Research Methods and Business Analytics Dr. C. Brahmaiah, Dr. D. Ch. Appa Rao, Md. Abdul Gaffar

RESEARCH METHODS AND BUSINESS ANALYTICS

KRISHNA UNIVERSITY
I M.COM - II SEMESTER
CBCS

Dr. C. Brahmaiah
Dr. D. Ch. Appa Rao
Md. Abdul Gaffar

RESEARCH METHODS AND BUSINESS ANALYTICS

KRISHNA UNIVERSITY, MACHILIPATNAM
CHOICE BASED CREDIT SYSTEM (CBCS)
M.Com,II SEMESTER *(With effect from 2016-17)*

Dr. C. Brahmaiah

Assistant Professor in Commerce

V.S.R. Govt. Degree & P.G. College

Movva, Krishna District

Dr.D.Ch. Appa Rao

Assistant Professor in Commerce

V.S.R. Govt. Degree & P.G. College

Movva, Krishna District

Md. Abdul Gaffar

Assistant Professor in Commerce

V.S.R. Govt. Degree & P.G. College

Movva , Krishna District

Published By

VRL

Publishers

First Edison: 2020

Price: 150/-

PREFACE

It gives us great pleasure to place the book ‘Research Methods and Business Analytics’ in your hands. The book is written strictly according to the M.Com. CBCS syllabus prescribed by Krishna University in Andhra Pradesh. The book has been thoroughly revised in the light of the new syllabus and latest examination question papers of different universities of Andhra Pradesh. In every chapter adequate theoretical explanation, short answer questions, essay answer questions and illustrations of M.Com, standard have been included to make the book more useful to the students. Sincere attempt is made to eliminate printing errors. we hope that the teachers and students will derive greater satisfaction from it.

We are thankful to Mr. Ch. kondal Rao, Mr.K. Brahma Reddy, Mr. M. Satyanarayana and M. Suniteja for the help rendered by them in publishing the book. We sincerely thanks to all lectures and students who have helped through by their valuable suggestions. We also thank Sri C.H. Anil Kumar VRL Publishers is a dedicated for the efficient and timely work done by them.

Constructive criticism and suggestions for improvement of the book are most welcome.

Movva Authors

Syllabus

Research Methods and Business Analytics

Unit – I: Introduction to Research

Nature and Scope of Research Methodology – Research design – Types of Research Problem

Formulation, Research Objectives – Hypotheses- Significance of Research in Commerce and

Management

Unit – II: Sources of Data and Sampling

Types of Sources: Primary and Secondary – Methods of Data Collection – Questionnaire –

Schedule-Observation - Attitude Measurement Techniques – Administration of Surveys –

Sample Design and Sampling Techniques.

Unit – III: Tabulation and Data Analysis

Tabulation and Cross Tabulation of Data: Univariate, Bivariate data - Analysis and Interpretation- Testing of Hypothesis-SPSS Packages and Applications.

Unit– IV: Multivariate Analysis

Advanced Techniques for Data Analysis: ANOVA, Discriminate Analysis, Factor Analysis,

Clustering Techniques, Report Writing.

Unit – V: Business Analytics

Evolution of Business Analytics - Master Data Management: Data Warehousing–Transformation and Up-loading of Data – Data Mining – Meta Data – Data Marts –Data Integration – OLTP and OLAP.

INDEX

Unit – III: Tabulation and Data Analysis

Unit– IV: Multivariate Analysis

Unit – V: Business Analytics

Unit – I: Introduction to Research

1.1. Introduction to Research:

The word research is composed of two syllables “Re” and “Search”. “Re” is the prefix meaning ‘Again or over again or a new’ and “Search” is the latter meaning ‘to examine closely and carefully’ or ‘to test and try’. Together they form, a careful, systematic, patient study and investigation in some field of knowledge undertaken to

establish principles / policies. Research normally is referred to as a search for knowledge. It is a systematic and scientific search for pertinent information in relation to a particular subject. It is an art of scientific and systematic investigation.

Research is an academic activity. It comprises of defining and redefining problems, formulating hypothesis or suggesting solutions, collecting, organizing and evaluating data, making deductions, reaching conclusions and sensibly testing the solutions to determine, whether they are in accordance to the hypothesis that have been formulated. Research is thus an original contribution to the existing stock of knowledge, making it appropriate for utilization. It is the pursuit of accuracy and facts, with the help of study, observation, comparison and experiment. It is the search for knowledge, through objective and systematic method of finding solutions to a problem.

1.2. **Meaning of Research**:

Research in common parlance refers to a search for knowledge. Once can also define research as a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation. The Advanced Learner's Dictionary of Current English lays down

the meaning of research as “a careful investigation or inquiry specially through search for new facts in any branch of knowledge.”

1.3. Definitions of Research:

Research work is traditionally defined as gathering of data to answer the questions and finding solution to problems. Some of the definitions of research are:

1. 1 Redman and Mory define research as “a systematized effort to gain new knowledge.”
2. According to Clifford Woody research comprises defining and redefining problems, formulating hypothesis or suggested solutions; collecting, organising and evaluating data; making deductions and reaching conclusions; and at last carefully testing the conclusions to determine whether they fit the formulating hypothesis.
3. D. Slesinger and M. Stephenson in the Encyclopaedia of Social Sciences define research as “the manipulation of things, concepts or symbols for the purpose of generalising to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art.”

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

1.4. Characteristics of Research:

The major characteristics of any research are;

1. Objectivity,
2. Precision,
3. Design and
4. Verifiability.

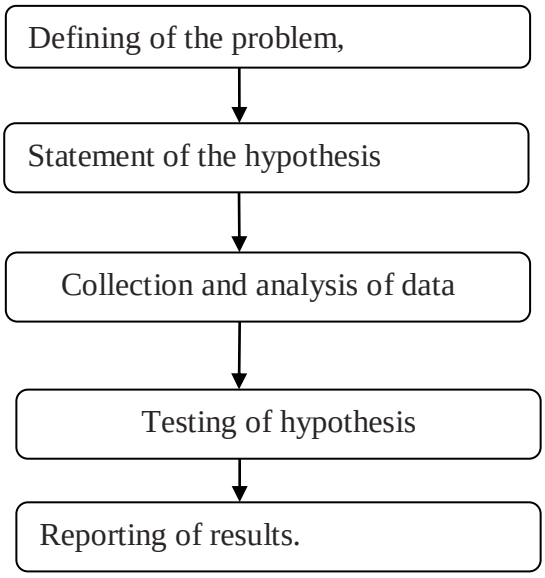
Let us look at these attributes more closely:

1. **Objectivity:** Ideally, research is beyond the subjective bias of the researcher. The researcher makes deliberate efforts to eliminate personal preference resisting the temptation to seek only such data that supports his/her hypothesis. The emphasis is on testing, rather than proving the hypothesis. The researcher is willing to suspend personal judgement and permit the data and logic to lead independently to a sound conclusion. Objectivity is achieved through standardisation of research instruments, choosing appropriate research

design and analytical tools and ensuring dependability of data.

2. **Precision:** Precision in scientific **r**esearch is achieved through the uses of statistical methods and techniques. As such, research conclusions convey the exact meaning to the reader, e.g. measures of central tendency, variability, correlation, regression etc. are the most precise expression in quantitative research which explains or represents the truth. Precise language describes the study accurately so that the study may be replicated or the results correctly used.
3. **Design:** In a scientific research, the researcher has to have a specified design of carrying out the investigation.

Process of research design



Only if the research has been carried out by using a specified process, it can be replicated for verification.

4. **Verifiability:** This is an important characteristic of every research. Research methods and findings presented to the professional community for other researchers to analyse, confirm or reject them. Verifiability is achieved primarily through two different approaches: first, analysing the same data on the same sample through alternative analytical tools (statistical methods), and the second, replicating the study on a different sample.

1.5. Nature and Scope of Research Methodology:

Research is systematic and critical investigation of a phenomenon. It identifies the variables, collects and analyses data on such variables to find answers to certain crucial questions. These answers contribute further to increase human knowledge. Orderliness is the hallmark of research. Research has to have an organic unity. This becomes essential if the knowledge which accrues from research is to be verified; for, it must be verifiable by anybody who takes the trouble to do so. In fact, research is considered to be a formal, systematic, intensive process of carrying on the scientific method of analysis. It involves a more systematic structure of

investigating, usually resulting in some sort of formal record of procedures and results or conclusions.

1.6. Research Design:

The most important step after defining the research problem is preparing the design of the research project, which is popularly known as the ‘research design’. A research design is a logical and systematic plan prepared for directing a research study. It specifies the objectives of the study, the methodology and techniques to be adopted for achieving the objectives.

The Research Design makes available the answers to various questions like

1) What is the design of a research?
2) What type of data is required?
3) What are the measures and scales used?
4) What is the form of data collected such as questionnaire, schedule, and interview?
5) What is the size of sample and sampling procedure?
6) What are the techniques of data analysis?

1.6.1. The Definition of a Research Design:

A research design is a plan, structure and strategy of investigation so conceived as to obtain answers to research

questions or problems. The plan is the complete scheme or program of the research. It includes an outline of what the investigator will do from writing the hypothesis and their operational implications to the final analysis of data. A traditional research design is blue print or detailed plan for how a research study is to be completed operational design variables so they can be measured, selecting a sample of interest to study. Collecting data to be used is a basis for testing hypothesis and analyzing the results. A research design is a procedural plan that is adopted by the researcher to answer questions validly objectively accurately and economically.

1. **According to Seltiz of at**, "A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose which economy in procedure".
2. **According to Kerlinger,** "Research design is the plan, structure and strategy and strategy of investigation conceived so as to obtain answers to research questions and to control variance"

3. **According to Bernard Philips** he research design "As a blue print for the collection, measurement and analysis of data".
4. **According to Claire Seltiz**, "A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure".
5. **According to S.L. Gupta and Hitesh**, "A research design provides a flow of activities from problem formulation to hypothesis development to data collection to data analysis to final results to implications".

1.6.2. Features of Research Design:

The important features of Research Design may be outlined as follows:

a) It is a plan that identifies the types and sources of information required for the research problem;
b) It is a strategy that specifies the methods of data collection and analysis which would be adopted; and
c) It also specifies the time period of research and monetary budget involved in conducting the study, which comprise the two major constraints of undertaking any research.

In view of the stated above research design decisions, the overall research design may be divided into the following (Kothari 1988):

a) The sampling design that deals with the method of selecting items to be observed for the selected study;
b) The observational design that relates to the conditions under which the observations are to be made;
c) The statistical design that concerns with the question of how many items are to be observed, and how the information and data gathered are to be analysed; and
d) The operational design that deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

1.6.3. Need for Research Design:

Research design is needed because it facilitates the smooth sailing of the various research operations,Thereby making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money. Just as for better, economical and attractive construction of a house, we need a blueprint (or what is commonly called the map of the house) well thought out and prepared by an expert architect, similarly we need a research

design or a plan in advance of data collection and analysis for our research project. Research design stands for advance planning of the methods to be adopted for collecting the relevant data and the techniques to be used in their analysis, keeping in view the objective of the research and the availability of staff, time and money. Preparation of the research design should be done with great care as any error in it may upset the entire project.Research design, in fact, has a great bearing on the reliability of the results arrived at and as such constitutes the firm foundation of the entire creation of the research work.

Even then the need for a well thought out research design is at times not realised by many. The importance which this problem deserves is not given to it. As a result many researches do not serve the purpose for which they are undertaken. In fact, they may even give misleading conclusions.Thoughtlessness in designing the research project may result in rendering the research exercise futile. It is, therefore, imperative that an efficient and appropriate design must be prepared before starting research operations. The design helps the researcher to organize his ideas in a form whereby it will be possible for him to look for flaws and

inadequacies. Such a design can even be given to others for their comments and critical evaluation. In the absence of such a course of action, it will be difficult for the critic to provide a comprehensive review of the proposed study.

1.6.4. Characteristics of a Good Research Design:

A good research design has the following traits:

1. Possesses the qualities of being flexible, suitable, efficient, and economical and so on.
2. It research design which minimizes bias and maximizes the reliability of the data collected and analysed (Kothari 1988).
3. Research design which does not allow even the smallest experimental error
4. Research design that yields maximum information and provides an opportunity of viewing the various dimensions of a research problem

Thus, the question of a good design relates to the purpose or objective and nature of the research problem studied. While a research design may be good, it may not be equally suitable to all studies. In other words, it may be lacking in one aspect or the other in the case of some other research problems.

Therefore, no single research design can be applied to all types of research problems.

If the research study happens to be an exploratory or a formulative one, wherein the major emphasis is on discovery of ideas and insights, the research design most appropriate must be flexible enough to permit the consideration of many different aspects of a phenomenon. But when the purpose of a study is accurate description of a situation or of an association between variables (or in what are called the descriptive studies), accuracy becomes a major consideration and a research design which minimises bias and maximises the reliability of the evidence collected is considered a good design.

Studies involving the testing of a hypothesis of a causal relationship between variables require a design which will permit inferences about causality in addition to the minimisation of bias and maximisation of reliability. But in practice it is the most difficult task to put a particular study in a particular group, for a given research may have in it elements of two or more of the functions of different studies.

It is only on the basis of its primary function that a study can be categorised either as an exploratory or descriptive or

hypothesis-testing study and accordingly the choice of a research design may be made in case of a particular study. Besides, the availability of time, money, skills of the research staff and the means of obtaining the information must be given due weightage while working out the relevant details of the research design such as experimental design, survey design, sample design and the like.

A research design suitable for a specific research problem would usually involve the following considerations:

- The methods of gathering the information;
- The skills and availability of the researcher and his/her staff, if any;
- The objectives of the research problem being studied;
- The nature of the research problem being studied; and
- The available monetary support and duration of time for the research work.

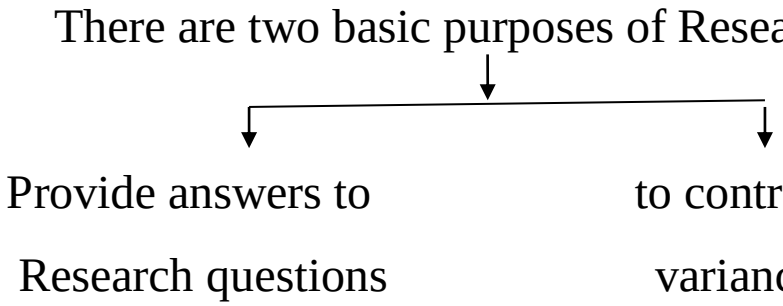
1.6.5. The nature of Research Design:

The Research Design is an outcome of decisions taken and put in a sequence in respect of the study and the types of data needed, the rational for the study being made, the sources

of data required, the study area, that time available for study, the material and the number cases need, the procedure of selection of the sample, the techniques of data Collection, the Analysis of data and the decisions articulated in a manner that the purpose of research is achieved with economy in money time and energy.

1. Convert the Research question and the stated assumptions/ hypotheses into operational variables that can be measured.
2. Specify the process that would be to follow to complete the task, as efficiently and economically as possible.
3. Specify the control mechanism that would be used to ensure that the effect of other variables that could impact the outcome of the study has been controlled.
4. Thus research designs have a critical and directive role to play in the research process. The execution details of the Research question to be investigated are referred to as the research design.

1.6.6. Purpose of Research Design



Research Design enables the researcher to answer research questions in a valid objective, accurate and economical way

- Research Design is worked out with care to yield reliable and valid answers to the research questions embodied by the hypotheses.
- Adequately planned and executed design assist greatly in allowing us to rely on both observation and inferences
- Research Design sets up the Framework for adequate tests of the relationship among variables

1.6.7. **Importance of Research Design:**

The need for a research design arises out of the fact that it facilitates the smooth conduct of the various stages of research. It contributes to making research as efficient as possible, thus yielding the maximum information with minimum effort, time and expenditure. A research design helps to plan in advance, the methods to be employed for collecting the relevant data and the techniques to be adopted

for their analysis. This would help in pursuing the objectives of the research in the best possible manner, provided the available staff, time and money are given. Hence, the research design should be prepared with utmost care, so as to avoid any error that may disturb the entire project. Thus, research design plays a crucial role in attaining the reliability of the results obtained, which forms the strong foundation of the entire process of the research work.

