs involved, and the time factor are few other criteria that must also be considered in selecting a problem.
vi) When the field of inquiry is relatively new, the selection of a problem must be preceded by a preliminary study.

The problem to be investigated must be defined unambiguously for that will help to discriminate relevant data from the irrelevant ones. If the subject for research is selected properly by observing the above mentioned points, the research will not be a boring drudgery. The selection of a research problem involves reading, discussing, and conceptualizing. Typically, the process is one of successive approximations to the problem as factors related to the problem are considered.

Necessity of defining the problem

A proper definition of a research problem will enable the researcher to be on the track whereas an ill-defined problem may create hurdles. Thus defining a research problem properly is a prerequisite for any study and is a step of the highest importance. In fact, formulation of a problem is often more essential than its solution.

TECHNIQUES IN DEFINING A RESEARCH PROBLEM.

The research problem should be defined in a systematic manner,

giving due attention to all relating points. The technique for this purpose involves the following general steps.

(i) Statement of the problem in a general way: First of all the problem should be stated in a broad general way, However, it should be noted that the problem stated in a broad general way may contain various ambiguities which must be resolved by cool thinking and rethinking over the problem.

(ii) Understanding the nature of the problem: The next step in defining the problem is to understand its origin and nature clearly. The best way of understanding the problem is to discuss it with those who first raised it in order to find out how the problem originally came about and with what objectives in view. If the researcher has stated the problem himself, he should consider once again all those points that induced him to make a general statement concerning the problem.

(iii) Surveying the available literature: All available literature concerning the problem at hand must necessarily be surveyed and examined before a definition of the research problem is given. This is done to find out what data and other materials, if any, are available for operational purposes. Surveying literature would also help a researcher to know if there are certain gaps in the theories,

(iv) Developing the ideas through discussions: Various new ideas can be developed through discussions. Hence, a researcher must discuss his problem with someone who have enough experience in the same area or in working on similar problems. Thus, discussions must be made not only with those who have a research experience but also with potential users of the research results.

(v) Rephrasing the research problem: Finally,Through rephrasing, the researcher puts the research problem in as specific terms as possible so that it may become operationally viable and may help in the development of working hypotheses.

CHAPTER THREE

DEVELOPING A RESEARCH PROPOSAL

After identifying and defining the problem, the researcher must arrange his ideas in order and write them in the form of an experimental plan or what can be described as a research proposal. A research proposal is also known as a work plan, prospectus, outline, statement of intent, or draft plan. It tells us what will be done, why it will be done, how it will be done, where it will be done, to whom it will be done, and what is the benefit of doing it.A **research proposal** is essentially a road map, showing clearly the location from which a journey begins, the destination to be reached, and the method of getting there.

THE PURPOSE OF PROPOSALS

The general purpose for writing proposals is
1. To present the problem to be researched and its importance.
2. To discuss the research efforts of others who have worked on related problems.
3. To suggest the data necessary for solving the problem and how the data will be gathered, treated, and interpreted.

Proposals are also useful both for sponsors and for researchers. **For sponsors,** it allows to assess the sincerity of the purpose, the clarity of the design, the extent of the background material, and fitness of the researcher to undertake the project. The proposal provides a document the sponsor can evaluate based up on the current organizational, scholastic, or scientific needs. It allows the research sponsor to assess both the researcher and the proposed design, to compare them against competing proposals, and to make the best selection for the project. Proposals that fail to convince sponsors that they will lead to practical outcomes are not likely to be funded. The proposal also provides a basis for the sponsor to evaluate the results of the project. By comparing the final product with the stated objectives, it is easy to decide if the research goals have been achieved.

For the researcher, proposal helps to have a tentative work plan that charts the logical steps needed to accomplish the stated goals. Additionally, there is the opportunity to spot flaws in the logic, errors in assumptions, or even problems that are not adequately addressed by the objectives and design. After the acceptance of the proposal, the

document serves as a guide for the researcher though out the investigation.

STRUCTURING A RESEARCH PROPOSAL / COMPONENTS

Cover page

The cover page contains introductory information for the proposal: the names of the proposed project (title), the author of the proposal or principal investigator, and the institution.

Title

The title should use the fewest possible words that adequately describe the content of the paper. In titles a researcher has to use specific, familiar, and short words. Use of 'waste words' like 'A study on …', 'An investigation on…', 'An observation on…' should be avoided.

Abstract/summery

It is a short summary of the research proposal which allows a busy manager or sponsor to understand quickly the thrust of the proposal without having to read the details.It should include a brief statement of the problem, the research objectives, and the benefits of the approach.

Background of the study

In the background, the researcher presents the background of what he wants to do and say briefly why he thinks the works should be done and relate it to what is already known about the problem.

Statement of the problem

Here a researcher has to capture the reader's attention by stating the problem, and its consequences. The researcher is expected to state his problem clearly without the use of idioms. Each word of the statement must be expressive, sharp, and indispensable.

Research objectives

This part addresses the purpose of the investigation. It is best to list the objectives either in order of importance or in general terms first, moving to specific terms.

Research hypothesis

The researcher has to clearly state a working hypothesis. A hypothesis is a tentative predictive statement that represents a very specific proposed answer to the problem statement.

Significance of the research

This part shows the explicit benefits and the beneficiaries of the results of the research being conducted. It states the importance or contribution that the study will have for different bodies.

Delimitations (Scope) of the study: This part indicates the scope (coverage) of the study. It should be delimited to manageable size.

Limitations of the study

This is a part that you will include some constraints or difficulties you think that they have influence on the results of your study becaus of the weaknesses in the methodology, lack of access to data, faulty instruments, sampling restrictions, lack of recent literature in the area etc.

Literature review

A section in our research that explains, interprets and discusses what has been researched and documented previously is known as review of related literature. The literature review section examines recent research studies, company data, or industry reports that act as a basis for the proposed study. Begin your discussion of the related literature and relevant secondary data from a comprehensive perspective, moving to more specific studies that are associated with your problem.

Avoid the extraneous details of the literature; do a brief review of the information, not a comprehensive report. Emphasis the important results and conclusions of other studies, the relevant data and trends from previous research, and particular methods or designs that could be duplicated or should be avoided. Discuss how the literature applies to the study you are proposing; show the weaknesses and faults in the design, discussing how you would avoid similar problems.

Always refer to the original source to avoid any errors of interpretation or transcription. Close the literature review section by summarizing the important aspects of the literature and interpreting them in terms of your problem.

Research methodology

Here, the researcher is expected to clearly state what he is going to do in technical terms. In this part we will indicate the variables that we will treat in our study, the subjects of our study, the sampling technique we will use, our method of data collection, the procedures we will follow for collecting the data and our method of data analysis.

As such, this section should include as many subsections as needed to show the phases of the project. Some of these subsections can be sampling techniques and sample size, types and sources of data, methods of data collection, and methods of analysis.

Budget/logistics and Work plan

Budget requirement

In order to carry out a given research you need to have money. The sources may be from government budget or from supporting agencies. Whatever the source may be, we are expected to assign a reasonable amount of money for the study in a form the sponsor requests. In addition, the money we have, need to be planned on how to spend it. Budget breakdown has to be prepared for all activities. It may include the details of travel expenses, per diems, and capital equipment purchase, and other costs together with allowance for contingencies. When the time comes to do the work, the researcher should know exactly how much money is budgeted for each particular task.

Work plan (Time Schedule)

We should also prepare a realistic time schedule for completing the study within the time available by dividing a study into phases and assigning dates for the completion of each phase. The work plan (schedule) should include the major phases of the project together with their timetables. Major phases may be questionnaire preparation, pre-testing, field interviews, editing and coding, data analysis, and report generation.

Bibliography/references

In a research proposal we should give a list of books, journals, and other documents that we have used in selecting the problem and which we may use while we conduct the study. For all projects that require literature review, a bibliography is necessary. Use the bibliographic format required by the sponsor. If none is specified, a standard style manual will provide the details necessary to prepare the bibliography.

Qualification of researchers

This consists of summaries of the experience, education, publications, and research activities of individuals who will work on the proposed project. It is customary to begin with the principal investigator and then to the co-investigators.