The research design helps the researcher to organize his/her ideas in a proper form, which in turn facilitates him/her to identify the inadequacies and faults in them.

A research without a pre-drawn plan is like an ocean voyage without Mariner's compass. The preposition of a research plan for a study aids in establishing direction to the study and in knowing exactly what has to be done and how and when it has to be done at every stage.

It enables the researcher to consider beforehand the various decisions to be made.

- What are the objectives of the study?
- What are the investigative questions?
- What are the sources of data?
- What is the universe of the study?

- What sampling method is appropriate? And so on.

Without a plan, research work becomes unfocussed and aimless empirical wandering the researcher would find it difficult, laborious and time-consuming to make adequate discriminations in the complex inter plan of facts before him, he may not be able to decide which is relevant and which is not, The use of a research design prevents such a blind search and indiscriminate gathering of data and guides him to proceed in the right direction.

1.6.8. Concepts Relating to Research Design:

Some of the important concepts relating to Research Design are discussed below:

1. Dependent and Independent Variables:

A variable is any characteristic, number, or quantity that can be measured or counted. A variable may also be called a data item. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, and eye colour and vehicle type are examples of variables.

2. Extraneous Variables:

The independent variables which are not directly related to the purpose of the study but affect the dependent variables, are known as extraneous variables. For instance, assume that a researcher wants to test the hypothesis that there is a relationship between children's school performance and their self-confidence, in which case the latter is an independent variable and the former, a dependent variable. In this context, intelligence may also influence the school performance.

3. Control:

One of the most important features of a good research design is to minimize the effect of extraneous variable(s). Technically, the term 'control' is used when a researcher designs the study in such a manner that it minimizes the effects of extraneous variables. The term 'control' is used in experimental research to reflect the restrain in experimental conditions.

4. Confounded Relationship:

The relationship between the dependent and independent variables is said to be confounded by an extraneous variable, when the dependent variable is not free from its effects.

5. Research Hypothesis:

When a prediction or a hypothesized relationship is tested by adopting scientific methods, it is known as research hypothesis. The research hypothesis is a predictive statement which relates to a dependent variable and an independent variable. Generally, a research hypothesis must consist of at least one dependent variable and one independent variable. Whereas, the relationships that are assumed but not to be tested are predictive statements that are not to be objectively verified, thus are not classified as research hypotheses.

6. Experimental and Non-Experimental Hypothesis Testing Research:

When the objective of a research is to test a research hypothesis, it is known as hypothesis-testing research. Such research may be in the nature of experimental design or non-experimental design. The research in which the independent variable is manipulated is known as 'experimental hypothesis-testing research', whereas the research in which the independent variable is not

manipulated is termed as ‘non-experimental hypothesis-testing research’.

7. Experimental and Control Groups:

When a group is exposed to usual conditions in an experimental hypothesis-testing research, it is known as ‘control group’. On the other hand, when the group is exposed to certain new or special condition, it is known as an ‘experimental group’. In the afore-mentioned example, Group A can be called as control group and Group B as experimental group. If both the groups, A and B are exposed to some special feature, then both the groups may be called as ‘experimental groups’. A research design may include only the experimental group or both the experimental and control groups together.

8. Treatments:

Treatments refer to the different conditions to which the experimental and control groups are subject to. In the example considered, the two treatments are the parents with regular earnings and those with no regular earnings. Likewise, if a research study attempts to examine through an experiment the comparative effect of three different types of fertilizers on the yield of rice crop, then

the three types of fertilizers would be treated as the three treatments.

9. Experiment:

Experiment refers to the process of verifying the truth of a statistical hypothesis relating to a given research problem. For instance, an experiment may be conducted to examine the yield of a certain new variety of rice crop developed. Further, Experiments may be categorized into two types, namely, 'absolute experiment' and 'comparative experiment'. If a researcher wishes to determine the impact of a chemical fertilizer on the yield of a particular variety of rice crop, then it is known as absolute experiment. Meanwhile, if the researcher wishes to determine the impact of chemical fertilizer as compared to the impact of bio-fertilizer, then the experiment is known as a comparative experiment.

10. Experimental Unit(s):

Experimental units refer to the pre-determined plots, characteristics or the blocks, to which different treatments are applied. It is worth mentioning here that

such experimental units must be selected with great caution.

1.6.9. Types of Research Design:

There are different types of research designs. They may be broadly categorized as:

(1) Exploratory Research Design;

(2) Descriptive and Diagnostic Research Design; and

(3) Hypothesis -Testing Research Design.

1. Exploratory Research Design:

The Exploratory Research Design is known as formulative research design. The main objective of using such a research design is to formulate a research problem for an in-depth or more precise investigation, or for developing a working hypothesis from an operational aspect. The major purpose of such studies is the discovery of ideas and insights. Therefore, such a research design suitable for such a study should be flexible enough to provide opportunity for considering different dimensions of the problem under study.

Usually, the following three methods are considered in the context of a research design for such studies. They are

(a) A survey of related literature;

(b) Experience survey; and

(c) Analysis of 'insight-stimulating' instances.

2. Descriptive and Diagnostic Research Design:

A Descriptive Research Design is concerned with describing the characteristics of a particular individual or a group. Meanwhile, a diagnostic research design determines the frequency with which a variable occurs or its relationship with another variable. In other words, the study analyzing whether a certain variable is associated with another comprises a diagnostic research study. On the other hand, a study that is concerned with specific predictions or with the narration of facts and characteristics related to an individual, group or situation, are instances of descriptive research studies.

The research design in such studies should be rigid and not flexible. Besides, it must also focus attention on the following:

a) Formulation of the objectives of the study,

b) Proper designing of the methods of data collection,

c) Sample selection,

d) Data collection,

e) Processing and analysis of the collected data, and

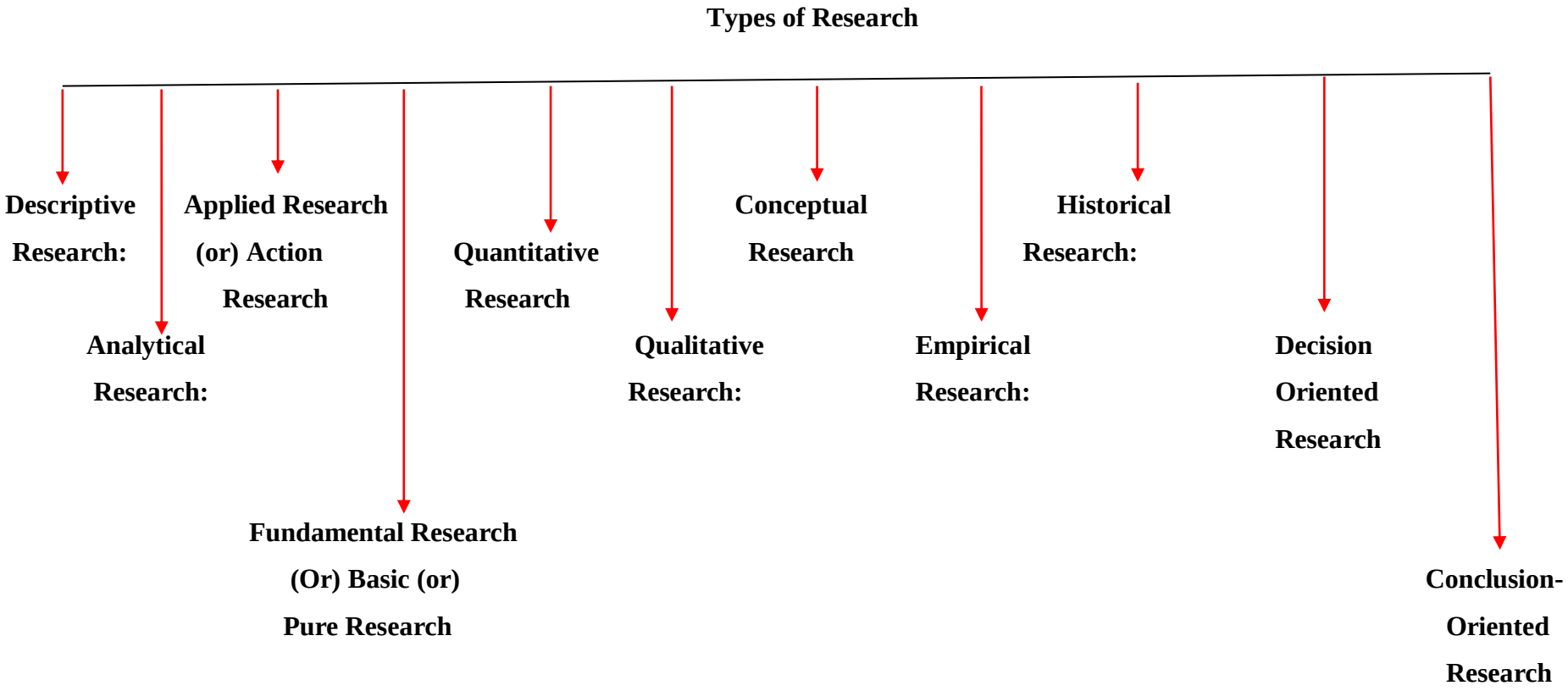
f) Reporting the findings.

3. Hypothesis-Testing Research Design:

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

1.7.Types of Research:

There are different types of research are as follows


Types of Research
Descriptive Research:
Analytical Research:
Applied Research (or) Action Research
Fundamental Research (Or) Basic (or) Pure Research
Quantitative Research
Qualitative Research:
Conceptual Research
Empirical Research:
Historical Research:
Decision Oriented Research
Conclusion-Oriented Research

1.7.1Descriptive Research:

Descriptive study is a fact-finding investigation with adequate interpretation. It is the simplest type of research. It is more specific that an exploratory study, as it has focus on particular aspects or dimensions of the problem studied.

Descriptive research consists of surveys and fact-finding enquiries of different types. The main objective of descriptive research is describing the state of affairs as it exists at the time of study. In social science and business research we quite often use. The term ‘ex post facto research’ is quite often used for descriptive research studies. The most distinguishing feature of this method is that the researcher has no control over the variables here. He/she has to only report what is happening or what has happened. Majority of the ex-post facto research projects are used for descriptive studies in which the researcher seeks to measure such items as, for example, the consumers’ preferences, frequency of purchases, shopping, etc.

1.7.2. Analytical Research:

The researcher has to use the already available facts or information, and analyse them to make a critical evaluation of

the subject.Analytical study is primarily concerned with testing hypothesis and specifying and interpreting relationships.

1.7.3. Applied Research (or) Action Research:

Applied research is carried on to find solution to a real life problem requiring an action or policy decision. It is thus problem oriented and action-directed. It seeks an immediate and practical result. The main objective of applied research is to find a solution to some pressing practical problem.

Applied research is an attempt to find a solution to an immediate problem encountered by a firm, an industry, a business organisation, or the society is known as applied research. Researchers engaged in such researches aim at drawing certain conclusions confronting a concrete social or business problem.

1.7.4. Fundamental Research (or) Basic (or) Pure Research:

Fundamental research mainly concerns generalizations and formulation of a theory. In other words, “Gathering knowledge for knowledge’s sake is termed ‘pure’ or ‘basic’ research” (Young in Kothari, 1988). Researches relating to pure mathematics or concerning some natural phenomenon are instances of Fundamental Research. Likewise, studies focusing

on human behaviour also fall under the category of fundamental research. The main objective of basic research is to find information with a broad base of application and add to the already existing organized body of scientific knowledge.

By developing principle, pure research offers solutions to many practical problems. For example, Maslow's theory of motivation serves as a guideline for formulating incentive schemes and approaches to motivating employees in organizations. Generalizations have many practical applications. In fact, nothing is as practical for the goals of diagnosis or treatment as good theoretical research.

Pure research develops many alternative solutions and thus enables us to choose the best solution. By applying scientific knowledge developed by pure researches, various appliances like radio, television, refrigerator, computer etc. have been invented. Continuous basic research in these fields has contributed to the manufacture of more effective and useful models at the least cost.

1.7.5. Quantitative Research :

Quantitative research is based on the measurement of quantity (or) amount. It is applicable to phenomena that can be

expressed in terms of quantity.Various available statistical and econometric methods are adopted for analysis in such research. Which includes correlation, regression and time series analysis etc.

1.7.6. Qualitative Research:

Qualitative research is concerned with qualitative phenomena, or more specifically, the aspects related to or involving quality or kind.

For example, an important type of qualitative research is 'Motivation Research', which investigates into the reasons for certain human behaviour.

The main aim of this type of research is discovering the underlying motives and desires of human beings by using in-depth interviews. The other techniques employed in such research are story completion tests, sentence completion tests, word association tests, and other similar projective methods.

Qualitative research is particularly significant in the context of behavioural sciences, which aim at discovering the underlying motives of human behaviour. Such research helps to analyse the various factors that motivate human beings to behave in a certain manner, besides contributing to an

understanding of what makes individuals like or dislike a particular thing. However, it is worth noting that conducting qualitative research in practice is considerably a difficult task. Hence, while undertaking such research, seeking guidance from experienced expert researchers is important.

1.7.7. Conceptual Research:

The research related to some abstract idea or theory is known as Conceptual Research. Generally, philosophers and thinkers use it for developing new concepts or for reinterpreting the existing ones.

1.7.8. Empirical Research:

Empirical Research relies on the observation or experience with hardly any regard for theory and system. Such research is data based, which often comes up with conclusions that can be verified through experiments or observation. Empirical research is also known as experimental type of research, in which it is important to first collect the facts and their sources, and actively take steps to stimulate the production of desired information. In this type of research, the researcher first formulates a working hypothesis, and then gathers sufficient facts to prove or disprove the stated hypothesis. He/she formulates the experimental

design, which according to him/her would manipulate the variables, so as to obtain the desired information.

This type of research is thus characterized by the researcher's control over the variables under study. In simple term, empirical research is most appropriate when an attempt is made to prove that certain variables influence the other variables in some way. Therefore, the results obtained by using the experimental or empirical studies are considered to be the most powerful evidences for a given hypothesis.

1.7.9. Historical Research:

Historical research As regards to historical research, sources like historical documents, remains, etc. Are utilized to study past events or ideas. It also includes philosophy of persons and groups of the past or any remote point of time.

1.7.10. Decision-Oriented Research:

The decision-oriented research is always carried out as per the need of a decision maker and hence, the researcher has no freedom to conduct the research according to his/her own desires.

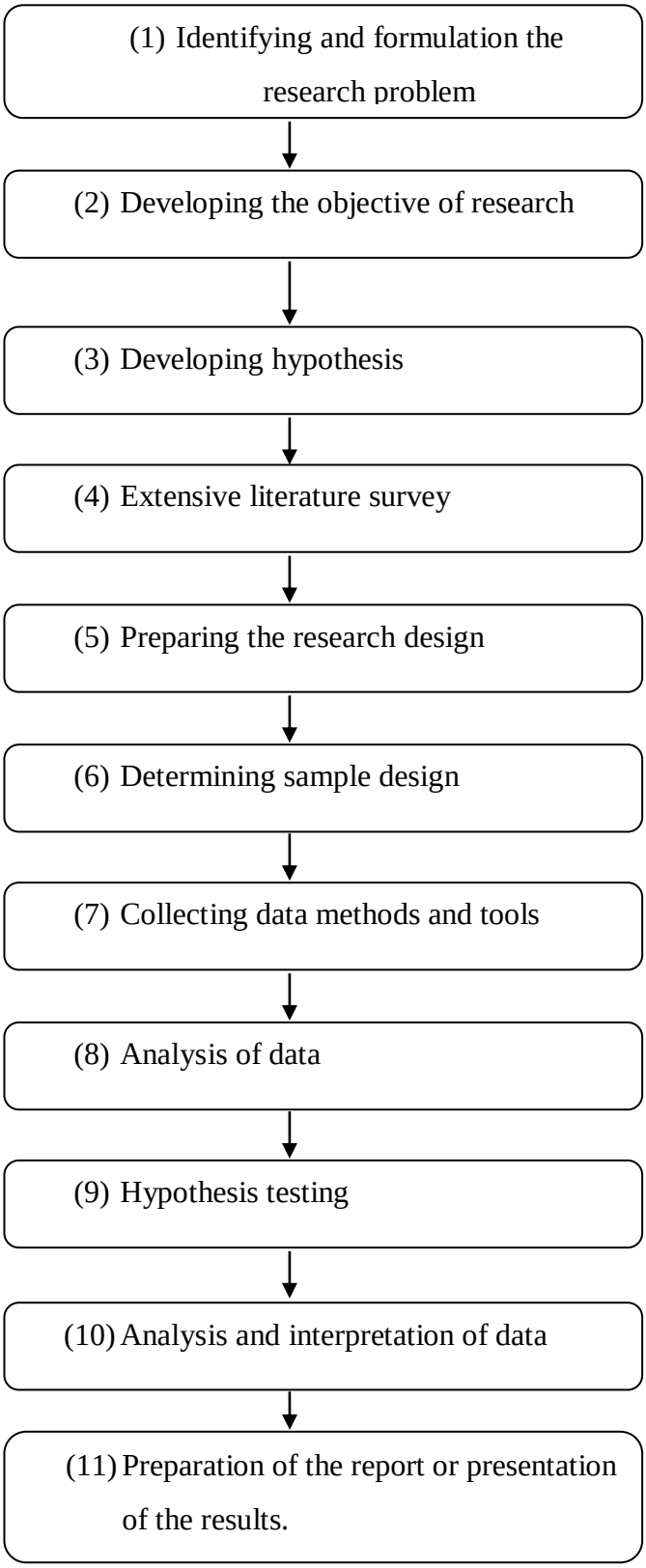
1.7.11. Conclusion-Oriented Research:

In the case of Conclusion-oriented research, the researcher is free to choose the problem, redesign the enquiry as it progresses and even change conceptualization as he/she wishes to. An operation research is a kind of decision-oriented research, where in scientific method is used in providing the departments, a quantitative basis for decision-making with respect to the activities under their purview.

1.8.Research Process:

Research process consists of a series of steps or actions required for effectively conducting research. The following are the steps that provide useful procedural guidelines regarding the conduct of research:

Research Process:



In other words, it involves the formal write-up of conclusions.

1.9. Research Problem:

The first and the foremost stage in the research process is to select and properly define the research problem. A researcher should first identify a problem and formulate it, so as to make it amenable or susceptible to research. In general, a research problem refers to an unanswered question that a researcher might encounter in the context of either a theoretical or practical situation, which he/she would like to answer or find a solution.

Thus, the components of a research problem may be summarized as:

i. There should be an individual or a group who have some difficulty or problem.
ii. There should be some objective(s) to be pursued. A person or an organization who wants nothing cannot have a problem.
iii. There should be alternative means the course of action for obtaining the objective(s) the researcher wants to attain. This implies that there should be more than one alternative means available to the researcher. This is because if the researcher

has no choice of alternative means, he/she would not have a problem.

iv. There should be some doubt in the mind of the researcher about the choice of alternative means. This implies that research should answer the question relating to the relative efficiency or suitability of the possible alternatives.

v. There should be a context to which the difficulty relates.

Thus, a research problem is one which requires a researcher to find out the best solution for thegiven problem, i.e., to find out by which course of action the objective can be attained optimally in the context of a given environment. There are several factors which may result in making the problem complicated.

For instance, the environment may change affecting the efficiencies of the courses of action or the values of the outcomes; the number of alternative courses of action may be very large; persons not involved in making the decision may be affected by it and react to it favourably or unfavourably, and similar other factors. All such elements (or at least the important ones) may be thought of in context of a research problem.

1.9.1. Technique Involved in Defining a Problem

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

(i) Statement of the problem in a general way;

(ii) Understanding the nature of the problem;

(iii) Surveying the available literature

(iv) Developing the ideas through discussions; and

(v) Rephrasing the research problem into a working proposition.

A brief description of all these points will be helpful.

(i) Statement of the problem in a general way: First of all the problem should be stated in a broad general way, keeping in view either some practical concern or some scientific or intellectual interest. For this purpose, the researcher must immerse himself thoroughly in the subject

matter concerning which he wishes to pose a problem. In case of social research, it is considered advisable to do some field observation and as such the researcher may undertake some sort of preliminary survey or what is often called *pilot survey*. Then the researcher can himself state the problem or he can seek the guidance of the guide or the subject expert in accomplishing this task.

(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. For a better understanding of the nature of the problem involved, he can enter into discussion with those who have a good knowledge of the problem concerned or similar other problems. The researcher should also keep in

view the environment within which the problem is to be studied and understood.

(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 means that the researcher must be well-conversant with relevant theories in the field, reports and records as also all other relevant literature. He must devote sufficient time in reviewing of research already undertaken on related problems. This is done to find out what data and other materials, if any, are available for operational purposes. "Knowing what data are available often serves to narrow the problem itself as well as the technique that might be used."

This would also help a researcher to know if there are certain gaps in the theories, or whether the existing theories applicable to the problem under study are in consistent with each other, or whether the findings of the different studies do not follow a pattern consistent with the theoretical expectations and so on.

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

(v) Rephrasing the research problem: Finally, the researcher must sit to rephrase the research problem into a working proposition. Once the nature of the problem has been clearly understood, the environment (within which the

problem has got to be studied) has been defined, discussions over the problem have taken place and the available literature has been surveyed and examined, rephrasing the problem into analytical or operational terms is not a difficult task. Through rephrasing, 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.

1.10. Objectives of Research:

The objective of research is to find answers to the questions by applying scientific procedures. In other words, the main aim of research is to find out the truth which is hidden and has not yet been discovered. Although every research study has its own specific objectives, the research objectives may be broadly grouped as follows:

1. To gain familiarity with new insights into a phenomenon (i.e.,exploratoryor formulative research studies);
2. To accurately portray the characteristics of a particular individual, group, or a situation (i.e., descriptive research studies);

3. To analyse the frequency with which something occurs (i.e., diagnostic research studies); and
4. To test a hypothesis of a causal relationship between two variables (such studies are known as hypothesis-testing research studies).

1.11. Hypothesis:

Hypothesis may be defined as "a proposition or a set of propositions set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional meet to guide some investigation in the light of established facts" (Kothari, 1988). A research hypothesis is quite often a predictive statement, which is capable of being tested using scientific methods that involve an independent and some dependent variables. For instance, the following statements may be considered:

i. "Students who take tuitions perform better than the others who do not receive tuitions" or,

ii. "The female students perform as well as the male students".

These two statements are hypotheses that can be objectively verified and tested. Thus, they indicate that a hypothesis states

what one is looking for. Besides, it is a proposition that can be put to test in order to examine its validity.

1.11.1 Characteristics of Hypothesis:

A hypothesis should have the following characteristic features:-

i. A hypothesis must be precise and clear. If it is not precise and clear, then the inferences drawn on its basis would not be reliable.

ii. A hypothesis must be capable of being put to test. Quite often, the research programmes fail owing to its incapability of being subject to testing for validity.

iii. A hypothesis must state relationship between two variables, in the case of relational hypotheses.

iv. A hypothesis must be specific and limited in scope.

v. As far as possible, a hypothesis must be stated in the simplest language, so as to make it understood by all concerned.

vi. A hypothesis must be consistent and derived from the most known facts. In other words, it should be consistent with a substantial body of established facts. That is, it must be in the form of a statement which is most likely to occur.

vii. A hypothesis must be agreeable to testing within a stipulated or reasonable period of time.

viii. A hypothesis should state the facts that give rise to the necessity of looking for an explanation. A hypothesis should explain what it actually wants to explain, and for this it should also have an empirical reference.

1.11.2. Concepts Relating to Testing of Hypotheses:

Testing of hypotheses requires a researcher to be familiar with various concepts concerned with it such as:

1) Null Hypothesis and Alternative Hypothesis:

In the context of statistical analysis, hypotheses are of two types viz., null hypothesis and alternative hypothesis. When two methods A and B are compared on their relative superiority, and it is assumed that both the methods are equally good, then such a statement is called as the null hypothesis. On the other hand, if method A is considered relatively superior to method B, or vice-versa, then such a statement is known as an alternative hypothesis. The null hypothesis is expressed as **H**0, while the alternative hypothesis is expressed as **H**1.

For example, if a researcher wants to test the hypothesis that the population mean (μ) is equal to the hypothesized mean (H0) = 100, then the null hypothesis should be stated

as the population mean is equal to the hypothesized mean 100. Symbolically it may be written as:-

H0: = μ = μ H0 = 100

If sample results do not support this null hypothesis, then it should be concluded that something else is true. The conclusion of rejecting the null hypothesis is called as alternative hypothesis H1. To put it in simple words, the set of alternatives to the null hypothesis is termed as the alternative hypothesis. If H0 is accepted, then it implies that Ha is being rejected. On the other hand, if H0 is rejected, it means that Ha is being accepted. For H0: μ = μ H0 = 100, the following three possible alternative hypotheses may be considered:

Alternative hypothesis	**To be read as follows**
H1: $\mu \neq \mu$ H0	The alternative hypothesis is that the population mean is not equal to 100, i.e., it could be greater than or less than 100
H1 : $\mu > \mu$ H0	The alternative hypothesis is that the population mean is greater than 100
H1 : $\mu < \mu$ H0	The alternative hypothesis is that the population mean is less than 100

Before the sample is drawn, the researcher has to state the null hypothesis and the alternative hypothesis. While formulating the null hypothesis, the following aspects need to be considered:

A. Alternative hypothesis is usually the one which a researcher wishes to prove, whereas the null hypothesis is the one which he/she wishes to disprove. Thus, a null hypothesis is usually the one which a researcher tries to reject, while an

alternative hypothesis is the one that represents all other possibilities.

B. The rejection of a hypothesis when it is actually true involves great risk, as it indicates that it is a null hypothesis because then the probability of rejecting it when it is true is α (i.e., the level of significance) which is chosen very small.

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

2) The Level of Significance:

In the context of hypothesis testing, the level of significance is a very important concept. It is a certain percentage that should be chosen with great care, reason and insight. If for instance, the significance level is taken at 5 per cent, then it means that H0 would be rejected when the sampling result has a less than 0.05 probability of occurrence when H0 is true. In other words, the five per cent level of significance implies that the researcher is willing to take a risk of five per cent of rejecting the null hypothesis, when (H0) is actually true. In sum, the significance level reflects the maximum value of the probability of rejecting H0 when it is actually true, and which is usually determined prior to testing the hypothesis.

3) Test of Hypothesis or Decision Rule:

Suppose the given hypothesis is H0 and the alternative hypothesis H1, then the researcher has to make a rule known as the decision rule. According to the decision rule, the researcher accepts or rejects H0. For example, if the H0 is that certain students are good against the H1 that all the students are good, then the researcher should decide the number of items to be tested and the criteria on the basis of which to accept or reject the hypothesis.

4) Type I And Type II Errors:

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

5) One-Tailed and Two-Tailed Tests:

These two types of tests are very important in the context of hypothesis testing. A two-tailed test rejects the null hypothesis, when the sample mean is significantly greater or lower than the hypothesized value of the mean of the population. Such a test is suitable when the null hypothesis is some specified value; the alternative hypothesis is a value that is not equal to the specified value of the null hypothesis.

Procedure of Hypothesis Testing:

The procedure of hypothesis testing includes all the steps that a researcher undertakes for making a choice between the two alternative actions of rejecting or accepting a null hypothesis.

The various steps involved in hypothesis testing are as follows:

1) Making a Formal Statement:

This step involves making a formal statement of the null hypothesis (H0) and the alternative hypothesis (Ha). This implies that the hypotheses should be clearly stated within the purview of the research problem. For example, suppose a school teacher wants to test the understanding capacity of the students

which must be rated more than 90 per cent in terms of marks, the hypotheses may be stated as follows:

Null Hypothesis H0: = 100 Alternative HypothesisH1 :> 100

2) Selecting a Significance Level:

The hypotheses should be tested on a pre-determined level of significance, which should be specified. Usually, either 5% level or 1% level is considered for the purpose. The factors that determine the levels of significance are:

(a) The magnitude of difference between the sample means;

(b) The sample size:

(c) The variability of measurements within samples; and

(d) Whether the hypothesis is directional or non-directional (Kothari, 1988).

In sum, the level of significance should be sufficient in the context of the nature and purpose of enquiry.

3) Deciding the Distribution to Use:

After making decision on the level of significance for hypothesis testing, the researcher has to next determine the appropriate sampling distribution. The choice to be made generally relates to normal distribution and the t-distribution. The rules governing

the selection of the correct distribution are similar to the ones with respect to estimation.

4) Selection of a Random Sample and Computing an Appropriate Value:

Another step involved in hypothesis testing is the selection of a random sample and then computing a suitable value from the sample data relating to test statistic by using the appropriate distribution. In other words, it involves drawing a sample for furnishing empirical data.

5) Calculation of the Probability:

The next step for the researcher is to calculate the probability that the sample result would diverge as far as it can from expectations, under the situation when the null hypothesis is actually true.

6) Comparing the Probability:

Another step involved consists of making a comparison of the probability calculated with the specified value of α, i.e. the significance level. If the calculated probability works out to be equal to or smaller than the α value in case of one-tailed test, then the null hypothesis is to be rejected. On the other hand, if the calculated probability is greater, then the null hypothesis is

to be accepted. In case the null hypothesis H0 is rejected, the researcher runs the risk of committing the Type I error. But if the null hypothesis H0 is accepted, then it involves some risk (which cannot be specified in size as long as H0 is vague and not specific) of committing the Type II error.

1.12. Significance of Research in Commerce and Management:

According to a famous Hudson Maxim, "All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry, and inquiry leads to invention". It brings out the significance of research, increased amount of which makes the progress possible.

1. Research encourages scientific and inductive thinking, besides promoting the development of logical habits of thinking and organisation.
2. The role of research in applied economics in the context of an economy or business is greatly increasing in modern times.
3. The increasingly complex nature of government and business has raised the use of research in solving operational problems.

4. Research assumes significant role in the formulation of economic policy for both, the government and business.
5. It provides the basis for almost all government policies of an economic system. Government budget formulation, for example, depends particularly on the analysis of needs and desires of people, and the availability of revenues, which requires research.
6. Research helps to formulate alternative policies, in addition to examining the consequences of these alternatives.
7. Research also facilitates the decision-making of policy-makers, although in it is not a part of research. In the process,
8. Research also helps in the proper allocation of a country's scarce resources.
9. Research is also necessary for collecting information on the social and economic structure of an economy to understand the process of change occurring in the country.

10. research as a tool of government economic policy formulation involves three distinct stages of operation:

Research is equally important to social scientists for analyzing the social relationships and seeking explanations to various social problems. It gives intellectual satisfaction of knowing things for the sake of knowledge. It also possesses the practical utility for the social scientist to gain knowledge so as to be able to do something better or in a more efficient manner. The research in social sciences is concerned with both knowledge for its own sake, and knowledge for what it can contribute to solve practical problems.

Short Questions

1. Objectives of research
2. Qualities of a good research.
3. Social research.
4. Research methodology
5. Scope of research in business decision making.
6. What is statistical hypothesis
7. Research approaches
8. Importance of research

9. Need for research
10. Research problem

Essay Questions

1. What is Research Design? Write the methodology of a descriptive study.
2. Define hypothesis. Explain the procedure of testing hypotheses.
3. What is research explain where is types of research
4. What is research? Explain the difference between descriptive and experimental research.
5. What is a research problem? Discuss how a research problem is formulated.
6. Discuss the examples of exploratory research, descriptive and experimental research
7. What is research? Explain the research process in detail

Unit – II: Sources of Data and Sampling

2.1. Data Meaning:

Information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer:

2.2. Types of Sources:

The task of data collection begins after a research problem has been defined and research design/plan chalked out. It is important for a researcher to know the sources of data which he requires for different purposes. Data are nothing but the information. There are two sources of information or data they are

2.2.1. Primary data: The primary data which was collected afresh and for the first time, and thus happen to be original in character. The data are name after the source. Primary data refers to the data collected for the first time.