CHAPTER FIVE

SAMPLING DESIGN

CENSUS AND SAMPLE SURVEY

All items in any field of inquiry constitute a 'universe' or a 'population'. A complete enumeration of all items in the population is known as a census inquiry. In such an inquiry, when all items are covered, no element of chance is left and highest accuracy is obtained. This type of inquiry, however, 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 beyond the reach of ordinary researchers. Government, in very rare cases, is the only institution which can get the complete enumeration carried out. For example, population census in our country is carried out once in a decade.

Undertaking a census survey, many a times, is not possible. Sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population, technically called samples. The process of selecting samples is called sampling technique. In sampling, however, the samples selected should be as representative of the total population as possible in order to produce a miniature cross-section.

A 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. Sample design is determined before data are collected.

THE NEED FOR SAMPLING

The following points summarize the benefits of studying samples.

1. There could be resource (time, finance, manpower, etc.) limitations which would make it difficult to study the whole population.
2. In some cases, tests may be destructive. For example, when we test the breaking strength of materials, we must destroy them. A census would mean complete destruction of materials. In such a case, we must sample.
3. Sampling provides much quicker results than does a census. When the time between the recognition of the need of information and the availability of that information is short, sampling helps not to miss the information.
4. Sampling is the only process possible if the population is infinite.

5. There is also an argument that the quality of a study is often better with sampling than with a census. The basis of the argument is that sampling possesses the possibility of better interviewing, more thorough investigation of missing, wrong, or suspicious information, better supervision, and better processing than is possible with complete coverage.

STEPS IN SAMPLING DESIGN

While developing a sample design, the researcher must pay attention to the following points.

1. **Type of universe:** The first step in developing any sample design is to define the universe. The universe can be finite or infinite. In **finite universe**, the number of items is certain. Examples can be the population of a city, the number of workers in a factory, etc. But in case of an **infinite universe**, the number of items is infinite. Examples for an infinite universe can be listeners of a specific radio program, number of stars in the sky, etc.
2. **Sampling Unit:** A decision has to be taken concerning a sampling unit before selecting sample. Sampling unit may be a geographical one such as district, *kebele*, village, etc., or a social unit such as family, school, etc., or it may be an individual.
3. **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 (for finite universe). A source 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.
4. **Size of sample:** This refers to the number of items to be selected from the universe to constitute a sample. The size of sample should neither be excessively large, nor too small. It should be optimum. In order to decide on the size of the sample to be selected, a researcher must take in to consideration the size of population variance, the size of population, the parameter of interest in the research study, and budgetary constraint.
5. **Parameters of interest:** In determining the sample design, one must consider the question of the specific population parameters which are of interest.
6. **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 non-probability samples.
7. **Sampling procedure:** Finally, the researcher must decide the type of sample he will use. He must decide about the technique to be used in selecting the items for the sample. There are several sample designs 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 FOR SELECTING A SAMPLING PROCEDURE

In this context one must remember that two costs are involved in a sampling analysis-the cost of collecting the data and the cost of an incorrect inference resulting from the data. There are two causes of incorrect inferences namely systematic bias and sampling error.

Causes of incorrect inferences

1. Systematic bias: Systematic bias results from errors in the sampling procedures, and it can not be reduced or eliminated by increasing the sample size. However, the causes responsible for these errors can be detected and corrected. Bias enters in when a sample fails to represent the population it was intended to represent. Usually a systematic bias is the result of one or more of the following factors:

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

ii) Defective measuring device: 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.

iii) Non-respondents: If we are unable to sample all the individuals initially included in the sample, there may arise a systematic bias.

iv) Indeterminancy 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.

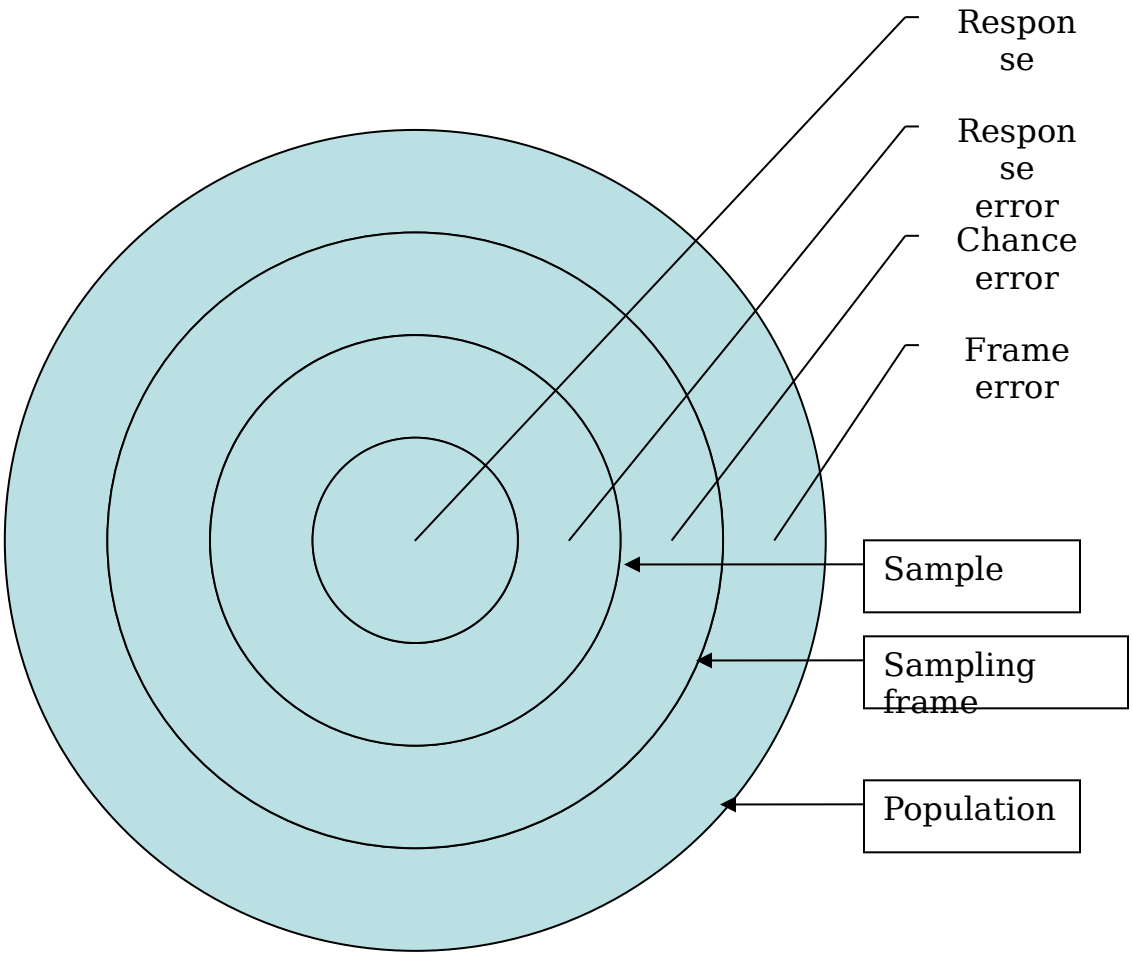
v) Natural bias in the reporting of data: This also leads to a systematic bias. There is usually a downward bias in the income data collected by government taxation department, whereas we find an upward bias in the income data collected by some social organization.

2. Sampling errors: 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. 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 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.



Sampling error = Frame error + Chance error + Response error

Characteristics of a good sample design:

Sample design must

a) result in a truly representative sample

b) result in small sampling error
c) be viable in the context of funds available for the research study
d) must enable to control the systematic bias in a better way
e) be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

TYPES OF SAMPLE DESIGNS

Sample designs are basically of two types-Probability sampling and non-probability sampling. Probability sampling is based on the concept of random selection, whereas non-probability sampling is non-random selection.

PROBABILITY SAMPLING TECHNIQUES

Probability sampling is also known as random sampling or chance sampling. Random sampling techniques can be divided in to simple random sampling and complex random sampling.