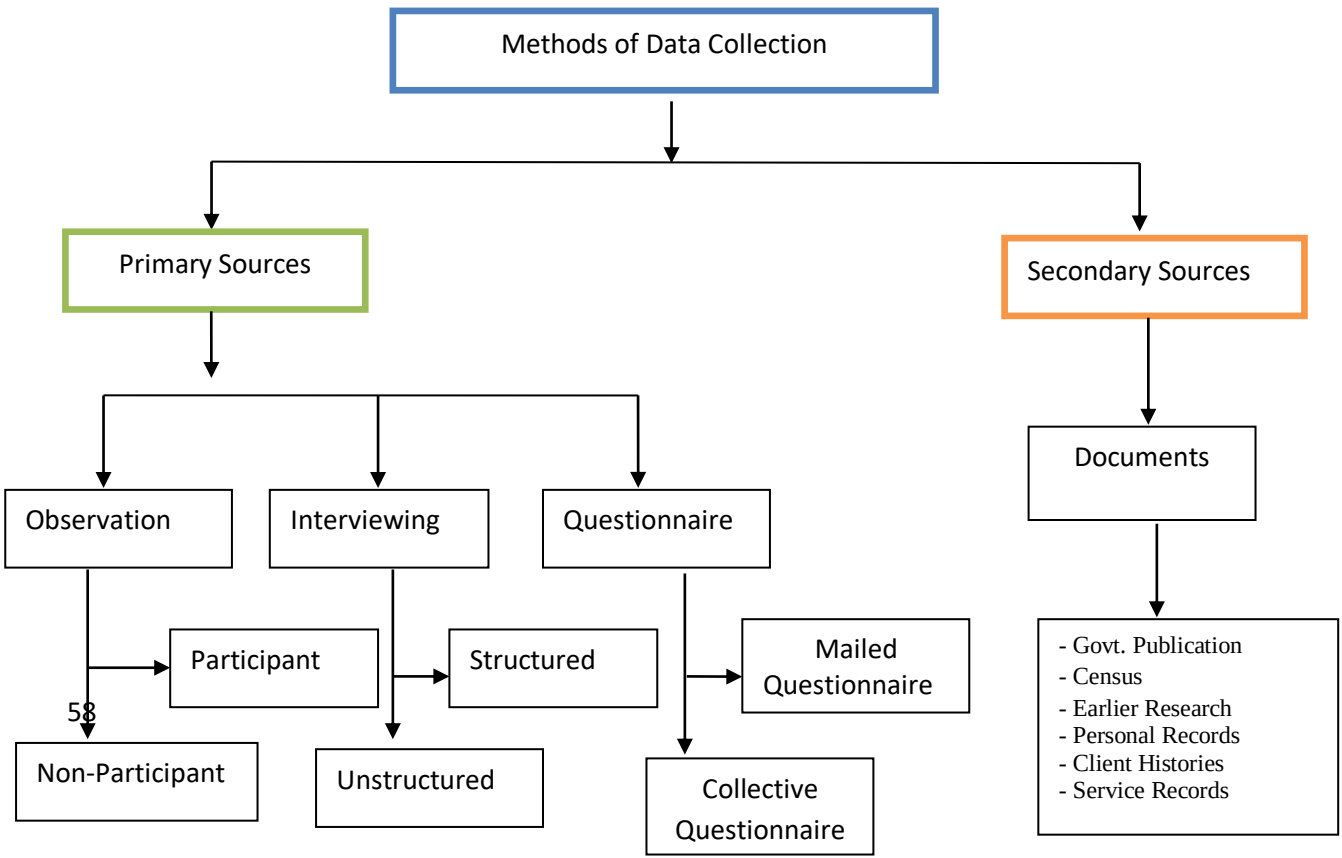
2.2.2. **Secondary data:** Secondary data refers to the data that have already been collected and used earlier by somebody or some agency and which have already been passed through the statistical process. The researcher would have to decide which sort of data he would be using (thus collecting) for his study and accordingly he will have to select one or the other method of data collection.

For example, the statistics collected by the Government of India relating to the population is primary data for the

Government of India since it has been collected for the first time. Later when the same data are used by a researcher for his study of a particular problem, then the same data become the secondary data for the researcher.

2.3. Methods of Data Collection:

There are two major approaches to gathering information about a situation, person, problem or phenomenon. When you undertake a research study, in most situations, you need to collect the required information; however, sometimes the information required is already available and need only be extracted. Based upon these broad approaches to information gathering, data can be categorized as:



Primary Sources:

Primary data is one which is collected by the investigator himself/herself first time for the purpose of a specific inquiry or study. Such data is original in character and is generated by surveys conducted by individual or research institutions**.**

2.3.1. Questionnaire:

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

- According to Bogardus, “A questionnaire is a list of questions sent to a number of persons for their answers and

which obtains standardized results that can be tabulated and treated statistically".

- The Dictionary of Statistical Terms defines it as a "Group of or sequence of questions designed to elicit information upon a subject or sequence of subjects from information."

A questionnaire should be designed or drafted with utmost care and caution so that all the relevant and essential information for the enquiry may be collected without any difficulty, ambiguity and vagueness. Drafting of a good questionnaire is a highly specialized job and requires great care skill, wisdom, efficiency and experience. No hard and fast rule can be laid down for designing or framing a questionnaire.

- **The mailed questionnaire:**– The most common approach to collecting information is to send the questionnaire to prospective respondents by mail. Obviously this approach presupposes that you have access to their addresses. Usually it is a good idea to send a prepaid, self-addressed envelope with the questionnaire as this might increase the response rate. A mailed questionnaire *must* be accompanied by a covering letter (see below for details). One of the

major problems with this method is the low response rate. In the case of an extremely low response rate, the findings have very limited applicability to the population studied.

- **Collective administration:**– One of the best ways of administering a questionnaire is to obtain a captive audience such as students in a classroom, people attending a function, participants in a programme or people assembled in one place. This ensures a very high response rate as you will find few people refuse to participate in your study. Also, as you have personal contact with the study population, you can explain the purpose, relevance and importance of the study and can clarify any questions that respondents may have. The author's advice is that if you have a captive audience for your study, don't miss the opportunity – it is the quickest way of collecting data, ensures a very high response rate and saves you money on postage.

The following general points designing or framing a questionnaire may be borne in mind:

1. Size of the Questionnaire should be small:

A researcher should try his best to keep the number of questions as small as possible, keeping in view the nature, objectives and scope of the enquiry. Respondent's time should not be wasted by asking irrelevant and unimportant questions. A large number of questions would involve more work for the investigator and thus result in delay on his part in collecting and submitting the information. A reasonable questionnaire should contain from 15 to 25 questions at large. If a still larger number of questions are a must in any enquiry, then the questionnaire should be divided into various sections or parts.

2. The questions should be clear:

The questions should be easy, brief, unambiguous, non-offending, and courteous in tone, corroborative in nature and to the point, so that much scope of guessing is left on the part of the respondents.

3. The questions should be arranged in a logical sequence:

Logical arrangement of questions reduces lot of unnecessary work on the part of the researcher because it

not only facilitates the tabulation work but also does not leave any chance for omissions or commissions.

4. Questions should be simple to understand:

The vague words like good, bad, efficient, sufficient, prosperity, rarely, frequently, reasonable, poor, rich etc., should not be used since these may be interpreted differently by different persons and as such might give unreliable and misleading information. Similarly the use of words having double meaning like price, assets, capital income etc., should also be avoided.

5. Questions should be comprehensive & easily answerable:

Questions should be designed in such a way that they are readily comprehensible and easy to answer for the respondents. They should not be tedious nor should they tax the respondents' memory. At the same time questions involving mathematical calculations like percentages, ratios etc., should not be asked.

6. Questions of personal & sensitive nature should not be asked:

There are some questions which disturb the respondents and he/she may be shy or irritated by hearing such questions. Therefore, every effort should be made to avoid such questions. For example, ‘do you cook yourself or your wife cooks?’ ‘Or do you drink?’ Such questions will certainly irk the respondents and thus be avoided at any cost. If unavoidable then highest amount of politeness should be used.

7. Types of Questions:

Under this head, the questions in the questionnaire may be classified as follows:

(a) Shut Questions:

Shut questions are those where possible answers are suggested by the framers of the questionnaire and the respondent is required to tick one of them. Shut questions can further be subdivided into the following forms:

(i) Simple Alternate Questions:

In this type of questions the respondent has to choose from the two clear cut alternatives like ‘Yes’ or ‘No’, ‘Right or Wrong’ etc. Such questions are also called as

dichotomous questions. This technique can be applied with elegance to situations where two clear cut alternatives exist.

(ii) Multiple Choice Questions:

Many a times it becomes difficult to define a clear cut alternative and accordingly in such a situation additional answers between Yes and No, like Do not know, No opinion, Occasionally, Casually, Seldom etc., are added.

For example, in order to find if a person smokes or drinks, the following multiple choice answers may be used:

Do you smoke?

(a) Yes regularly [ ] (b) No never [ ]

(c) Occasionally [ ] (d) Seldom [ ]

Multiple choice questions are very easy and convenient for the respondents to answer. Such questions save time and also facilitate tabulation. This method should be used if only a selected few alternative answers exist to a particular question.

8. Leading questions should be avoided:

Questions like ‘why do you use a particular type of car, say Maruti Suzuki Dzire car’ should preferably be framed into two questions-

(i) Which car do you use? (ii) Why do you prefer it?

It gives smooth ride [ ]

It gives more mileage [ ]

It is cheaper [ ]

It is maintenance free [ ]

9 Cross Checks:

The questionnaire should be so designed as to provide internal checks on the accuracy of the information supplied by the respondents by including some connected questions at least with respect to matters which are fundamental to the enquiry.

10 Pre Testing the Questionnaire:

It would be practical in every sense to try out the questionnaire on a small scale before using it for the given enquiry on a large scale. This has been found extremely useful in practice. The given questionnaire can be improved or modified in the light of the drawbacks, shortcomings and problems faced by the investigator in the pre test.

11 A Covering Letter:

A covering letter from the organizers of the enquiry should be enclosed along with the questionnaire for the purposes regarding definitions, units, concepts used in the questionnaire, for taking the respondent's confidence, self-addressed envelope in case of mailed questionnaire, mention about award or incentives for the quick response, a promise to send a copy of the survey report etc.

2.3.2. Observation: The observation method is the most commonly used method especially in studies relating to behavioral sciences. This method implies the collection of information by way of investigator's own observation, without interviewing the respondents. The information obtained relates to what is currently happening and is not complicated by either the past behavior or future intentions or attitudes of respondents. This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

Observation is one way to collect primary data. Observation is a purposeful, systematic and selective way of watching and listening to an interaction or phenomenon as it

takes place. There are many situations in which observation is the most appropriate method of data collection; for example, when you want to learn about the interaction in a group, study the dietary patterns of a population, ascertain the functions performed by a worker, or study the behavior or personality traits of an individual. It is also appropriate in situations where full and/or accurate information cannot be elicited by questioning, because respondents either are not co-operative or are unaware of the answers because it is difficult for them to detach themselves from the interaction. In summary, when you are more interested in the behavior than in the perceptions of individuals, or when subjects are so involved in the interaction that

They are unable to provide objective information about it, observation is the best approach to collect the required information.

Types of observation

There are two types of observation:

1. Participant Observation;
2. Non-Participant Observation.

1. Participant Observation is when you, as a researcher, participate in the activities of the group being observed in the same manner as its members, with or without their knowing that they are being observed.

For example, you might want to examine the reactions of the general population towards people in wheelchairs. You can study their reactions by sitting in a wheelchair yourself. Or you might want to study the life of prisoners and pretend to be a prisoner in order to do this.

2. Non-participant observation, on the other hand, is when you, as a researcher, do not get involved in the activities of the group but remain a passive observer, watching and listening to its activities and drawing conclusions from this.

For example, you might want to study the functions carried out by nurses in a hospital. As an observer, you could watch, follow and record the activities as they are performed. After making a number of observations, conclusions could be drawn about the functions nurses carry out in the hospital. Any occupational group in any setting can be observed in the same manner.

2.3.3. Interview Method: The interview method of collecting data involves presentation of oral-verbal stimuli and reply interns of oral-verbal responses. This method can be used through personal interviews and, if possible, through telephone interviews.

(a) Personal interviews: Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons. (At times the interviewee may also ask certain questions and the interviewer responds to these, but usually the interviewer initiates the interview and collects the information.) This sort of interview may be in the form of direct personal investigation or it may be indirect oral investigation. In the case of direct personal investigation the interviewer has to collect the information personally from the sources concerned. He has to be on the spot and has to meet people from whom data have to be collected.

This method is particularly suitable for intensive investigations. But in certain cases it may not be possible or worthwhile to contact directly the persons concerned or on account of the extensive scope of enquiry, the direct

personal investigation technique may not be used. In such cases an indirect oral examination can be conducted under which the interviewer has to cross-examine other persons who are supposed to have knowledge about the problem under investigation and the information, obtained is recorded. Most of the commissions and committees appointed by government to carry on investigations make use of this method.

The method of collecting information through personal interviews is usually carried out in structured way.

As such we call the interviews as *structured interviews.* Such interviews involve the use of a set of predetermined questions and of highly standardized techniques of recording. Thus, the interviewer in a structured interview follows a rigid procedure laid down, asking questions in a form and order prescribed. As against it, the **unstructured interviews**are characterized by a flexibility of approach to questioning.

Unstructured interviews do not follow a system of pre-determined questions and standardized techniques of recording information. In a non-structured interview, the

interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so requires. He may even change the sequence of questions. He has relatively greater freedom while recording the responses to include some aspects and exclude others. But this sort of flexibility results in lack of comparability of one interview with another and the analysis of unstructured responses becomes much more difficult and time-consuming than that of the structured responses obtained in case of structured interviews.

Unstructured interviews also demand deep knowledge and greater skill on the part of the interviewer. Unstructured interview, however, happens to be the central technique of collecting information in case of exploratory or formulate research studies. But in case of descriptive studies, we quite often use the technique of structured interview because of its being more economical, providing a safe basis for generalization and requiring relatively lesser skill on the part of the interviewer.

We may as well talk about focused interview, clinical interview and the non-directive interview.

Focused interview: is meant to focus attention on the given experience of the respondent and its effects. Under it the interviewer has the freedom to decide the manner and sequence in which the questions would be asked and has also the freedom to explore reasons and motives. The main task of the interviewer in case of a focused interview is to confine the respondent to a discussion of issues with which he seeks conversance. Such interviews are used generally in the development of hypotheses and constitute a major type of unstructured interviews.

The **clinical interview**is concerned with broad underlying feelings or motivations or with the course of individual's life experience. The method of eliciting information under it is generally left to the interviewer's discretion. In case of **non-directive interview,** the interviewer's function is simply to encourage the respondent to talk about the given topic with a bare minimum of direct questioning.

The interviewer often acts as a catalyst to a comprehensive expression of the respondents' feelings and beliefs and of

the frame of reference within which such feelings and beliefs take on personal significance.

Despite the variations in interview-techniques, the major advantages and weaknesses of personal interviews can be enumerated in a general way.

(A) **Merits of the interview method:** The chief merits of the interview method are as follows:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Merits of Telephone interviews: The main merits of such a system are:

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

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

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

7. Interviewer can explain requirements more easily.

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

9. No field staff is required.

10. Representative and wider distribution of sample is possible.

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

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

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

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

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

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

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

2.3.4. Collection of Data through Schedules

This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose.

This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected. The enumerators should be trained to perform their job well and the nature and scope of the investigation should be explained to them thoroughly so that they may well understand the implications of different questions put in the schedule.

This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however,

very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organizations. Population census all over the world is conducted through this method.

Distinction (or) Difference between Questionnaires and Schedules:

S.No.	Points of Difference	Questionnaire	Schedule
1	Nature of Data Collection	Indirect method (Impersonal) sent through mail	Direct Method i.e., Researcher himself / herself Collection the schedule
2	Cost	Relatively cheap as it money only preparation of questionnaire and mailing the same	More expensive as it money not only preparation of schedule but also appoint of enumerators
3	Non-response	High	Low
4	Target Respondent	Clearly known	Clearly not known
5	Time	very slow	Speed (well in time)
6	Personal Contact	Not Possible	Possible
7	Applicability	Literate	Illiterate
8	Scope	Wider	Narrow
9	Risk of Collecting Incomplete Data	More	Very less
10	Success	Depends on quality of questionnaire itself	Depends on honesty and competence of enumerators
11	Appearance (physical) of questionnaire	Attractive should be	Not necessary

12	Combination of methods	No Combination possible	Observation method also possible along with questionnaires

2.4. Measurement in Research:

In our daily life we are said to measure when we use some yardstick to determine weight, height, or some other feature of a physical object. We also measure when we judge how well we like a song, a painting or the personalities of our friends. We, thus, measure physical objects as well as abstract concepts. Measurement is a relatively complex and demanding task, specially so when it concerns qualitative or abstract phenomena. By measurement we mean the process of assigning numbers to objects or observations, the level of measurement being a function of the rules under which the numbers are assigned.

It is easy to assign numbers in respect of properties of some objects, but it is relatively difficult in respect of others. .

Technically speaking, measurement is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence. In measuring, we devise some form of scale in the range (in terms of set theory, range may refer to some set) and then transform or map the

properties of objects from the domain (in terms of set theory, domain may refer to some other set) onto this scale.

For example, in case we are to find the male to female attendance ratio while conducting a study of persons who attend some show, then we may tabulate those who come to the show according to sex. In terms of set theory, this process is one of mapping the observed physical properties of those coming to the show (the domain) on to a sex classification (the range). The rule of correspondence is: If the object in the domain appears to be male, assign to "0" and if female assign to "1".

Similarly, we can record a person's marital status as 1, 2, 3 or 4, depending on whether the person is single, married, widowed or divorced. We can as well record "Yes or No" answers to a question as "0" and "1" (or as 1 and 2 or perhaps as 59 and 60). In this artificial or nominal way, categorical data (qualitative or descriptive) can be made into numerical data and if we thus code the various categories, we refer to the numbers we record as nominal data. *Nominal data* are numerical in name only, because they do not share any of the properties of the numbers we deal in ordinary arithmetic.

For instance if we record marital status as 1, 2, 3, or 4 as stated above, we cannot write $4 > 2$ or $3 < 4$ and we cannot write $3 - 1 = 4 - 2$, $1 + 3 = 4$ or $4 \div 2 = 2$. In those situations when we cannot do anything except set up inequalities, we refer to the data as ordinal data.

2.4.1 Measurement Scales:

From what has been stated above, we can write that scales of measurement can be considered in terms of their mathematical properties.

The most widely used classification of measurement scales are:

(a) Nominal scale

(b) Ordinal scale

(c) Interval scale and

(d) Ratio scale.

(a) Nominal scale: Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as

such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved.

For example, one cannot usefully average the numbers on the back of a group of football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one group with the numbers assigned to another. The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measure of dispersion for nominal scales.

Chi-square test is the most common test of statistical significance that can be utilized, and for the measures of correlation, the contingency coefficient can be worked out. Nominal scale is the least powerful level of measurement. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes differences between things by assigning them to categories. Nominal data are, thus, counted data. The scale wastes any information that we may have about varying degrees of

attitude, skills, understandings, etc. In spite of all this, nominal scales are still very useful and are widely used in surveys and other *ex-post-facto* research when data are being classified by major sub-groups of the population.

(b) Ordinal scale: The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statement about scores based on ordinal scales. For instance, if Ram's position in his class is 10 and Mohan's position is 40, it cannot be said that Ram's position is four times as good as that of Mohan. The statement would make no sense at all.

Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise

comparisons cannot be made. Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less.

The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

(c) Interval scale: In the case of interval scale, the intervals are adjusted in terms of some rule that has been established as a basis for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based. Interval scales can have an arbitrary zero, but it is not possible to determine for them what may be called an absolute zero or the unique origin.

The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the

complete absence of a trait or characteristic. The Fahrenheit scale is an example of an interval scale and shows similarities in what one can and cannot do with it. One can say that an increase in temperature from 30° to 40° involves the same increase in temperature as an increase from 60° to 70°, but one cannot say that the temperature of 60° is twice as warm as the temperature of 30° because both numbers are dependent on the fact that the zero on the scale is set arbitrarily at the temperature of the freezing point of water.

The ratio of the two temperatures, 30° and 60°, means nothing because zero is an arbitrary point. Interval scales provide more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of interval. As such more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is the most widely used measure of dispersion. Product moment correlation techniques are appropriate and the generally used tests for statistical significance are the‘t’ test and ‘F’ test.

(d) Ratio scale: Ratio scales have an absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be. We can conceive of an absolute zero of length and similarly we can conceive of an absolute zero of time. For example, the zero point on a centimeter scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unobtainable and it remains a concept existing only in the scientist's mind.

The number of minor traffic-rule violations and the number of incorrect letters in a page of type script represent scores on ratio scales. Both these scales have absolute zeros and as such all minor traffic violations and all typing errors can be assumed to be equal in significance. With ratio scales involved one can make statements like "Jyoti's" typing performance was twice as good as that of "Reetu."

The ratio involved does have significance and facilitates a kind of comparison which is not possible in case of an interval scale. Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, etc. are examples. Generally, all statistical

techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above.

Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated. Thus, proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly. If the nature of the variables permits, the researcher should use the scale that provides the most precise description. Researchers in physical sciences have the advantage to describe variables in ratio scale form but the behavioral sciences are generally limited to describe variables in interval scale form, a less precise type of measurement.

2.4.2. Comparison of nominal scale, ordinal scale, interval scale and ratio scale.

Scale	Basic characteristics	Common examples	Marketing examples
Nominal	Numbers identity and classify objects	Social security nos., numbering of football players	Brand nos., Store types
Ordinal	Numbers indicate the relative positions of objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference Rankings, Market position, Social class
Interval	differences between objects can be compared, zero point is arbitrary	Temperature (Fahrenheit) (Celsius)	Attitudes, Opinions, Index numbers
Ratio	Zero point is fixed, ratios of scale values can be compared	Length, Weight.	Age, Sales, Income, costs.

2.4.3. Scale construction techniques: Following are the five main techniques by which scales can be developed.

(i) Arbitrary approach: It is an approach where scale is developed on ad hoc basis. This is the most widely used approach. It is presumed that such scales measure the concepts for which they have been designed, although there is little evidence to support such an assumption.

(ii) Consensus approach: Here a panel of judges evaluate the items chosen for inclusion in the instrument in terms of

whether they are relevant to the topic area and unambiguous in implication.

(iii) Item analysis approach: Under it a number of individual items are developed into a test which is given to a group of respondents. After administering the test, the total scores are calculated for everyone. Individual items are then analyzed to determine which items discriminate between persons or objects with high total scores and those with low scores.

(iv) Cumulative scales approach: are chosen on the basis of their conforming to some ranking of items with ascending and descending discriminating power. For instance, in such a scale the endorsement of an item representing an extreme position should also result in the endorsement of all items indicating a less extreme position.

(vi) **Factor scales approach:** may be constructed on the basis of inter correlations of items which indicate that a common factor accounts for the relationship between items. This relationship is typically measured through factor analysis method.

2.5. Administration of survey:

2.5.1. Differential Scales (or Thurstone-type Scales)

The name of L.L. Thurstone is associated with differential scales which have been developed using consensus scale approach. Under this approach the selection of items is made by a panel of judges who evaluate the items in terms of whether they are relevant to the topic area and unambiguous in implication.

The detailed procedure is as under:

(a) The researcher gathers a large number of statements, usually twenty or more, that express various points of view toward a group, institution, idea, or practice (i.e., statements belonging to the topic area).

(b) These statements are then submitted to a panel of judges, each of whom arranges them in eleven groups or piles ranging from one extreme to another in position. Each of the judges is requested to place generally in the first pile the statements which he thinks are most unfavourable to the issue, in the second pile to place those statements which he thinks are next most unfavourable and he goes on doing so in this manner till in the eleventh pile he puts the statements which he considers to be the most favourable.

(c) This sorting by each judge yields a composite position for each of the items. In case of marked disagreement between the judges in assigning a position to an item, that item is discarded.

(d) For items that are retained, each is given its median scale value between one and eleven as established by the panel. In other words, the scale value of any one statement is computed as the 'median' position to which it is assigned by the group of judges.

(e) A final selection of statements is then made. For this purpose a sample of statements, whose median scores are spread evenly from one extreme to the other is taken. The statements so selected, constitute the final scale to be administered to respondents. The position of each statement on the scale is the same as determined by the judges.

After developing the scale as stated above, the respondents are asked during the administration of the scale to check the statements with which they agree. The median value of the statements that they check is worked out and this establishes their score or quantifies their opinion. It may be noted that in the actual instrument the statements are arranged in random order of

scale value. If the values are valid and if the opinionnaire deals with only one attitude dimension, the typical respondent will choose one or several contiguous items (in terms of scale values) to reflect his views. However, at times divergence may occur when a statement appears to tap a different attitude dimension.

Merits of Thurstone Method:

- This method has been widely used for developing differential scales which are utilized to measure attitudes towards varied issues like war, religion, etc.
- Such scales are considered most appropriate and reliable when used for measuring a single attitude.

Demerits of Thurstone Method:

- To their use is the cost and effort required to develop them.
- Another weakness of such scales is that the values assigned to various statements by the judges may reflect their own attitudes.
- The method is not completely objective; it involves ultimately subjective decision process.

2.5.2. Summated Scales (or Likert-type Scales) or Likert- 5 point scale:

Summated scales (or Likert-type scales) are developed by utilizing the item analysis approach wherein a particular item is evaluated on the basis of how well it discriminates between those persons whose total score is high and those whose score is low. Those items or statements that best meet this sort of discrimination test are included in the final instrument.

Thus, summated scales consist of a number of statements which express either a favourable or unfavourable attitude towards the given object to which the respondent is asked to react. The respondent indicates his agreement or disagreement with each statement in the instrument. Each response is given a numerical score, indicating its favourableness or unfavourableness, and the scores are totaled to measure the respondent's attitude. In other words, the overall score represents the respondent's position on the continuum of favourable-unfavourableness towards an issue.

Most frequently used summated scales in the study of social attitudes follow the pattern devised by Likert. For this reason they are often referred to as Likert-type scales. In a Likert scale,

the respondent is asked to respond to each of the statements in terms of several degrees, usually five degrees (but at times 3 or 7 may also be used) of agreement or disagreement. For example, when asked to express opinion whether one considers his job quite pleasant, the respondent may respond in any one of the following ways:

(i) strongly agree - 5
(ii) agree - 4
(iii) Undecided - 3
(iv) Disagree - 2
(v) Strongly disagree - 1

We find that these five points constitute the scale. At one extreme of the scale there is strong agreement with the given statement and at the other, strong disagreement, and between them lie intermediate points. We may illustrate this as under: Each point on the scale carries a score. Response indicating the least favourable degree of job satisfaction is given the least score (say 1) and the most favourable is given the highest score (say 5).

These score—values are normally not printed on the instrument but are shown here just to indicate the scoring pattern. The Likert scaling technique, thus, assigns a scale value to each of the five responses. The same thing is done in respect of each and every statement in the instrument. This way the instrument yields a total score for each respondent, which would then measure the respondent's favourableness toward the given point of view. If the instrument consists of, say 30 statements, the following score values would be revealing.

30 × 5 = 150 Most favourable response possible,

30 × 4 = 120 favorable response

30 × 3 = 90 A neutral attitude

30 × 2 = 60 unfavourableattitude

30 × 1 = 30 Most unfavourable attitude.

The scores for any individual would fall between 30 and 150. If the score happens to be above 90, it shows favourable opinion to the given point of view, a score of below 90 would mean unfavourable opinion and a score of exactly 90 would be suggestive of a neutral attitude.

Procedure: The procedure for developing a Likert-type scale is as follows:

(i) As a first step, the researcher collects a large number of statements which are relevant to the attitude being studied and each of the statements expresses definite favourableness or unfavourableness to a particular point of view or the attitude and that the number of favourable and unfavourable statements is approximately equal.

(ii) After the statements have been gathered, a trial test should be administered to a number of subjects. In other words, a small group of people, from those who are going to be studied finally, are asked to indicate their response to each statement by checking one of the categories of agreement or disagreement using a five point scale as stated above.

(iii) The response to various statements are scored in such a way that a response indicative of the most favourable attitude is given the highest score of 5 and that with the most unfavourable attitude is given the lowest score, say, of 1.

(iv) Then the total score of each respondent is obtained by adding his scores that he received for separate statements.

(v) The next step is to array these total scores and find out those statements which have a high discriminatory power. For this purpose, the researcher may select some part of the highest and the lowest total scores, say the top 25 per cent and the bottom 25 per cent. These two extreme groups are interpreted to represent the most favourable and the least favourable attitudes and are used as criterion groups by which to evaluate individual statements. This Stronglyagree (1)Agree(2)Undecided(3)Disagree(4)Stronglydisagree (5) way we determine which statements consistently correlate with low favourability and which with high favourability.

(vi) Only those statements that correlate with the total test should be retained in the final instrument and all others must be discarded from it.

Advantages: The Likert-type scale has several advantages. Mention may be made of the important ones.

(a) It is relatively easy to construct the Likert-type scale in comparison to Thurstone-type scale because Likert-type scale can be performed without a panel of judges.

(b) Likert-type scale is considered more reliable because under it respondents answer each statement included in the

instrument. As such it also provides more information and data than does the Thurstone-type scale.

(c) Each statement, included in the Likert-type scale, is given an empirical test for discriminating ability and as such, unlike Thurstone-type scale, the Likert-type scale permits the use of statements that are not manifestly related (to have a direct relationship) to the attitude being studied.

(d) Likert-type scale can easily be used in respondent-centred and stimulus-centred studies

i.e., through it we can study how responses differ between people and how responses differ between stimuli.

(e) Likert-type scale takes much less time to construct, it is frequently used by the students of opinion research. Moreover, it has been reported in various research studies* that there is high degree of correlation between Likert-type scale and Thurstone-type scale.

Limitations: There are several limitations of the Likert-type scale as well.

- Likert- 5 point scale, we can simply examine whether respondents are more or less favourable to a topic, but we cannot tell how much more or less they are.

- There is no basis for belief that the five positions indicated on the scale are equally spaced. The interval between 'strongly agree' and 'agree', may not be equal to the interval between "agree" and "undecided". This means that Likert scale does not rise to a stature more than that of an ordinal scale, whereas the designers of Thurstone scale claim the Thurstone scale to be an interval scale.
- One further disadvantage is that often the total score of an individual respondent has little clear meaning since a given total score can be secured by a variety of answer patterns.
- It is unlikely that the respondent can validly react to a short statement on a printed form in the absence of real-life qualifying situations.
- Moreover, there "remains a possibility that people may answer according to what they think they should feel rather than how they do feel."

2.6. Sample Design:

A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the

procedure the researcher would adopt in selecting items for the sample. Sample design may as well lay down the number of items to be included in the sample i.e., the size of the sample. Sample design is determined before data are collected. Researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

Steps in Sample Design:

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

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

(ii) **Sampling unit:** A decision has to be taken concerning a sampling unit before selecting sample. Sampling unit may be a

geographical one such as state, district, village, etc., or a construction unit such as house, flat, etc., or it may be a social unit such as family, club, school, etc., or it may be an individual. The researcher will have to decide one or more ofsuch units that he has to select for his study.

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

(iv) **Size of sample:** This refers to the number of items to be selected from the universe to constitute a sample. This is a major problem before a researcher. The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility. While deciding the size of sample, researcher must determine the desired precision as also an acceptable confidence level for the estimate. The size of population variance needs tobe considered

as in case of larger variance usually a bigger sample is needed. The size of population must be kept in view for this also limits the sample size. The parameters of interest in a research study must be kept in view, while deciding the size of the sample. Costs to dictate the size of sample that we can draw. As such, budgetary constraint must invariably be taken into consideration when we decide the sample size.