I. Simple Random Sampling

Individuals are randomly drawn from the population at large. Under this method each unit in the universe has the same chance of being included in the sample. Random sampling needs a 'sampling frame' or a 'sampling unit', i.e. complete and up-to-date list of all members of the population. For a homogeneous type of population, simple random sampling is reliable. Under this sampling design, every item of the universe has an equal chance of inclusion in the sample. It is blind chance alone that determines whether one item or the other is selected. Random sampling is considered as the best technique of selecting a representative sample. The problem with this method is that it is time consuming.

Simple random sampling gives: advantage

a) each element in the population an equal probability of getting into the sample; and all choices are independent of one another
b) each possible sample combination has an equal probability of being chosen.

We can define a simple random sample from a finite population as a sample which is chosen in such a way that each of the ${}^{N}C_{n}$ possible samples has the same probability, $1/{}^{N}C_{n}$, of being selected. For example, if a finite population consists of 4 elements (say a, b, c, d) i.e. N=4. Suppose that we want to take a sample of n=2 from it. Then there are ${}^{4}C_{2}=6$ possible distinct samples of the required size, and they consist of the elements ab, ac, ad, bc,bd, and cd. If we choose one of these samples in such a way that each has a probability 1/6 of being chosen, we will then call this a random sample.

With regard to how to take a random sample, we can use a lottery method. To do this, we first write the name of each element of a finite

population on a slip of paper, and then put the slips of paper in to a box and mix them thoroughly and then draw the required number of slips for the sample one after the other without replacement. Using the above example where we want to draw two from 4 elements, the probability of drawing any one element for our sample in the first draw is 2/4, the probability of drawing one more element in the second draw is 1/3 (the first element drawn is not replaced). Since these draws are independent, the joint probability of the two elements which constitute our sample is the product of their individual probabilities and this works out to 2/4 X 1/3 =1/6. This verifies our earlier calculation.

Another method of selecting random samples is to use random number tables. Random numbers are often tabulated on some standard books. Sampling is then just by picking numbers at random from the table until the required sample size is obtained.

II. Complex random sampling

Probability sampling under restricted sampling techniques may result in complex random sampling designs. It is sometimes also called 'mixed sampling designs' for many of such designs may represent a combination of probability and non-probability procedures in selecting a sample. Some of the popular complex random sampling designs are as follows:

(i) Systematic sampling: In some instances, the most practical way of sampling is to select every i^{th} item on a list. An element of randomness is introduced into this kind of sampling by using random numbers to pick up the unit with which to start.

The following **steps** will help us to make systematic sampling:

- Assign a sequence number to each member of the population.
- Determine the skip interval by dividing the number of units in the population by the sample size. I=P/S where I is skip interval, P is population size, and S is sample size.
- Select a starting point in a random digit table (it must be between 1 and I).
- include that item in a sample and select every i^{th} item thereafter until total sample has been selected.

For example, if we want to take 20 samples from a population of 100 members, our skip interval is 5 (i.e 100/20). Our starting point must be selected randomly from the interval 1 to 5. Then every fifth item will be our sample. If our starting point is 2, then our sample must include members with sequence numbers of 2, 7, 12, 17, 22, 25, …, 97.

The advantage of this sampling technique

- the samples will spread evenly over the entire population.
- It is also an easier and less costly method of sampling and can be conveniently used even in case of large populations.

However, if there is a hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling.

(ii) Stratified sampling: If a population from which a sample is to be drawn does not constitute a homogeneous group, stratified sampling is generally applied. Under stratified sampling, the population is divided into several subpopulations (strata) that are individually more homogeneous than the total population and then we select items from each stratum to constitute a sample. Stratified sampling results in a more reliable and detailed information. The basic steps for stratified sampling are:

a) Divide the population to be surveyed in to strata of similar study units or into areas with which similar social, environmental, or economic conditions exist.
b) Make a separate and complete list of the stratum and from each stratum draw a separate random sample of study units using these lists.
c) A similar survey is then done on the sample of study units in each of the strata i.e. the same questionnaire is used.

The main advantages of stratified sampling are:

(i) more reliable information is obtained for the same sample size if the population is stratified than they are for the population as a whole.

(ii) Comparisons between strata are easy. This is so because a separate but similar survey is done in each stratum.

The following questions are highly relevant in the context of stratified sampling:

How to form strata?

The strata are formed on the basis of common characteristic(s) of the items to be put in each stratum. This means that various strata be formed in such a way as to ensure elements being most homogeneous within each stratum and most heterogeneous between the different strata. Strata are purposively formed usually based on past experience and personal judgment of the researcher. At times, pilot study may be conducted for determining a more appropriate and efficient stratification plan.

How should items be selected from each stratum?

For selection of items from each stratum, we may use simple random sampling. Systematic sampling can also be used if it is considered more appropriate in certain situations.

How many items be selected from each stratum or how to allocate the sample size to each stratum?

We usually follow the method of proportional allocation under which the sizes of the samples from the different strata are kept proportional to the sizes of the strata. That is, if P_i represents the proportion of population included in stratum ***i***, and ***n*** represents the total sample size, the number of elements selected from stratum ***i*** is $\boldsymbol{n.P_i}$.

For example, Suppose we want to take a sample of size n=30 to be drawn from a population of size N=800 which is divided into three strata of size N_1=400, N_2=240, and N_3=160. The sample size for stratum with N_1=400 is n_1=30(400/800)=15.

The sample size for stratum with N_2=240 is n_2=30(240/800)=9.

The sample size for stratum with N_3=160 is n_3=30(160/800)=6.

In cases where strata differ not only in size but also in variability and is considered reasonable to take larger samples from the more variable strata and smaller samples from the less variable strata, we can then account for both (differences in stratum size and differences in stratum variability) by using disproportionate sampling design by using the formula:

$$n_i = \frac{n.N_i.\sigma_i}{N_1\sigma_1 + N_2\sigma_2 + \ldots + N_k\sigma_k}$$

Where σ_i denote the standard deviations for the i^{th} stratum, N_i denote the size of the i^{th} stratum, and n_i denote the sample size of the i^{th} stratum.

For example, assume a population is divided into three strata so that N_1=5000, N_2=2000, and N_3=3000.Respective standard deviations are $\sigma_1 = 15, \sigma_2 = 18, \sigma_3 = 5$. How should a sample of size n=84 be allocated to the three strata , if we want optimum allocation using disproportionate sampling design?

The solution is $n_1 = \frac{84(5000)(15)}{(5000)(15) + (2000)(18) + (3000)(5)} = 50.$

$$n_2 = \frac{84(2000)(18)}{(5000)(15) + (2000)(18) + (3000)(5)} = 24.$$

$$n_3 = \frac{84(3000)(5)}{(5000)(15) + (2000)(18) + (3000)(5)} = 10$$

NB. 1. Each stratum, in stratified sampling, is homogeneous internally and

heterogeneous with other strata.

2. The more strata used, the closer you come to maximizing interstrata differences and minimizing intrastratum variances.

(iii) Cluster sampling: If the total area of interest happens to be a big one, a convenient way in which a sample can be taken is to divide the area in to a number of smaller non-overlapping areas and then to randomly select a number of these smaller areas (clusters), with the ultimate sample consisting of all units in these small areas or clusters.

In cluster sampling, the total population is divided into a number of relatively small subdivisions which are themselves clusters of still smaller units and then some of these clusters are randomly selected for inclusion in the overall sample.

Advantage

Cluster sampling reduces cost by concentrating surveys in selected clusters. But certainly it is less precise than simple random sampling.

Differences between stratified sampling and cluster sampling

Stratified sampling	**Cluster sampling**
1. We divide the population into a few subgroups, each with many elements in it. The subgroups are selected according to some criterion that is related to the variables under study	1. We divide the population into many subgroups, each with a few elements in it. The subgroups are selected according to some criterion of ease or availability in data collection
2. We try to secure homogeneity within subgroups and heterogeneity between subgroups	2. We try to secure heterogeneity within subgroups and homogeneity between subgroups, but we usually get the reverse
3. We randomly choose elements from within each subgroup	3. We randomly choose a number of subgroups, which we then typically study in toto

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

Multi-stage random sampling: Multi-stage sampling is a further development of the principle of cluster sampling. Suppose we want to investigate the working efficiency of health stations in Ethiopia, and we want to take a sample of few health stations for this purpose. The first stage is to select large primary sampling unit such as regional states in country. Then we may select certain zones in the second stage. Then we may again select districts (*woredas*) and interview all

health stations in the selected districts. This would represent a three-stage sampling design.

If instead of taking a census of all health stations within the selected districts, we randomly sample health stations from each selected districts, then it is a case of a four stage sampling plan. If we select randomly at all stages, we will have what is known as multi-stage random sampling design.

Sampling with probability proportional to the cluster size: In case the cluster sampling units do not have the same number or approximately the same number of elements, it is considered appropriate to use a random selection process where the probability of each cluster being included in the sample is proportional to the size of the cluster. For this purpose, we have to list the number of elements in each cluster irrespective of the method of ordering the cluster. Then we must sample systematically the appropriate number of elements from the cumulative totals.

For illustration, consider the following are the number of departmental stores in 10 cities: 35, 17, 10, 32, 80, 18, 26, 19, 26, and 57. If we want to select a sample of 8 stores, using cities as clusters and selecting within clusters proportional to size, how many stores from each city should be chosen? (Use a starting point of 8)
Let us put the information as follows:

City number	***Number of departmental stores***	***Cumulative total***	***Sample***
1	35	35	8
2	17	52	48
3	10	62	
4	32	94	88
5	80	174	128, 168
6	18	192	
7	26	218	208
8	19	237	
9	26	263	248
10	57	320	288

Since in the given problem, we have 320 departmental stores from which we have to select a sample of 8 stores, the appropriate sampling interval is 40. As we have to use the starting point of 8, so we add successively increments of 40 till 8 numbers have been selected. The numbers, thus, obtained are: 8, 48, 88, 128, 168, 208, 248, and 288. Two stores should be selected randomly from city number 5 and one each from city number 1, 2, 4, 7, 9, and 10.

Comparison of probability sampling designs

Design	Random selection	Other characteristics
Simple random sampling	Sample members individually from the population	Each population element has an equal chance to be selected *Disadv*.-requires a listing of population elements, -expensive and requires more time to implement
Systematic sampling	The initial sample member is individually selected	Designation of the initial sample member determines the entire sample. *Disadv*. –periodicity within the population may skew the sample and the results
Stratified	Sample members individually within each of the subpopulations or strata	All strata are represented in the sample most frequently by proportional allocation *Disadv*. –creating strata on the population is expensive
Cluster sampling	Clusters of members selected from the larger population of clusters	All members of a selected clusters are included in the sample Not all clusters are included *Disadv*. –often lower statistical efficiency (more error) due to subgroups being homogeneous rather than heterogeneous

NON-PROBABILITY SAMPLING TECHNIQUES

While useful for many studies, non-probability sampling procedures provide only a weak basis for generalization. In reality, the conclusions drawn from a study of a non-probability sample are limited to that sample and cannot be used for further generalization. In this type of sampling, items for the sample are selected deliberately by the researcher; his choice concerning the items remains supreme. For instance, if economic conditions of people living in a state are to be studied, a few towns and villages may be purposively selected for intensive study on the principle that they can be representative of 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. Sampling error in this type of sampling cannot be estimated and the element of bias, great or small, is always there. Some of the major non-probability sampling techniques are the following:

a) Judgment (purposive) sampling - The researcher uses his judgment to select people that he feels are representative of the population to have a particular expertise or knowledge which makes them suitable.

b) Convenience (accidental) sampling - the most convenience population is chosen, which may be the researchers friends, work colleagues or students from a nearby college. This method is often used to save time and resources.

c) Quota sampling - The researcher selects a predetermined number of individuals from different group (i.e. based on age, gender, etc). This is perhaps the most popular non-probability sampling method used. Quota sampling is useful when a particular group or characteristics is relatively rare in the population. By setting a quota, and selecting people until the quota is filled the investigator ensures that the group or characteristics he/she wants in the sample are there.

d) Referral Sampling: this is a non-probability sampling technique which utilizes some form of referral, wherein respondents who are initially contacted are asked to supply the names and addresses of members of the target population.

CHAPTER FOUR

THE SOURCES AND COLLECTION OF DATA

TYPES AND SOURCES OF DATA

The task of data collection begins after a research problem has been defined and research design chocked out. **Data** are records of the actual state of some measurable aspect of the universe at a particular point in time. Data are not abstract; they are concrete, they are measurements or the tangible and countable features of the world. In general, data could be **quantitative** (expressed in numerical form) or **qualitative** (expressed in the form of verbal descriptions rather than numbers).

When choosing whether to collect quantitative or qualitative data, the following factors need to be considered:

1. The purpose for which the data is required:***quantitative data*** is necessary if one requires a high degree of precision or wants to perform statistical analysis, while ***qualitative data*** is useful for providing a detailed or vivid impression of the issue or characteristic concerned.

2. The subject matter: some kinds of subject matter (eg. Production, export levels, prices, imports, income, etc.) are relatively easily presented in ***quantitative form***, while others (eg. Attitude to a new product, religious beliefs, etc) tend to be more appropriately presented in ***qualitative form.***

3. The method of data collection: the collection of ***quantitative data*** is based on statistically designed survey procedures, while the collection of ***qualitative data*** relies primarily on detailed observation or interview.

4. The method of data presentation: ***qualitative data*** can often be 'translated' into a quantitative form if it can be 'scaled' in some way; for example, information on attitudes can be grouped into categories (eg strongly agree, agree, neutral, disagree, strongly disagree) which can then be subjected to statistical analysis.

Depending on the source, the type of data collected could be primary or secondary in nature.

Type of data

Primary data are those which are collected afresh and for the first time, and thus happen to be original in character. Its advantage is its

relevance to the user, but it is also likely to be expensive in time and money terms to collect.

Secondary data are those which have already been collected by someone else and which have already been passed through the statistical process. It is information extracted from an existing source, probably published or held on a computer database. From Practical point of view this type of information is collected for any purpose other than the current research objectives and is not always up-to-date. For this reason it may not precisely meet the needs of the secondary user. However, it is less expensive and time-consuming to obtain. Therefore, it provides a good starting point and very often can help the investigator to formulate and generate ideas which can later be refined further by collecting primary data.

COLLECTION OF PRIMARY DATA

Primary data can be collected through

- experimentation in experimental research, or
- through surveys, whether sample surveys or census surveys.

An experiment is a special form of research, which sets out to examine the relationship between two factors by manipulating one whilst measuring changes in the other.

- There are two types of experiments: *Field experiments* and *laboratory experiments*.
- In fact, experimentation is a means of obtaining data with relatively high precision in measurement of the variables.

Survey refers to the method of securing information concerning a phenomenon under study from all or a selected number of respondents of the concerned universe.

Survey designs can be broadly divided into cross-sectional and longitudinal designs.

A cross-sectional survey collects data at one time. The researcher can generalize findings from such one-shot studies to the sampled population only at the time of the survey.

A longitudinal survey takes place over time with two or more data collections and has the benefit of measuring change over time. The following are the types of longitudinal surveys:

i) **A trend survey** is a longitudinal survey in which a general population is studied over time. Usually the population is sampled and random samples are measured.

ii) **A cohort survey** is a longitudinal survey in which a specific population is studied over time.

iii) **A panel survey** is a longitudinal survey in which the same sample is measured two or more times. The samples can represent either a specific or a general population.

- Survey includes several methods of collecting primary data such as **observation**, **interview**, through **questionnaires**, and other methods.

OBSERVATION METHOD

Observation is the most commonly used method of data collection especially, in behavioral studies. This method could be used both for cross checking information obtained using other methods and for understanding processes which are difficult to grasp in an interview context. This method is useful when studying subjects who are not capable of giving verbal reports of their feelings for one reason or another.

In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves the formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under this method, the information is sought by way of the investigator's own direct observation without asking from the respondent.

Advantages of observation method:

1. subjective bias is eliminated, if observation is done accurately
2. the information obtained relates to what is currently happening; it is not complicated by either the past behavior or future intentions or attitudes
3. it is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method.