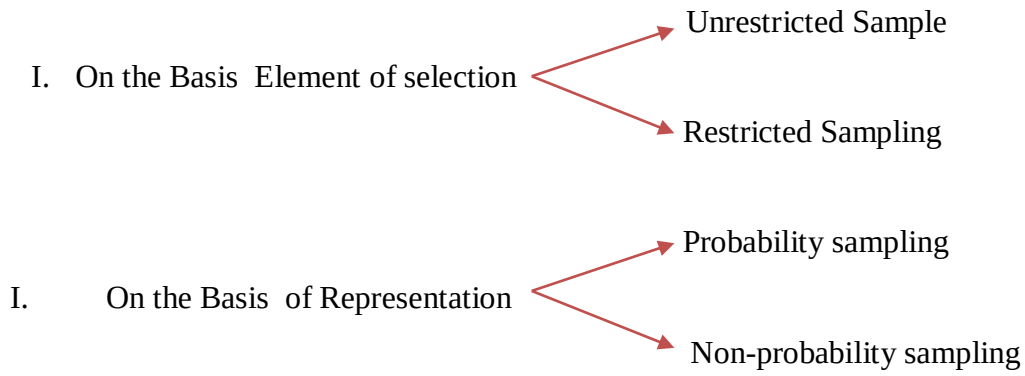
(v) **Parameters of interest:** In determining the sample design, one must consider the question of the specific population parameters which are of interest. For instance, we may be interested in estimating the proportion of persons with some characteristic in the population, or we may be interested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

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

(vii) **Sampling procedure:** Finally, the researcher must decide the type of sample he will use

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

Different Types of Sample Designs:



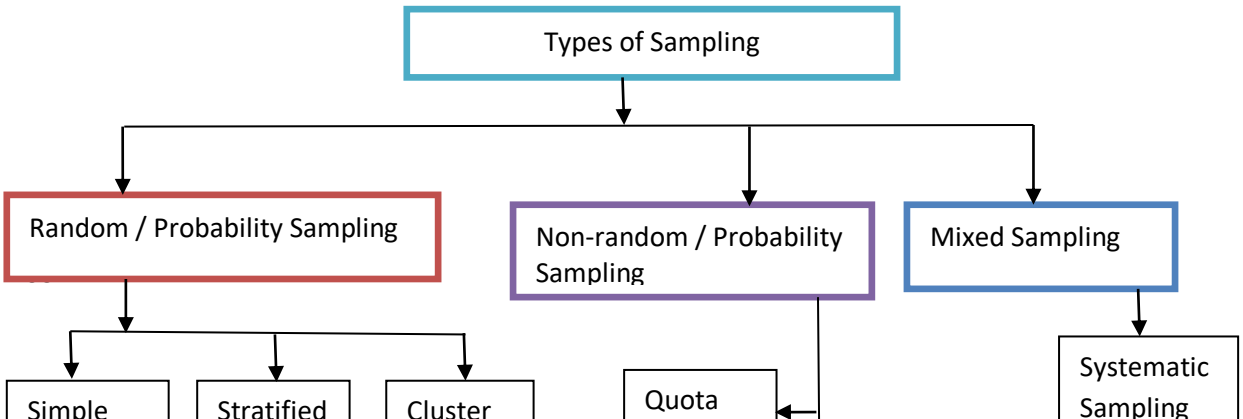
The following chart exhibits the sample designs as explained above. Thus, sample designs are basically of three types viz.

1) Probability sampling,
2) Non-Probability sampling and
3) Mixed Sampling

Probability Sampling:

Probability sampling is also known as 'choice sampling' or 'random sampling'. Under this sampling design, every item of the universe has an equal chance of being included in the sample. In a way, it is a lottery method under which individual units are selected from the whole group, not deliberately, but by using some mechanical process. Therefore, only chance would determine whether an item or the other would be included in the sample or not. The results obtained from probability or random sampling would be assured in terms of probability. That is, the researcher can measure the errors of estimation or the significance of results obtained from the random sample. This is the superiority of random sampling design over the deliberate sampling design.

Random sampling satisfies the law of statistical regularity, according to which if on an average the sample chosen is random, then it would have the same composition and characteristics of the universe.



This is the reason why the random sampling method is considered the best technique of choosing a representative sample.The following are the implications of the random sampling:

- it provides each element in the population an equal probable chance of being chosen in the sample, with all choices being independent of one another and
- It offers each possible sample combination an equal probable opportunity of being selected.

Simple Random Sampling (SRS) :– The most commonly used method of selecting a probability sample. In line with the definition of randomization, whereby each element in the

population isgiven an equal and independent chance of selection, let us take example of the class. There are 80 students in the class, and so the first step is to identify each student by a number from 1 to 80. Suppose you decide to select a sample of 20 using the simple random sampling technique. Use the fishbowl draw, the table for random numbers or a computer program to select the 20 students. These 20 students become the basis of your enquiry.

Method of Selecting a Random Sample:

The process of selecting a random sample involves writing the name of each element of a finite population on a slip of paper and putting them into a box or a bag. Then they have to be thoroughly mixed and then the required number of slips for the sample can be picked one after the other without replacement. While doing this, it has to be ensured that in successive drawings each of the remaining elements of the population has an equal chance of being chosen. This method results in the same probability for each possible sample

Stratified random sampling: – As discussed, the accuracy of your estimate largely depends on the extent of variability or heterogeneity of the study population with respect to the

characteristics that have a strong correlation with what you are trying to It follows; therefore, that if the heterogeneity in the population can be reduced by some means for a given sample size you can achieve greater accuracy in your estimate. Stratified random sampling is based upon this logic.

In stratified random sampling the researcher attempts to stratify the population in such a way that the population within a stratum is homogeneous with respect to the characteristic on the basis of which it is being stratified. It is important that the characteristics chosen as the basis of stratification are clearly identifiable in the study population. For example, it is much easier to stratify a population on the basis of gender than on the basis of age, income or attitude. It is also important for the characteristic that becomes the basis of stratification to be related to the main variable that you are exploring.

Once the sampling population has been separated into no overlapping groups, you select the required number of elements from each stratum, using the simple random sampling technique. There are two types of stratified sampling:

Proportionate Stratified Sampling:-With proportionate stratified sampling, the number of elements from each stratum in

relation to its proportion in the total population is selected, For example, if the final year MBA students of the management faculty of a university consist of the following specialization groups:

Specialization Stream	No. of Students	Proportion of each stream.
Production	40	0.4
Finance	20	0.2
Marketing	30	0.3
Rural Development	10	0.1
Total	100	1.0

The researcher wants to draw an overall sample of 30, then the strata sample sizes would be:

Strata	Sample Size
Production	30 x 0.4 = 12
Finance	30 x 0.2 = 6
Marketing	30 x 0.3 = 9
Rural Development	30 x 0.1 = 3
Total	30

Thus, proportionate sampling gives proper representation to each stratum and its statistical efficiency is generally higher. This method is, therefore, very popular.

Disproportionate Stratified Sampling:

In disproportionate stratified sampling, consideration is not given to the size of the stratum.

Complex Random Sampling Designs:

Under restricted sampling technique, the probability sampling may result in complex random sampling designs. Such designs are known as mixed sampling designs. Many of such designs may represent a combination of non-probability and probability sampling procedures in choosing a sample.

Some of the prominent complex random sampling designs are as follows:

Stratified Sampling:

When a population from which a sample is to be selected does not comprise a homogeneous group, stratified sampling technique is generally employed for obtaining a representative sample. Under stratified sampling, the population is divided into many sub-populations in such a manner that they are individually more homogeneous than the rest of the total population. Then, items are selected from each stratum to form a sample. As each stratum is more homogeneous than the remaining total population, the researcher is able to obtain a more precise estimate for each stratum and by estimating each of the component parts more accurately; he/she is able to obtain a better estimate of the whole. In sum, stratified sampling method yields more reliable and detailed information.

Cluster Sampling:

When the total area of research interest is large, a convenient way in which a sample can be selected is to divide the area into a number of smaller non-overlapping areas and then randomly selecting a number of such smaller areas. In the process, the ultimate sample would consist of all the units in these small areas or clusters. Thus in cluster sampling, the total population is sub-divided into numerous relatively smaller subdivisions, which in themselves constitute clusters of still smaller units. And then, some of such clusters are randomly chosen for inclusion in the overall sample. Some illustrations of clusters are:

Population	**Elements**	**Cluster or sampling units**
City	Households	Blocks
City	Individuals	Households
Affiliating University	Students	Affiliated colleges
Rural areas	Households	Villages
Industrial areas	Industrial unit	Industrial estates

Area Sampling:

When clusters are in the form of some geographic subdivisions, then cluster sampling is termed as area sampling. That is, when the primary sampling unit represents a cluster of units based on geographic area, the cluster designs are distinguished as area sampling. The merits and demerits of cluster sampling are equally applicable to area sampling.

Multi-Stage Sampling:

A further development of the principle of cluster sampling is multi-stage sampling. When the researcher desires to investigate the working efficiency of nationalized banks in India and a sample of few banks is required for this purpose, the first stage would be to select large primary sampling unit like the states in the country. Next, certain districts may be selected and all banks interviewed in the chosen districts. This represents a two-stage sampling design, with the ultimate sampling units being clusters of districts.

On the other hand, if instead of taking census of all banks within the selected districts, the researcher chooses certain towns and interviews all banks in it, this would represent three-stage sampling design. Again, if instead of taking a census of all banks within the selected towns, the researcher randomly selects

sample banks from each selected town, then it represents a case of using a four-stage sampling plan. Thus, if the researcher selects randomly at all stages, then it is called as multi-stage random sampling design.

Sampling with Probability Proportional to Size:

When the case of cluster sampling units does not have exactly or approximately the same number of elements, it is better for the researcher to adopt a random selection process, where the probability of inclusion of each cluster in the sample tends to be proportional to the size of the cluster. For this, the number of elements in each cluster has to be listed, irrespective of the method used for ordering it. Then the researcher should systematically pick the required number of elements from the cumulative totals.

The actual numbers thus chosen would not however reflect the individual elements, but would indicate as to which cluster and how many from them are to be chosen by using simple random sampling or systematic sampling. The outcome of such sampling is equivalent to that of simple random sample. The method is also less cumbersome and is also relatively less expensive.

Thus, a researcher has to pass through various stages of conducting research once the problem of interest has been selected. Research methodology familiarizes a researcher with the complex scientific methods of conducting research, which yield reliable results that are useful to policy-makers, government, industries etc. in decision-making.

Non-Probability Sampling:

Non-probability sampling is the sampling procedure that does not afford any basis for estimating the probability that each item in the population would have an equal chance of being included in the sample. Non-probability sampling is also known as deliberate sampling, judgment sampling and purposive sampling. Under this type of sampling, the items for the sample are deliberately chosen by the researcher; and his/her choice concerning the choice of items remains supreme. In other words, under non-probability sampling the researchers select a particular unit of the universe for forming a sample on the basis that the small number that is thus selected out of a huge one would be typical or representative of the whole population.

For example, to study the economic conditions of people living in a state, a few towns or village may be purposively selected for an intensive study based on the principle that they are representative of the entire state. In such a case, the judgment of the researcher of the study assumes prime importance in this sampling design.

Quota Sampling:

Quota sampling is also an example of non-probability sampling. Under this sampling, the researchers simply assume quotas to be filled from different strata, with certain restrictions imposed on how they should be selected. This type of sampling is very convenient and is relatively less expensive. However, the samples selected using this method certainly do not satisfy the characteristics of random samples. They are essentially judgment samples and inferences drawn based on that, would not be amenable to statistical treatment in a formal way.

Accidental sampling:

Accidental sampling is also based upon convenience in accessing the sampling population. Whereas quota sampling attempts to include people possessing an obvious/visible characteristic, accidental sampling makes no such attempt. You

stop collecting data when you reach the required number of respondents you decided to have in your sample. This method of sampling is common among market research and newspaper reporters. It has more or less the same advantages and disadvantages as quota sampling but, in addition, as you are not guided by any obvious characteristics, some people contacted may not have the required information.

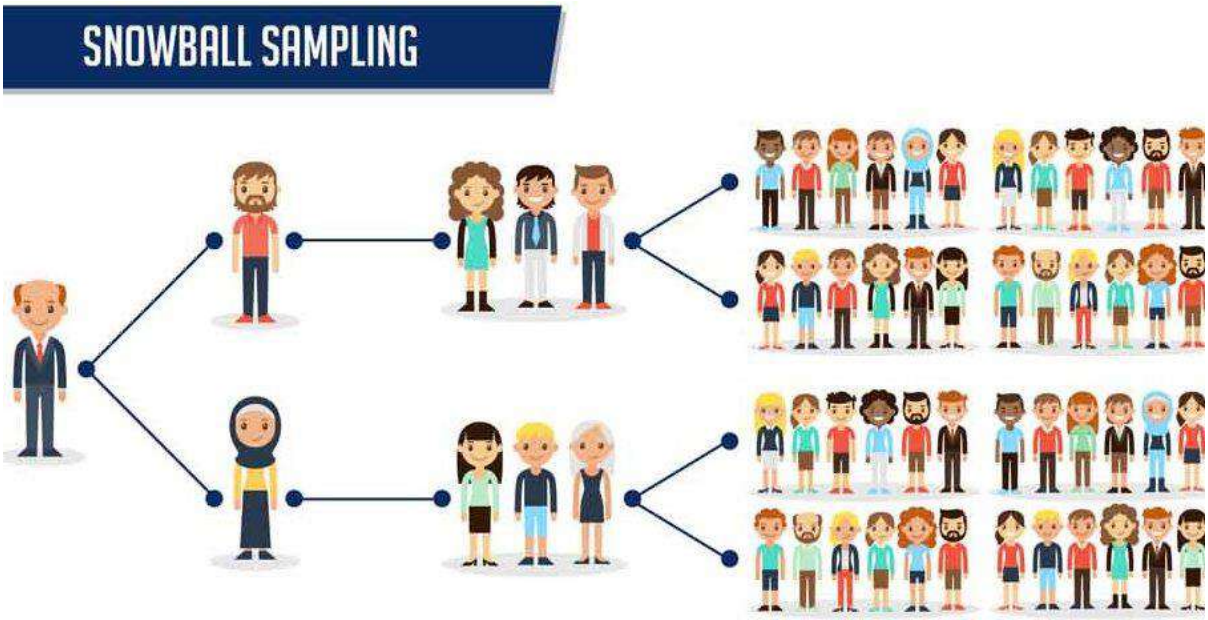
Judge mental or purposive sampling:

The primary consideration in purposive sampling is your judgment as to who can provide the best information to achieve the objectives of your study. You as a researcher only go to those people who in your opinion are likely to have the required information and be willing to share it with you. This type of sampling is extremely useful when you want to construct a historical reality, describe a phenomenon or develop something about which only a little is known. This sampling strategy is more common in qualitative research, but when you use it in quantitative research you select a predetermined number of people who, in your judgement, are best positioned to provide you the needed information for your study.

Snowball sampling:

Snowball sampling is the process of selecting a sample using networks. To start with, a few individuals in a group or organization are selected and the required information is collected from them. They are then asked to identify other people in the group or organization, and the people selected by them become a part of the sample. Information is collected from them, and then these people are asked to identify other members of the group and, in turn, those identified become the basis of further data collection. This process is continued until the required number or a **saturation point** has been reached, in terms of the information being sought. This sampling technique is useful if you know little about the group or organisation you wish to study, as you need only to make contact with a few individuals, who can then direct you to the other members of the group. This method of selecting a sample is useful for studying communication patterns, decision making or diffusion of knowledge within a group. There are disadvantages to this technique, however. The choice of the entire sample rests upon the choice of individuals at the first stage. If they belong to a particular faction or have strong biases, the study may be biased.

Also, it is difficult to use this technique when the sample becomes fairly large.



Expert sampling:

The only difference between judgmental sampling and expert sampling is that in the case of the former it is entirely your judgement as to the ability of the respondents to contribute to the study. But in the case of expert sampling, your respondents must be known experts in the field of interest to you. This is again used in both types of research but more so in qualitative research studies. When you use it in qualitative research, the number of people you talk to is dependent upon the data saturation point whereas in quantitative research you decide on the number of experts to be contacted without considering the

saturation point. You first identify persons with demonstrated or known expertise in an area of interest to you, seek their consent for participation, and then collect the information either individually or collectively in the form of a group.

Systematic Sampling:

In some cases, the best way of sampling is to select every first item on a list. Sampling of this kind is called as systematic sampling. An element of randomness is introduced in this type of sampling by using random numbers to select the unit with which to start. For example, if a 10 per cent sample is required out of 100 items, the first item would be selected randomly from the first low of item and thereafter every 10th item. In this kind of sampling, only the first unit is selected randomly, while rest of the units of the sample is chosen at fixed intervals.