Limitations:

1. expensive;
2. the information obtained is limited

3. Sometimes unforeseen factors may interfere with the observational task.

Types of observation:

Structured observation: the observer has a clear definition of the units to be observed, the style of recording the observed information, the selection of the pertinent data of observation, etc.
Non-structured observation: the opposite of what is mentioned under structured observation

Depending on the nature of the **observer, we can classify observations** into three basic forms:

Secretive - where the subjects of the study are unaware that they are being observed
Non-participant: where the subjects of the study are aware that they are being observed but the observer takes no part in the behavior being observed.
Participant: when the subject and the observer interact.

INTERVIEW METHOD

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

Personal interviews: This method requires a person (interviewer) asking questions in a face-to-face contact to the interviewee.

If the interview is carried out in a structured way, it is called

- ***structured interview***. This involves the use of a set of predetermined questions and highly standardized techniques of recording. The interviewer in a structured interview follows a rigid procedure laid down, asking questions in a form and order prescribed.
- ***unstructured interviews.*** are characterized by a flexibility of approach to questioning. In unstructured interview, the interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so requires. He may even change the sequence of questions. But this sort of flexibility results in lack of comparability of one interview with another and the analysis of unstructured responses becomes much more difficult and

time consuming than that of the structured responses obtained in case of structured interviews.

Advantages of personal interviews:

1. More information and in greater depth can be obtained
2. The interviewer by his own skill can overcome the resistance, if any, of the respondents
3. There is greater flexibility especially in case of unstructured interviews
4. personal information can be obtained easily
5. samples can be controlled effectively as there arises no difficulty of missing returns; non-response generally remains very low
6. the language of the interview can be adopted to the ability or educational level of the person interviewed

Some of the weaknesses of the personal interview method:

1. It is very expensive, specially when large and widely spread geographical sample is taken
2. The possibility of the bias of interviewer as well as that of the respondent
3. Certain types of respondents may not be easily approachable (eg. Important officials or executives, people in high income groups)
4. It is relatively more time consuming

- For successful implementation of the interview method,
- interviewers should be carefully selected, trained and briefed.
- They should be honest, sincere, hardworking, impartial and must possess the technical competence and necessary practical experience.
- Occasional field checks should be made to ensure that interviewers are neither cheating nor deviating from instructions given to them for performing their job efficiently.

Telephone interviews: This method of collecting information consists in contacting respondents on telephone itself. It is not a very widely used method, but plays important part in industrial surveys, particularly in developed countries.

Some of the chief merits of telephone interview are:

1. It is faster than other methods
2. It is cheaper than personal interview method; the cost per response is relatively low
3. Recall is easy; callbacks are easy and economical
4. Replies can be recorded without causing embarrassment to respondents
5. No field staff is required

Some of the demerits of telephone interview are:

1. Little time is given to respondents for considered answers
2. Surveys are restricted to respondents who have telephone facilities
3. It is not suitable for intensive surveys where comprehensive answers are required
4. Questions have to be short and to the point; probes are difficult to handle

DATA COLLECTION THROUGH QUESTIONNAIRE

This method is quite popular, particularly in case of big inquiries. Service evaluations of hotels, restaurants, transportation providers, and other service providers are good examples of self-administered questionnaire. Often a short questionnaire is left to be completed by the respondent in a convenient location. In a mail survey, a questionnaire can also be sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire.

A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself.

The merits of questionnaire method are:

1. it is free from the bias of the interviewer; answers are in respondents' own words
2. respondents have adequate time to give well thought out answers
3. respondents who are not easily approachable can also be reached conveniently

The main demerits of questionnaire system can be:

1. it can be used only when respondents are educated and cooperating
2. the control over questionnaire may be lost once it is sent
3. there is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been dispatched
4. There is also possibility of ambiguous replies or omission of replies altogether to certain questions

Schedules: data collection through schedules is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules are being filled in by the enumerators who are specially appointed for the purpose. These enumerators along

with schedules go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same.

Here, enumerators should be very carefully selected and trained to perform the job well. Enumerators should be intelligent and must possess the capacity of cross-examination in order to find out the truth. Above all, they should be honest, sincere, hardworking and should have patience.

Essentials of a good questionnaire: To be successful,

- questionnaire should be comparatively short and simple.
- Questions should proceed in logical sequence moving from easy to more difficult questions.
- Personal questions should be left to the end.
- Technical terms and vague expressions capable of different interpretations should be avoided.
- Questions may be *dichotomous* (yes or no answers), *multiple choice* (alternative answers listed), or *open-ended* (inviting free response).
- The later type of questions are often difficult to analyze and hence should be avoided in a questionnaire to the extent possible.

CASE STUDY METHOD

It is a method of in depth study rather than in breadth. Under case study, more emphasis is given on the full analysis of a limited number of events or conditions and their interrelations. It is an intensive and comprehensive investigation of the particular unit under consideration. The objective is to locate the factors that account for the behavioral patterns of the given unit as an integrated totality.

Characteristics of case study method:

1. A researcher can take one single social unit or more of such units for his study purpose
2. The selected unit is studied intensively, i.e. in minute details
3. A researcher has to make complete study of the social unit covering all facets
4. The approach happens to be qualitative and not quantitative
5. The effort is to know the mutual interrelationship of causal factors
6. The behavioral pattern of the concerning unit is studied directly and not by an indirect and abstract approach

7. It results in fruitful hypothesis along with the data which may be helpful in testing them, and thus it enables the generalized knowledge to get richer and richer.

Major phases involved in case study:

1. Recognition and determination of the status of the phenomenon to be investigated or the unit of attention
2. Collection of data, examination and history of the given phenomenon
3. Diagnosis and identification of causal factors as a basis for remedial or development treatment
4. Application of remedial measures or treatments
5. Follow-up program to determine effectiveness of the treatment applied.

COLLECTION OF SECONDARY DATA

The use of existing data (secondary data) in a research activity is termed as desk research simply because the person carrying it out can usually gather such data with out leaving his/her desk. In any type of study, it is advisable to assess the availability of secondary data before embarking upon a primary data collection exercise, since the latter is expensive in terms of time, money and manpower.

Sources of Secondary data:

- Different Central Statistical Authority Publications;
- Different Publications by Regional Governments;
- Various publications by the different Ministries;
- Publications of the National Bank of Ethiopia;
- On-line and Electronic Data Bases;
- Reports and publications of various associations, business organizations, etc
- Various publications of international, multilateral and non-governmental Organizations;
- Report of research scholars and consultants;
- Historical documents, archives, maps, photographs, letters, biographies, autobiographies, diaries, textbooks, periodicals;
- Popular media (Newspapers, magazines, Radio and television).

Researcher must be very careful in using secondary data.

- The researcher, before using secondary data, must see that they posses the following characteristics:

1. **Reliability of data:** reliability can be tested by answering questions like who collected them. What were the sources of data? What methods were used to collect them? At what time were they collected? How they were analyzed etc.
2. **Suitability of data:** Data must be evaluated whether they could serve for another purpose other than the one for which they were collected. This should be seen in terms of definitions of various terms and units of collection used at the time of collecting the data from primary source originally. Similarly, the object, scope and nature of the original inquiry must also be studied.
3. **Adequacy of data:** This should be done in terms of area coverage, level of accuracy, number of respondents, etc.

SELECTION OF APPROPRIATE METHOD FOR DATA COLLECTION

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

1. **Nature, scope and object of inquiry:** The method selected should be such that it suits the type of inquiry 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:** 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:** 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.

However, one must always remember that each method of data collection has its uses and none is superior in all situations.

CHAPTER SIX

DATA PROCESSING, ANALYSIS AND INTERPRETATION

The data, after collection, has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research proposal.

DATA PROCESSING OPERATIONS

Technically speaking, processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis.

1) Editing: Editing is a process of examining the collected raw data to detect errors and omissions and to correct these when possible. Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as complete 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 forms by the investigator for completing what the interviewer (enumerator) has written in abbreviated and/or in illegible form at the time of recording the respondents' responses. This type 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 questionnaires have been completed and returned to the office. 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. At times the respondents can be asked 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.

Editing is done to assure that the data are:

- Accurate
- Concise with other information
- uniformly entered
- As complete as possible
- Arrange to facilitate coding and tabulation.

2) Coding: Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. These classes must possess the characteristic of exhaustiveness (there must be a class for every data item) and also that of mutual exclusivity which means that a specific answer can be placed in one and only one cell in a given category set.

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. In case of hand coding, it is possible to code on the margin of the questionnaire with colored pencil or to transcribe the data from the questionnaire to a coding sheet.

3)Classification: Classification is the process of arranging data in groups or classes on the basis of common characteristics, especially for studies with large volume of raw data. 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 according to attributes based on descriptive information (such as literacy, sex, honesty, etc.). A researcher, based on attributes, can classify in to 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.

Classification can also be according to class-intervals.Numerical characteristics relating to income, production, age, weight, etc. can be classified based on class-intervals. For instance, persons whose incomes, say, are within Br 201 to Br 400 can form one group, those whose incomes are within Br 401 to Br 600 can form another group and so on. The difference between the two class limits is known as ***class magnitude***. The number of items which fall in a given class is known as the ***frequency* of the given class**.

4) Tabulation: When a mass of data has been assembled, it becomes necessary for 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 summarizing raw data and displaying the same in compact form for further analysis.
Tabulation is essential because of the following reasons:
i) It conserves space and reduces explanatory and descriptive statement to a minimum
ii) It facilitates the process of comparison

iii) It facilitates the summation of items and the detection of errors and omissions
iv) It provides a basis for various statistical computations.

Some problems in processing

A) *The problem concerning 'Don't know' (or DK) responses:* While processing the data, the researcher may come across DK responses. 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 a question arises: Is the question which elicited DK response useless? The answer depends on two points namely; 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.

B) *Use of percentages:* Percentages are often used in data presentation for they simplify numbers, reducing all of them to a 0 to 100 range. 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
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 causal-factor in case of two dimension tables and for this purpose we must select the more significant factors as the causal factor.

DATA ANALYSIS

The term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. Analysis involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences. Analysis can be categorized as descriptive analysis and inferential (statistical) analysis.

Descriptive analysis is largely the study of distribution of one variable. The characteristics of location, spread, and shape describe distributions. Their definitions, applications, and formulas fall under the heading of descriptive statistics. The common *measures of*

location, often called *central tendency*, include mean, median, and mode. The common *measures of spread*, alternatively called *measures of dispersion*, are variance, standard deviation, and range. The common *measures of shape* are skewness and kurtosis.

Inferential analysis includes two topics, estimation of population values and testing statistical hypothesis.

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 (regression) 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.

Measurement Scales

Before analyzing data, it is important to identify the measurement scales of the data type. There are four basic measurement scales: nominal, ordinal, interval, and ratio. The most accepted basis for scaling has three characteristics:

1. Numbers are ordered. One number is less than, greater than, or equal to another number.
2. Differences between numbers are ordered. The difference between any pair of numbers is greater than, less than, or equal to the difference between any other pair of numbers.
3. The number series has a unique origin indicated by the number zero.

Combination of these characteristics of order, distance, and origin provide the following widely used classification of measurement scales.

Nominal Scales: When we use nominal scale, we partition a set into categories that are mutually exclusive and collectively exhaustive. The counting of members is the only possible arithmetic operation and as a result the researcher is restricted to the use of the mode as the measure of central tendency. If we use numbers to identify categories, they are recognized as labels only and have no quantitative value. Nominal scales are the least powerful of the four types. They suggest no order or distance relationship and have no arithmetic origin. Examples can be respondents' marital status, gender, students' Id number, etc.

Ordinal Scales: Ordinal scales include the characteristics of the nominal scale plus an indicator of order. The use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) with out stating how much greater or less. Thus the real difference between ranks 1 and 2 may be more or less than the difference between ranks 2 and 3. The appropriate measure of central tendency for ordinal scales is the median. Examples of ordinal scales include opinion or preference scales.

Interval Scales: The interval scale has the powers of nominal and ordinal scales plus one additional strength: It incorporates the concept of equality of interval (the distance between 1 and 2 equals the distance between 2 and 3). When a scale is interval, you use the arithmetic mean as the measure of central tendency. Calendar time is such a scale. For example, the elapsed time between 4 and 6 A.M. equals the time between 5 and 7 A.M. One cannot say, however, 6 A.M is twice as late as 3 A.M. because zero time is an arbitrary origin. Centigrade and Fahrenheit temperature scales are other examples of classical interval scales.

Ratio Scales: Ratio scales incorporate all of the powers of the previous ones plus the provision for absolute zero or origin. The ratio scale represents the actual amounts of a variable. Multiplication and division can be used with this scale but not with the other mentioned. Money values, population counts, distances, return rates, weight, height, and area can be examples for ratio scales.

Summary of measurement scales

Type of scale	*Characteristics*	*Basic empirical operation*
Nominal	No order, distance, or origin	Determination of equality
Ordinal	Order but no distance or unique origin	Determination of greater or lesser values
Interval	Both order and distance but no unique origin	Determination of equality of intervals or differences
Ratio	Order, distance, and unique origin	Determination of equality of ratios

CORRELATION ANALYSIS

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 workers devote for leisure is somewhat related to their income,

to age, to sex, to education level or to similar other factors. We may ask 'Is there any association or correlation between the two (or more) variables? If yes, of what degree?' These questions are answered by the use of correlation analysis.

Karl Pearson's coefficient of correlation (or simple correlation): Is the most widely used method of measuring the degree of relationship between two variables. It can be worked out as:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1) S_X . S_Y}$$

Correlation coefficients reveal the magnitude and direction of relationships. Pearson's correlation coefficient varies over a range of +1 through 0 to -1. The sign signifies the direction of relationship.

There are two **basic assumptions for Pearson's correlation coefficient.**

- *linearity*. When r =0, no pattern is evident that could be described with a single line. It is possible to find coefficients of zero where the variables are highly related but in a non-linear form.
- *bivariate normal distribution*. That is, the data are from a random sample of a population where the two variables are normally distributed in a joint manner. If this assumption is not met, one should select a nonparametric measure of association.

Spearman's coefficient of correlation (or rank correlation): When the data are not available to use in numerical form but the information is sufficient to rank the data as first, second, third, and so forth, we quite often use the rank correlation method. In fact, the rank correlation coefficient is a measure of correlation that exists between two sets of ranks.

For calculating rank correlation coefficient, rank the observations by giving 1 for the highest value, 2 to the next highest value, and so forth. If two or more values happen to be equal, then the average of the ranks which should have been assigned to such values had they been all different, is taken and the same rank is given to concerning values. The next step is to record the difference between ranks ('*d*') for each pair of observations, then square these differences to obtain a total of such differences. Finally, Spearman's rank correlation coefficient can be worked out as:

$$r_r = 1 - \left\{ \frac{6 \sum d_i^2}{n(n^2 - 1)} \right\}$$

The value of Spearman's rank correlation coefficient will always vary between -1 and 1, where 1 indicates a perfect positive correlation and -1 indicates a perfect negative correlation.

II.REGRESSION ANALYSIS

The statistical tool with the help of which we are in a position to estimate (or predict) the unknown values of one variable from known values of another variable is called regression. For example if we know that advertising and sales are correlated, we may find out the expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales. If we take two variables x and y , we shall have two regression lines as under.

a. Regression Equation of X on Y
b. Regression equation of Y on X

III. TEST OF HYPOTHESIS

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In general, it is a mere assumption or some supposition to be proved or disproved. But for the researcher, hypothesis is a formal question that he intends to resolve.

For Example,

1) Students who receive counseling will show a greater increase in creativity than students not receiving counseling.
2) The automobile A is performing as well as automobile B.

CHARACTERISTICS OF HYPOTHESIS

1) It should be clear and precise.
2) It should be capable being tested.
3) It should state relationship between variables.
4) It should be limited in scope and must be specific,
5) Hypothesis should be stated as far as possible in most simple terms. So that the same is easily understandable by all.
6) Hypothesis should be consistent with most known facts.
7) Hypothesis should be amenable to testing within a reasonable time.
8) Hypothesis must explain the facts that gave rise to the need for explanations.