Short Questions

1. Secondary data
2. What is a schedule? Explain
3. What is data classification?

4. Types of errors
5. Kinds of data
6. Sample design
7. Random sample
8. Need for survey
9. Exploratory research
10. Quota sampling
11. Qualities of a good questionnaire
12. Experimental research
13. Systematic sampling
14. Explain the sources of information
15. How can you determine sample size?
16. Sources of information

Essay Questions

1. What is a questionnaire? Explain the qualities of a good questionnaire.
2. What is sampling? Explain the probability sampling techniques in detail
3. Elucidate the ex- post facto research
4. Distinguish between measurement and evaluation.
5. Difference between questionnaire and schedule
6. What is data? Explain primary data collection methods in briefly
7. What is a questionnaire? Explain the general structure of a questionnaire.
8. What is the need for data collection? Kinds of data and point out where is a source of information.
9. What is the schedule? Discuss its validity of schedule as a data.
10. Explain various sampling methods.
11. Explain primary data collection methods in detail

Unit – III: Tabulation and Data Analysis

3.1. Tabulation and Cross Tabulation of Data

3.2. Univariate,

3.3. Bivariate

3.4 Data - Analysis and Interpretation

3.5. Testing of Hypothesis-

3.6. SPSS Packages and Applications

3.1 Tabulation of Data: Tabulation is a systematic & logical presentation of numeric data in rows and columns, to facilitate comparison and statistical analysis. To put it in other words, the method of placing organised data into a tabular form is called as tabulation. Tables can be classified according to their purpose and characteristics used.

Tabulation comprises sorting of the data into different categories and counting the number of cases that belong to each category. The simplest way to tabulate is to count the number of responses to one question. This is also called univariate tabulation. The analysis based on just one variable is obviously meager. Where two or more variables are involved in tabulation,

it is called bivariate or multivariate tabulation. In marketing research, projects generally both types of tabulation are used.

3.1.1 Definition:

- **Prof. Neiswanger** has defined a statistical table as "In a systemic organisation of data in columns and rows."
- **L. K Connor** has defined tabulation as the orderly and systematic presentation of numerical data in a form designed to elucidate the problem under consideration

3.1.2 Objectives of Tabulation:

The following are the main objects of tabulation.

- To make the purpose of enquiry clear tabulation in the general scheme of statistics investigation is to arrange in easily accessible form.
- To make significance clear by arranging in form of table the significance of data, is made very clear. This is because table permits the observation of the whole data in one glance. The total information is clear to the view and the significance of different parts can easily be followed.
- To express the data in least space table also permits the data to be represented in least possible space, making the whole information clear to the view. If it is expressed in

form of a passage it would not only be difficult to follow, but would require more space too.

- To make comparison easy mainly because of the arrangement of figures in it. When two sets of figures are given side by side, it is much easier to form a comparative idea of their significance

3.1.3 Classification of tabulation:

1. Simple or one-way table
2. Two way table
3. Manifold table or Cross Tabulation:

1. Simple Tabulation or one-way table:

- A simple or one-way table is the simplest table which contains data of one characteristic only.
- It gives information about one or more groups of independent questions. This results, in one way table, provides information of one characteristics of data.
- A simple table is easy to construct and simple to follow.

The Number of adults in different occupations in a locality

Occupations	No. of Adults

Farmers	100
Other Farmers	100
Total	200

2. Two way table

- A table, which contains data on two characteristics, is called a two way table.
- Either stub or caption is divided into two co-ordinate parts
- For example the caption may be further divided in respect of 'sex'. This subdivision is shown in two-way table, which now contains two characteristics namely, occupation and sex.

The Number of adults in a locality in respect of Occupation and Sex

Occupation	**No. of Adults**		Total
	Male	Female	
Labors	50	50	100
Total	50	50	100

3. Manifold table or Complex Tabulation or Cross Tabulation:

More and more complex tables can be formed by including other characteristics.

- For example, we may further classify the caption sub-headings in the above table in respect of “marital status”, “religion” and “socio-economic status” etc. A table, which has more than two characteristics of data, is considered as a manifold table. For instance, table shown below shows three characteristics namely, occupation, sex and marital status.
- In this type of tabulation, the data is divided in two or more categories which gives information regarding more sets of interrelated question.

The Number of adults in a locality in respect of Occupation, Sex and Marital status

Occupation	**No. of Adults**						Total
	Male			Female			
Marital status	M	U	Total	M	U	Total	

Labors	30	20	50	40	10	50	100
Total	30	20	50	40	10	50	100

Foot note: M Stands for Married and U Stands for Unmarried

3.2. Univariate data analysis

Univariate data: This type of data consists of **only one variable**. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it. The example of a univariate data can be height.

Height (In Cm.)	160	163	169.3	171	175	180	185

suppose that the heights of seven students of a class is recorded (figure 1),there is only one variable that is height and it is not dealing with any cause or relationship. The description of patterns found in this type of data can be made by drawing

conclusions using central tendency measures (mean, median and mode), dispersion or spread of data (range, minimum, maximum, quartiles, variance and standard deviation) and by using frequency distribution tables, histograms, pie charts, frequency polygon and bar charts.

3.3. Bivariate data: This type of data involves **two different variables**. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables. Example of bivariate data can be temperature and ice cream sales in summer season.

Temperature (In Celsius)	Ice Cream Sales (No.)
23	1700
26	2400
36	4800
47	7100

Suppose the temperature and ice cream sales are the two variables of a bivariate data(the above Table). Here, the relationship is visible from the table that temperature and sales

are directly proportional to each other and thus related because as the temperature increases, the sales also increase. Thus bivariate data analysis involves comparisons, relationships, causes and explanations. These variables are often plotted on X and Y axis on the graph for better understanding of data and one of these variables is independent while the other is dependent.

Differences between univariate and bivariate data.

Univariate Data	Bivariate Data
➢ involving a single variable ➢ does not deal with causes or relationships ➢ the major purpose of univariate analysis is to describe ➢ central tendency - mean, mode, median dispersion - range, variance, max, min, ➢ quartiles, standard deviation. frequency distributions ➢ bar graph, histogram, pie chart, line ➢ graph, box-and-whisker plot	➢ involving two variables ➢ deals with causes or relationships ➢ the major purpose of bivariate analysis is to explain ➢ analysis of two variables simultaneously correlations ➢ comparisons, relationships, causes, ➢ explanations tables where one variable is contingent ➢ On the values of the other variable. independent and dependent variables

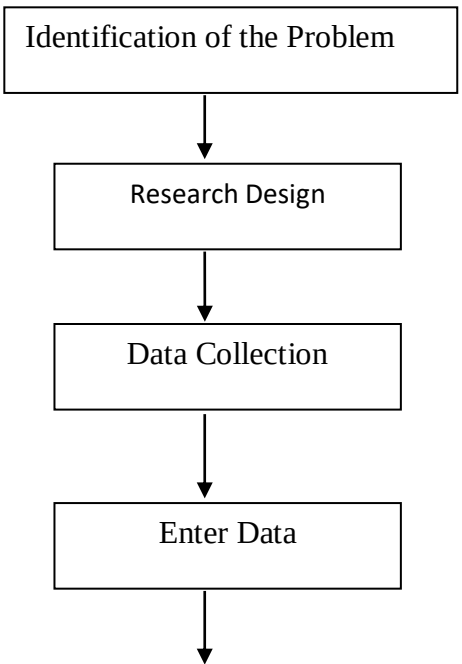
3.4. Analysis of Data:

The data collected may or may not in numerical form. Even if data is not in numerical form still we can carry out qualitative analysis based on the experiences of individual participants.

When data is collected in numerical form than through descriptive statistics findings can be summarized. This includes measure of central tendency like mean range etc. Another way to summarized finding is by means of graphs and charts. In any of the research study there is experimental hypothesis or null hypothesis one the basis of data of both hypothesis, various test have been devised to take decision. Where decision is taken on the basis statistical test, it is subject to error, and such correct decision is difficult. But some standard procedures followed to arrived at proper decision.

Analysis involves estimating the values of unknown parameters of the population and testing hypothesis for drawing inferences. The processing of data includes various operations which will be explained here.

Data Analysis and Interpretation



3.4.1. Data interpretation is a crucial step in research. Data should not be studies in isolation of all factors which might influence the origin of the data. Any decision taken needs a holistic overview of paraphernalia factors associated with the subject in concern.

The data which is collected for the purpose of the study itself cannot several any things. This being a raw data it is required to process and analyzed in order to have desired result. The data which is collected cannot be directly use for making analysis. Before analysis data is requires to be processed.

3.4.2. Data processing is a crucial stage in research. After collecting the data from the field, the researcher has to process and analyze them in order to arrive at certain conclusion which may confirm or invalidate the hypothesis which he had formulated towards the beginning of his research worth. The

mass of data collected during the field work is to be processed with a view to reducing them to manageable proportions. The processing of data includes editing, coding, classification and tabulation. The data which is collected is analyze with an object checking of data, and reducing this data into manageable proportions. Data collected should be organized in such way so that table charts can be prepared for presentation. The processing of data includes various operations which will be explained here.

3.4.3. Meaning of Interpretation:

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz.,

(i) the effort to establish continuity in research through linking the results of a given study with those of another, and

(ii) The establishment of some explanatory concepts. “In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to

include the results of other research, theory and hypotheses.

Thus, interpretation is the device through which the factors that seem to explain what has been observed by researcher in the course of the study can be better understood and it also provides a theoretical conception which can serve as a guide for further researches.

3.4.4. Technique of Interpretation:

The task of interpretation is not an easy job, rather it requires a great skill and dexterity on the part of researcher. Interpretation is an art that one learns through practice and experience. The researcher may, at times, seek the guidance from experts for accomplishing the task of interpretation.

The technique of interpretation often involves the following steps:

1) Researcher must give reasonable explanations of the relations which he has found and he must interpret the lines of relationship in terms of the underlying processes and must try to find out the thread of uniformity that lies under the surface layer of his diversified research findings. In

fact, this is the technique of how generalization should be done and concepts be formulated.

2) Extraneous information, if collected during the study, must be considered while interpreting the final results of research study, for it may prove to be a key factor in understanding the problem under consideration.
3) It is advisable, before embarking upon final interpretation, to consult someone having insight into the study and who is frank and honest and will not hesitate to point out omissions and errors in logical argumentation. Such a consultation will result in correct interpretation and, thus, will enhance the utility of research results.
4) Researcher must accomplish the task of interpretation only after considering all relevant factors affecting the problem to avoid false generalization. He must be in no hurry while interpreting results, for quite often the conclusions, which appear to be all right at the beginning, may not at all be accurate.

3.5. Hypothesis:

A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.One

simply means a mere assumption or some supposition to be proved or disproved. But for a researcher hypothesis is a formal question that he intends to resolve. Thus a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional conjuncture to guide some investigation or accepted as highly probable in the light of established facts. Quite often a research hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variable to some dependent variable. For example, consider statements like the following ones:

"Students who receive counselling will show a greater increase in creativity than students not receiving counselling" Or "The automobile *A* is performing as well as automobile *B*."

These are hypotheses capable of being objectively verified and tested. Thus, we may conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test to determine its validity.

3.5.1 Characteristics of hypothesis: Hypothesis must possess the following characteristics:

- Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.
- Hypothesis should be capable of being tested.
- Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.
- Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.
- Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.
- Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.
- Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent

hypothesis, if the same cannot be tested in reasonable time for one cannot spend a life-time collecting data to test it.

- Hypothesis must explain the facts that gave rise to the need for explanation.

3.5.2 Basic Concepts Concerning Testing Of Hypotheses

Basic concepts in the context of testing of hypotheses need to be explained.

3.5.3 Null hypothesis and alternative hypothesis:

In the context of statistical analysis, we often talk about null hypothesis and alternative hypothesis. If we are to compare method *A* with method *B* about its superiority and if we proceed on the assumption that both methods are equally good, then this assumption is termed as the null hypothesis. As against this, we may think that the method *A* is superior or the method *B* is inferior, we are then stating what is termed as alternative hypothesis. The null hypothesis is generally symbolized as $H0$ and the alternative hypothesis as *Ha*.

Suppose we want to test the hypothesis that the population mean (μ) is equal to the hypothesized mean (μ $H0$) = 100.

Then we would say that the null hypothesis is that the population mean is equal to the hypothesized mean 100 and symbolically we can express as:

$$H0: \mu = \mu\ H0 = 100$$

If our sample results do not support this null hypothesis, we should conclude that something else is true. What we conclude rejecting the null hypothesis is known as alternative hypothesis. In other words, the set of alternatives to the null hypothesis is referred to as the alternative hypothesis. If we accept $H0$, then we are rejecting *Ha* and if we reject $H0$, then we are accepting *Ha*. For $H0$: μ =μ $H0$ = 100 we may consider three possible alternative hypotheses as follows.

Alternative hypothesis	To be read as follows
Ha : μ ≠μ $H0$	(The alternative hypothesis is that the population mean is not equal to 100 i.e., it may be more or less than 100)

Ha : μ >μ *H*0	(The alternative hypothesis is that the population mean is greater than 100)
Ha : μ <μ *H*0	(The alternative hypothesis is that the population mean is less than 100)

The null hypothesis and the alternative hypothesis are chosen before the sample is drawn (the researcher must avoid the error of deriving hypotheses from the data that he collects and then testing the hypotheses from the same data). In the choice of null hypothesis, the following considerations are usually kept in view:

i. Alternative hypothesis is usually the one which one wishes to prove and the null hypothesis is the one which one wishes to disprove. Thus, a null hypothesis represents the hypothesis we are trying to reject, and alternative hypothesis represents all other possibilities.

ii. 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 is (the level of significance) which is chosen very small.

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

Generally, in hypothesis testing we proceed on the basis of null hypothesis, keeping the alternative hypothesis in view. Why so? The answer is that on the assumption that null hypothesis is true, one can assign the probabilities to different possible sample results, but this cannot be done if we proceed with the alternative hypothesis. Hence the use of null hypothesis (at times also known as statistical hypothesis) is quite frequent.

3.5.4 The level of significance*:* This is a very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should be chosen with great care, thought and reason. In case we take the significance level at 5 per cent, then this implies that *H*0 will be rejected when the sampling result (i.e., observed evidence) has a less than 0.05 probability of occurring if *H*0 is true. In other words, the 5 per cent level of significance means that researcher is willing to take as much as a 5 per cent risk of rejecting the null hypothesis when it (*H*0) happens to be true. Thus the significance level is the maximum

value of the probability of rejecting *H*0 when it is true and is usually determined in advance before testing the hypothesis.

3.5.5 Decision rule or test of hypothesis*:* Given a hypothesis *H*0 and an alternative hypothesis *Ha*, we make a rule which is known as decision rule according to which we accept *H*0 (i.e., reject *Ha*) or reject *H*0 (i.e., accept *Ha*). For instance, if (*H*0 is that a certain lot is good (there are very few defective items in it) against *Ha*) 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 for accepting or rejecting the hypothesis. We might test 10 items in the lot and plan our decision saying that if there are none or onlyb1 defective item among the 10, we will accept *H*0 otherwise we will reject *H*0 (or accept *Ha*). This sort of basis is known as decision rule.

3.5.6 **Type I and Type II errors***:* In the context of testing of hypotheses, there are basically two types of errors we can make. We may reject *H*0 when *H*0 is true and we may accept *H*0 when in fact *H*0 is not true. The former is known as Type I error and the latter as Type II error. In other words, Type I error means rejection of 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 the level of significance of test; and Type II error is denoted by β (beta) known as β error. In a tabular form the said two errors can be presented as follows:

	Decision	
	Accept *H*0	Reject *H*0
*H*0 (true)	Correct Decision	Type I error (α error)
*H*0 (false)	Type II error (β error)	Correct Decision

The probability of Type I error is usually determined in advance and is understood as the level of significance of testing the hypothesis. If type I error is fixed at 5 per cent, it means that there are about 5 chances in 100 that we will reject *H*0 when *H*0 is true. We can control Type I error just by fixing it at a lower level. For instance, if we fix it at 1 per cent, we will say that the maximum probability of committing Type I error would only be 0.01.