BASIC CONCEPTS CONCERNING TEST OF HYPOTHESIS

The following are the important basic concepts of hypothesis,

1) Null hypothesis and alternative hypothesis

If we are compare method A with method B about its superiority and we proceed on the assumption that both methods are equally good, then this assumption is termed as the null hypothesis. As against that, we may think that method A is superior or method B is inferior, we are then stating what is termed as alternative hypothesis.
The null hypothesis termed as $=H_O$
The alternative hypothesis termed as$=H_1$

If we accepting $H_{0,}$ then we are rejecting H_1 and if we reject $H_{0,}$ then we are accepting H_1.
The null hypothesis and alternative hypothesis are chosen before the sample drawn.
In the choice of null hypothesis, the following points should be considered.

a) Alternative hypothesis is usually to prove and null hypothesis is usually to disprove. Thus, a null hypothesis should be rejecting and alternative hypothesis represents all other possibilities.

b) If the rejection of a certain hypothesis when it is actually true involves great risk, it is taken as null hypothesis because then the probability of rejecting it when it is true (level of significance) which is chosen very small.

c) Null should always be specific hypothesis i.e. it should not state about or approximately a certain value.

2) The level of significance

It is very important in hypothesis testing. It is always some % (usually5%), which should be chosen. The 5% level of significance means that the researcher is willing to take as much as a5% risk of rejecting the null hypothesis when it (H_o) happens to be true. Thus the level of significance is the maximum value of the probability of rejecting H_o when it is true and is usually determined in advance before testing the hypothesis.

3) DECISION RULE OR TEST OF HYPOTHESIS

When make a rule, which is known as decision rule according to we accept H_o (i.e. reject H_1) or reject H_o(accept H_1). If H_O is that a certain lot is good (there are very few items in it) against H_1 that the lot is not good (there are too many defective items in it), then we must

decide the number of items to be tested and the criterion. We might test 10 items in the lot and plan our decision saying that if there are none or only 1 defective item among the 10, we will accept H_o otherwise we will reject H_O (or accept H_1). It is known as decision rule.

4) TYPE I AND TYPE II ERRORS

There are two type of error in the hypothesis testing. We may reject H_o when H_o is true and we may accept H_o when it is not true. The former is known as type 1 error and latter is known as type II error. In other words, type I error means rejecting the hypothesis, which should have been accepted, and type II error means accepting the hypothesis, which should have been rejected.

Type I error is denoted by α (alpha) known as α error, also called as level of significance of the test. Type II error is denoted by (Beta) known as β error. The following table will explain the two errors as clearly.

	DECISION	DECISION
	ACCEPT H_o	REJECT H_O
H_O TRUE	CORRECT DECISION	TYPE I ERROR (α error)
H_O FALSE	TYPE II ERROR (β error)	CORRECT DECISION

The probability of type I error is usually determined in advance and it is the level of significance of testing the hypothesis. If type I error is fixed at 5%, it means that there are about 5 chances in 100 that we will reject H_O when Ho is true. We can control the type I error just by fixing at lower level. Suppose we fix at 1%, we will say that the maximum probability of committing type I error would only be 0.01.

When we try to reduce type I error, the probability of committing type II error will increase. Both types of errors cannot be reduced simultaneously. So it should be trade- off between two types of error. Therefore, we are usually fixing the level of significance (Type I error) is 5% for hypothesis testing.

5) TWO TAILED AND ONE TAILED ERROR.

A two tailed test rejects the null hypothesis if, say, the sample mean is significantly higher or lower than the hypothesized value of the mean of population. Symbolically, the two-tailed test is appropriate when we have $H_{O:}\ \mu = \mu_H$ and $H_1: \mu \neq \mu_H$ which may mean. Thus in two-tailed test, there are two rejection region or critical region.

STEPS OR PROCEDURE FOR HYPOTHESIS TESTING

The various steps involved in hypothesis testing are stated below:

1) Making a formal statement

It is making a formal statement of the null hypothesis (H_O) and the alternative hypothesis (H_1). For instance, Ato. William of a Civil Engineer wants to test the load of an old bridge, which must be more than 10 tonus, his hypothesis should be,

Null hypothesis $H_O: \ \mu = 100$ tons
Alternative hypothesis H1 : $\mu > 10$ tons

2) Selecting a significance level;

The hypothesis is tested on a predetermined level of significance. Generally, in practice, either 5% level or 1% level is adopted for the purpose.

The factors that affect the level of significance are,

a) The magnitude of the difference between sample means.
b) The size of the samples
c) The variability of measurement within samples
d) Whether hypothesis directional or non-directional.

3) Deciding the distribution to use:

After deciding the level of significance, the next step is to determine the appropriate sampling distribution. The choice generally remains between normal distribution and t-distribution

4) Selecting a random sample and computing an appropriate value;

Select a random sample and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution, In other words, draw a sample to furnish empirical data.

5) Calculation of the probability

The next step is to calculate the probability that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact.

6) Comparing the probability

Comparing the calculated value with specified value of α, the significance level. If the calculated value of probability is equal to or less than the value in case of one-tailed test (and $\alpha/2$ in case of two-tailed test), then reject the null hypothesis and accept the alternative hypothesis. If the calculated probability is greater than, then accept null hypothesis and reject the alternative hypothesis.

The following are the **different types of test in hypothesis**

a) Hypothesis testing of mean
b) Hypothesis testing of difference between two means
c) Hypothesis testing of proportion
d) Hypothesis testing of difference between two proportions

IV. ANOVA

Analysis of variance (ANOVA) is useful technique in research field. Prof. R.A. Fisher was the first man to use the term variance and in fact, he was developed a very elaborate theory concerning ANOVA. It is essentially a procedure for test the difference among different groups of data for homogeneity. ANOVA is useful to a manager of a big concern, which can analyze the performance of various salesman of his concern in order to know whether their performances differ significantly.

If we take only one factor and investigate the differences among its various categories, it is said to be one-way ANOVA. In case, we investigate two factors at the same time, then we have to use two-way ANOVA. We have to make 2 estimate of population variance i.e. one based on between samples variance and other based on within sample variance. The two estimate of population are compared with F-test by

F= Estimate of population variance based on between sample variance

Estimate of population variance based on within sample variance

Techniques of ANOVA (One-way/single factor ANOVA)

We consider only one factor and taking several types of samples can occur within that factor. We can determine if there is difference within that factor. This technique involves the following steps;

1) Take the total of the values of individual items in all the samples i.e

   Workout ΣX_{ij} and it is called as T.

   i=1, 2, 3 j=1, 2, 3

2) Workout the correction factor as under

$$\text{Correction factor} = \frac{T^2}{n}$$

3) Find out the square of all the item value one by one and then take its total.
Subtract the correction factor from this total and the result is the sum of squares for total variance. It can be written as

$$\text{Total SS} = \Sigma X_{ij}^2 - \frac{T^2}{n}$$

i=1, 2, 3, j=1, 2.3,

4) Obtain the square of each sample total (T_J^2) and divide such square value of each sample by the number of items in the concerning sample and take the total of the result obtained. Subtract the correction factor from this total and the result is the sum of squares for variance between the samples. It can be written as;

$$\text{SS between} = \sum \frac{(T_j)^2}{n_j} - \frac{(T)^2}{n}$$

i=1, 2, 3,
j=represents different samples.

5) Divide the result of the fourth step by the degree of freedom between samples to obtain variance or mean square (MS) between samples. It can be written as;

$$\text{MS between} = \frac{\text{SS between}}{(k-1)}$$

(k-1) represents degrees of freedom between samples

6) Obtain the deviations of the values of the samples items for the entire sample from corresponding means of the samples and calculate the squares of such deviations and then obtain their total. This total is known as sum of squares for variance within samples (SS within)It can be written as;

$$\text{SS within} = \Sigma (X_{1i} - X_1)^2 + \Sigma (X_{2i} - X_2)^2 + \ldots\ldots\Sigma (X_{ki} - X_k)^2$$

i=1, 2, 3,.....i.e. each item of samples.

7) Divide the result of sixth step by degrees of freedom within samples to obtain the variance or mean square (MS) within samples. It can be written as

$$\text{MS within} = \frac{\text{SS within}}{n-k}$$

(n-k) represents degrees of freedom within samples
n= total number of items in all the samples i.e. $n_{1+}\ n_{2+}.....n_{K.}$
k=number of samples

8) Verification;-

SS for total variance = SS between + SS within

The degrees of freedom for total variance will be equal to the number of items in all samples minus one (n-1). It can be verified by

n-1 = (k-1) + (n-k)

9) Finally, F- ratio may be worked as follows;

$$\text{F-ratio} = \frac{\text{MS between}}{\text{MS within}}$$

If the calculated value is less than the F- table value, there is no difference between sample mean or insignificant and vice versa.

V. CHI-SQUARE TEST

Chi-square test is an important test among the several tests of significance developed by statisticians. It is written as χ^2 (chi-square). It is useful to make comparison between theoretical population and actual data when categories are used. It is possible for all researchers to

a) Test the goodness of fit.
b) Test the significance of association between two attributes.
c) Test the significance of population variance.

Characteristics of chi-square test.

1) It is based on frequencies and not on the parameters like mean and S.D.
2) It is used for testing the hypothesis and not for estimation.
3) It is useful to the research work.
4) This test can also be applied to a complex contingency table with several classes.
5) It have less mathematical details are involved.
6) No rigid assumptions are necessary regarding the type of population.

Types of chi-square test

Based on classification of data, the chi-square test may be three types;

a) Test for goodness of fit.
b) Test of independence.
c) Test of equality of several population proportions.

VI. OTHER METHODS

1. Index number:

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. One 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 date. 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 typo of average, meant to study the changes in the effect so 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 given 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 labeled as 'Time Series'. More clearly it can be stated that series of successive observations f 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:

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

a. 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.
b. Seasonal fluctuations (or S) are of short duration occurring in 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 tend to be repeated form year to year. Cyclical fluctuations and seasonal fluctuations taken together constitute short-period regular fluctuations.

c. 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 analyses 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 form the series. The given series is, thus, left with the effects of one type of factor only.
For analyzing time series, we usually have tow models; (1) multiplicative model; and (2) additive model. Multiplicative model assumes that the various components interact in 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 form the given series viz., the free hand method, semi-average 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 short-term 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.

CHAPTER SEVEN
RESEARCH REPORT WRITING AND PRESENTATION

Purpose 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 favors 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 form 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. The following are the important purpose of the writing the reports;

1) Presenting the results
2) Easy to understand the findings to all
3) To take the modification in the research work
4) Future reference
5) Guidance to other researchers
6) Formal completion
7) Written evidence
8) Finding new idea

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 basic of mental connections and associations between the one thing and another by means of analysis. Logical treatment often consists in developing the material form 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 of 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 loses bricks."[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.

Example

Kothari, C.R., *Quantitative Techniques,* New Delhi, Vikas Publishing House Pvt. Ltd., 1978.

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.

Example

Robert V. Roosa, “Coping with short-term International Money flows”, *The Banker,* London, September, 1971, p.995.

The above examples are just the samples for bibliography entries and may be used, but one should also remember that they are not the only acceptable forms. The only thing important is that, whatever method one selects, it must remain consistent.

Writing the final draft: This constitutes the last. 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 the 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 or 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 abut 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 financing are to be taken. For this purpose there is 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.

(A) *Preliminary Pages*

In its preliminary pages the report should carry a *little 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.

(B) *Main Text*

The main text provides the complete outline of the research report along with all details. Tile 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: (i) Introduction; (ii) Statement of findings and recommendations; (iii) The results; (iv) The implications drawn from the results; and (v) the summary.

(i) *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.

(ii) *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.

(iii) *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 splitter 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 judgments 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 bases for his conclusions.”[5]

(iv) *Implications of the results:* To ward 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 answerers for them.

It is considered a good practice to finish the report with a short conclusion which summaries and recapitulates the main points of the study. The conclusion drawn from the study should be clearly related to 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.

(v) *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 form the research results.

(C) *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.

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 unrolled paper 81/2” X 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 an 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 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 worked, footnotes are meant for cross-references, citation of authorities and sources, acknowledgement and elucidation to 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 the case of statistical tables and other numerical materials, 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).

Example

John Gassner, *Masters of the Drama,* New York: Dover Publications, Inc. 1954. p. 315.

(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. In such cases the order is illustrated as under:

Example 1

"Salamanca," *Encyclopedia Britannica,* 14th Edition.

Example 2

"Mary Wollstonecraft Godwin," *Dictionary of national biography.* But if there should be a detailed reference to a long encyclopedia article, volume and pagination reference may be found necessary.

(iv) *Regarding periodical 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.

Example
J.F. Jones, *Life in Polynesia,* P. 16. Quoted in *History of the Pacific Ocean area,* by R.B.Abel, p. 191.

(vii) *Case of multiple authorship*

in 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 other".

Subsequent references to the same work need not be as 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 is referred to as PP. If there are several pages referred to at as trench, 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) is two of the very convention 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 ahs 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. Information concerning the edition is given next. This entry is followed by a comma. The place of publication is then state; it may be mentioned in an abbreviated form, if the place happens to be a famous one such as Land. 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 clawed 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. The following is a partial list of the most common abbreviations frequently used in report-writing (the researcher should learn to recognize them as well as he should learn to use them):

anon.,	anonymous
ante.,	before
art.,	article
aug.,	augmented
bk.,	book
bull.,	bulletin
cf.,	compare
ch.,	chapter
col.,	column
diss.,	dissertation
ed.,	editor, edition, edited
ed.cit	edition cited
e.g.,	exempli gratia: for example
eng.,	enlarged
et.al.,	and others
et seq.,	et sequens: and the following
ex.,	example
f., ff.,	and the following
fig(s).,	figure(s)
fn.,	footnote

ibid., ibidem:	in the same place (when two or more successive footnotes refer to the same work, it is not necessary to repeat completer reference for the second footnote. ibid. may be used. if different pages are referred to, pagination must be shown).
id., idem:	the same

ill., illus., or illust(s).	illustrated, illustration(s)
intor. intro.,	introduction
l, or 11,	line(s)
loc. cit.,	in the place cited; used as op.cit., (when new reference
loco citato:	is made to the same pagination as cited in the previous note)
MS. MSS.,	manuscript or manuscripts
N.B., nota bene:	note well
n.d.,	no date
n.p.,	no place
no pub.,	no publisher
no(s).,	number(s)
o.p.,	out of print
op.cot	in the work cited (if reference has been made to a work and new
opera citato	reference is to be made, ibid., may be used; if intervening reference has been made to different works, op.cit. must be used. The name of the author must precede.
p. or pp.,	page(s)
passim:	here and there
post:	after
rev.,	revised
tr., trans.,	translator, translated, translation
vid or vide	see, refer to
viz.,	namely
vol. or vol(s).,	volume(s)
vs., versus:	against

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 on 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 emending the difficulties, can be an invaluable aid in achieving the goal of adequate communication.”[6]

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.

PRESENTING THE RESULTS

The primary task in the research is the data presented in tabular form should first be discussed and explained. At the same time the researcher should avoid repeating the whole data that the table contains. While explaining the tables, it is always necessary to cite relevant column numbers and row numbers fro the convenience of the readers. The statistical materials presented in the tables should be systematically checked for accuracy. Besides explaining the tables, the researcher has to interpret the data on the basis of evidences. Such interpretation is the basis of generalization. Arrangement of ideas and the methodical and systematic presentation is an aid to draw the attention and evoke the interest of the readers. Heading should be specific, brief and represent the content. Such headings and sub-headings call the attention of readers as to the materials presented. There must be uniformity and consistency in presentation of ideas, punctuation, capitalization, abbreviations, footnotes, quotations, acknowledgement and bibliography, typical and summary sentences at the beginning end f the paragraph, capitalization of key phrases and sentences will make the presentation meaningful and logical.