But with a fixed sample size, *n*, when we try to reduce Type I error, the probability of committing Type II error

increases. Both types of errors cannot be reduced simultaneously. There is a trade-off between two types of errors which means that the probability of making one type of error can only be reduced if we are willing to increase the probability of making the other type of error. To deal with this trade-off in business situations, decision-makers decide the appropriate level of Type I error by examining the costs or penalties attached to both types of errors. If Type I error involves the time and trouble of reworking a batch of chemicals that should have been accepted, whereas Type II error means taking a chance that an entire group of users of this chemical compound will be poisoned, then in such a situation one should prefer a Type I error to a Type II error. As a result one must set very high level for Type I error in one's testing technique of a given hypothesis.2 Hence, in the testing of hypothesis, one must make all possible effort to strike an adequate balance between Type I and Type II errors.

3.6. Statistical Package for Social Science (SPSS)

Introduction: SPSS is a software package used for conducting statistical analysis, manipulating data, and generating table and

graphs that summarize data. SPSS performs statistical analysis range from basic descriptive statistics, such as average and prevalence, to advanced inferential statistical, such as regression model, analysis of variance (ANOVA), factor analysis etc. SPSS also contains several tools for manipulating data, including functions for recording data, macros programming on visual basic editor, merging data, and aggregating complex data sets.

SPSS stands for Statistical Package for the Social Sciences}

- SPSS Incorporated is leading worldwide provider of predictive analytics software and solutions.
- First version of SPSS was released in 1968, after being developed by Norman H. Nie, Dale H. Bentand C. Hadlai Hull.
- The company announced on July 28, 2009 that it was being acquired by IBM for US$1.2 billion.
- Company Logo SPSS is now owned by IBM.
- Between 2009-10 the prime Vender of SPSS was called PASW (Predictive Analytics Software)

- IBM SPSS Statistics 21.0 – Released on August 2012 Latest version
- Compare groups to determine if there are significant differences between these groups example t-test, ANOVA etc.
- Examine Relationships between variables example correlation, regression, factor analysis etc.
- Describe data using descriptive statistics example frequency, mean, minimum and maximum.

3.6.1 Features and Benefits

- For Small And Medium Enterprises (SME)
- Resources & best practices.
- Techniques for cleaning data.
- Access Data in Relational Databases.
- For Large Scale Enterprises (LSE)
- Real time processing and online mapping.
- Online Database connectivity different networks.

- Data import & export data from different medium like SAS, Statistics.
- Compliances and Validations.

3.6.2Strengths

- Very robust statistical software
- Many complex statistical tests available
- Good “stats coach” help with interpreting results
- Easily and quickly displays data tables
- Can be expanded Using the syntax feature , Purchasing add-ins

3.6.3 Limitations

- Can be expensive
- Not intuitive to use
- Typically requires additional training to maximize features (at a cost)
- Graphing feature not as simple as Excel

Short Questions

1. Define nominal and ordinal scale
2. Explain how SPSS is useful in Analysis of data.
3. Likert scale
4. Univariate
5. Level of significance
6. Tabulation
7. Null hypothesis

Essay Questions

1. Explain the concept of editing, coding and tabulation in detail
2. Discuss the role of SPSS in data analysis
3. Discuss various techniques used in data analysis
4. Explain briefly the testing of hypothesis
5. What is tabulation? Explain editing and coding
6. The procedure of testing hypotheses.
7. Explain the advantages of using SPSS package
8. Define scaling? Explain different types of scales

9. What is hypothesis? Explain the process of testing hypotheses.
10. How we use SPSS in analysis and interpretation of a study

Unit– IV: Multivariate Analysis

4.1. Advanced Techniques for Data Analysis

4.2. ANOVA,

4.3. Discriminate Analysis,

4.4. Factor Analysis,

4.5. Clustering Techniques,

4.6. Report Writing.

4.1. Multivariate Analysis (MVA) orAdvanced Techniques for Data Analysis:

This is exactly what **Multivariate analysis** is all about. So, analysis of multiple variables simultaneously would result in a better picture to arrive at inferences instead of multiple uni-variate analyses done with the individual variables. Statistical Techniques that simultaneously analyze multiple measurements of the observed variables are known as Multivariate analysis (MVA). We may perform MVA by using multiple variables in a single relationship or in multiple relationships. In a truly multivariate scenario, all variables must be random in nature, be inter-related and cannot be interpreted in isolation. Now, let us remember this example involving the variables for computing the fuel efficiency since we would be referring to the variables mentioned in the example more than once in this session.

Since we have been using the term variables, let see what we mean by it. Well..... the dictionary says that a variable is "an element, feature, or factor that is liable to vary or change". It does make sense doesn't it? All the above

factors or variables can take different values depending on the measurement. For example the value of fuel efficiency would change when the values of the influencing variables change. So it is justified to call them variables. But the term **variate** is different. A computed *single* value of a **variate** consists of a set of variables in a linear relationship and each variable has an associated weight that is empirically determined. We can simply say that a variate result is a unitary value that represents a combination of the entire set of variables chosen to define the objective of the analysis.

The choice of the statistical technique depends on the nature of research problem or question and the nature of data set. The research questions to solve a research gap or problem may be related to identifying the degree of relationships among variables, checking for significance of group differences, predicting of group memberships or structure, or it could be time related. In-order to identify associations between two or more variables, depending on whether their nature of being parametric or non-parametric, correlation and regression or chi-square techniques may be adopted. This can be done as a Bivariate correlation and regression, multiple correlation and regression,

Canonical correlation, Multiple Discriminant Analysis and Logit regression. The bivariate correlation is a good starting point to identify the degree of relationship between two continuous variables, such as engine capacity fuel efficiency where either of them can be treated as a DV and IV as the research question may be. But bi-variate regression would require one of them to be defined as the DV and the other as the IV. Although these are not multivariate techniques they form the basis of the MVA and are part of the General Linear Model (GLM).

4.1.1. Classification of Multivariate analysis

Multivariate analysis can be classified as

4.1.2. Dependence techniques (used when there is one or more dependent variable and independent variables. Eg. Multiple regression analysis)

- Multiple regression and multiple correlation
- Multiple Discriminate Analysis (MDA) and Logistic Regression
- Canonical Correlation Analysis
- Multivariate Analysis of Variance and Covariance
- Conjoint Analysis

- Structural Equation Modelling (SEM) and Confirmatory Factor Analysis (CFA)

4.1.3. Interdependence techniques (absence of dependent or independent variables but involves techniques to simultaneously analyse all variables together in the set. Eg. Factor analysis)

- Factor Analysis (both Principal components analysis (PCA) and common factor
- analysis)
- Cluster Analysis
- Perceptual Mapping (also called as Multidimensional Scaling)
- Correspondence Analysis

4.1.4. Nature of data

The following table gives a summary of the nature of the data

Name of the Multivariate Technique	Nature of the Data	
	DV	IV

Canonical Correlation	Metric, non-metric	Metric, non-metric
MANOVA	Metric	Non-metric
ANOVA	Metric	Non-metric
MDA	Non-metric	Metric
Multiple Regression	Metric	Metric, Non-metric
Conjoint Analysis	Non-metric, Metric	Non-metric
SEM	Metric	Metric, Non-metric

(Metrics are measures of quantitative assessment commonly used for assessing, comparing, and tracking performance or production)

4.1.5. Some generic tips to perform Multivariate Analyses

While performing MVA on the research problem, it would help if the researcher would observe the following tips.

1. Ensure that both statistical and practical significance exist in the research being done.

2. The sample size should be adequate but neither under sized or over sized.

3. Clearly understand the nature of the data.

4. Use minimum number of variables in the model to obtain the desired results.\

5. Identify and eliminate errors and

6. Ensure a fool-proof validation of the results

4.2. Analysis of variance (ANOVA): is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among group means in a sample. **ANOVA** was developed by statistician and evolutionary biologist Ronald Fisher.

4.2.1 The ANOVA Test

An **ANOVA** test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.

Basically, **you're testing groups to see if there's a difference between them.**

Examples of when you might want to test different groups:

- A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.
- A manufacturer has two different processes to make Ceiling fan. They want to know if one process is better than the other.

- Students from different colleges take the same exam. You want to see if one college better performs than the other college.

4.2.2. What Does "One-Way" or "Two-Way Mean

One-way or **two-way** refers to the number of independent variables (IVs) in your Analysis of Variance test.

- One-way has one independent variable (with 2 levels). For example: **brand of cereal,**
- Two-way has two independent variables (it can have multiple levels). For example:

- **brand of cereal, calories.**

Breakfast Cereals		
Description	**Serving**	**Calories (kcal)**
Cereals ready-to-eat, corn flakes, plain, single **brand**	1 cup	360

Cereals ready-to-eat, corn, rice, wheat, oats, presweetened, with fruit and almonds, single **brand**	1.25 cup	383
Cereals ready-to-eat, corn, whole wheat, rolled oats, presweetened, single **brand**	0.75 cup	394

4.2.3. What are "Groups" or "Levels?"

Groups or levels are different groups within the same independent variable. In the above example, your levels for "brand of cereal" might be Lucky Charms, Raisin Bran, and Cornflakes -a total of three levels. Your levels for "Calories" might be: sweetened, unsweetened - a total of two levels.

Let's say you are studying if an alcoholic support group and individual counseling combined is the most effective treatment for lowering alcohol consumption. You might split the study participants into three groups or levels:

- Medication only,
- Medication and counseling,
- Counseling only.

Your dependent variable would be the number of alcoholic beverages consumed per day.

If your groups or levels have a hierarchical structure (each level has unique subgroups), then use a nested (placed or stored one inside the other) for the analysis.

4.2.4. What Does "Replication" Mean?

It's whether you are replicating (i.e. duplicating) **or (i.e. replication**. DNA **replication** is the process by which a double-stranded DNA molecule is copied to produce two identical DNA molecules. **Replication** is an essential process because, whenever a cell divides, the two new daughter cells must contain the same genetic information, or DNA, as the parent cell) your test(s) with multiple groups. With a two wayANOVA with replication, you have two groups and individuals within that group are doing more than one thing (i.e. two groups of students from two colleges taking two tests). If you only have one group taking two tests, you would use **without replication.**

4.2.5. Types of Tests.

There are two main types: one-way and two-way. Two-way tests can be with or without replication.

- One-way ANOVA between groups: used when you want to test **two groups** to see if there's a difference between them.
- Two way ANOVA without replication: used when you have **one group** and you're **double-testing** that same group. For example, you're testing one set of individuals before and after they take a medication to see if it works or not.
- Two way ANOVA with replication: **Two groups** and the members of those groups are **doing more than one thing**. For example, two groups of patients from different hospitals trying two different therapies.

4.2.6. One Way ANOVA

A one way ANOVA is used to compare two means from two independent (unrelated) groups using the F-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two means are unequal.

Examples of when to use a one way ANOVA

- Situation 1: You have a group of individuals randomly split into smaller groups and completing different tasks. For example, you might be studying the effects of tea on weight loss and form three groups: green tea, black tea, and no tea.
- Situation 2: Similar to situation 1, but in this case the individuals are split (i.e. break) into groups based on an attribute they possess. For example, you might be studying leg strength of people according to weight. You could split participants into weight categories (obese, overweight and normal) and measure their leg strength on a weight machine.

4.2.7. Limitations of the One Way ANOVA

A one way ANOVA will tell you that at least two groups were different from each other. But it won't tell you which groups were different. If your test returns a significant f-statistic, you may need to run an ad hoc test (like the Least Significant Difference test) to tell you exactly which groups had a difference in means.

4.2.8. Two Way ANOVA

A Two Way ANOVA is an extension of the One Way ANOVA. With a One Way, you have one independent variable affecting a dependent variable. With a Two Way ANOVA, there are two independents. Use a two way ANOVA when you have one measurement variable (i.e. a quantitative variable) and two nominal variables. In other words, if your experiment has a quantitative outcome and you have two categorical explanatory variables, a two way ANOVA is appropriate.

For example, you might want to find out if there is an interaction between income and gender for anxiety level at job interviews. The anxiety level is the outcome, or the variable that can be measured. Gender and Income are the two categorical variables. These categorical variables are also the independent variables, which are called factors in a Two Way

The factors can be split into levels. In the above example, income level could be split into three levels: low, middle and high income. Gender could be split into three levels: male, female, and transgender. Treatment groups are all possible combinations of the factors. In this example there would be 3 x 3 = 9 treatment groups.

4.2.9. Main Effect and Interaction Effect

The results from a Two Way ANOVA will calculate a main effect and an interaction effect. The main effect is similar to a One Way ANOVA: each factor's effect is considered separately. With the interaction effect, all factors are considered at the same time. Interaction effects between factors are easier to test if there is more than one observation in each cell. For the above example, multiple stress scores could be entered into cells. If you do enter multiple observations into cells, the number in each cell must be equal.

Two null hypotheses are tested if you are placing one observation in each cell. For this example, those hypotheses would be: H_{01}: All the income groups have equal mean stress. H_{02}: All the gender groups have equal mean stress.

For multiple observations in cells, you would also be testing a third hypothesis:

H_{03}: The factors are independent **or** the interaction effect does not exist.

An F-statistic is computed for each hypothesis you are testing.

4.2.10. Assumptions for Two Way ANOVA

- The population must be close to a normal distribution.
- Samples must be independent.
- Population variances must be equal.
- Groups must have equal sample sizes

4.3. Multiple Discriminant Analysis (MDA) and Logistic Regression Analysis

If the dependent variable is dichotomous (Yes/No, Men / Women) type then MDA is an appropriate technique. The independent variables need to be metric data. MDA helps to understand group differences and to predict the possibility that an observation or object would belong to a specific group. In example that we had discussed in MR in the previous section, suppose we had data on the engine capacity and un-laden weight of about 100 plus cars (that run on the same type of fuel) and if we want to class them as Big and Small cars then MDA would be a relevant technique.

Logistic Analysis also known as Logistics regression is a combination of MR and MDA. Although the regression principle is similar to that of MR, the DV in Logistic regression need not be metric as in the case of MR but can be

a dichotomous variable as in MDA. Another, distinguishing fact of Logistic regression is that it can accommodate both metric and non-metric IVs and overlooks the multivariate normality assumption

4.4. Factor Analysis

The objective of factor analysis is to reduce the number of measured variables into meaningful factors (or variates) with minimal loss of information. This can either be done by the PCA (Principal components analysis) method or by common factor analysis. Suppose a prospective car buyer is considering the colour of the car, the aerodynamic design, body coloured bumpers, height adjustable steering column, driver seat height adjustment, touch screen for infotainment, ABS (Anti-lock Braking System) and Airbags. If the opinion of the car buyer is captured using a 7 point Likert' scale, either PCA or common factor analysis may group these eight variables in three groups, namely external features (colour of the car, the aerodynamic design, body coloured bumpers), internal features (height adjustable steering column, driver seat height adjustment, touch screen for infotainment) and safety features (ABS and Airbags).

So factor analysis helps us to reduce eight variables into three meaningful factors (variates).

4.4.1. Procedure of factor Analysis:

The relationship of each variable to the underlying factor is expressed by the so-called factor loading. Here is an example of the output of a simple factor analysis looking at indicators of wealth, with just six variables and two resulting factors.

Variables	Factor 1	Factor 2
Income	0.65	0.11
Education	0.59	0.25
Occupation	0.48	0.19
House value	0.38	0.60
Number of public parks in neighborhood	0.13	0.57
Number of violent crimes per year in neighborhood	0.23	0.55

The variable with the strongest association to the underlying latent variable. Factor 1, is income, with a factor loading of 0.65. Since factor loadings can be interpreted like standardized regression coefficients, one could also say that the variable income has a correlation of 0.65 with Factor 1. This would be considered a strong association for a factor analysis in most research fields. Two other variables, education and occupation, are also

associated with Factor 1. Based on the variables loading highly onto Factor 1, we could call it “Individual socioeconomic status.”

House value, number of public parks, and number of violent crimes per year, however, have high factor loadings on the other factor, Factor 2. They seem to indicate the overall wealth within the neighbourhood, so we may want to call Factor 2 “Neighbourhood socioeconomic status.”

Notice that the variable house value also is marginally important in Factor 1 (loading = 0.38). This makes sense, since the value of a person’s house should be associated with his or her income.

4.5. Cluster Analysis

Cluster analysis is a statistical method used to group similar objects into respective categories. It can also be referred to as segmentation analysis, organization analysis, or clustering. The goal of performing a cluster analysis is to sort different objects or data points into groups in a manner that the degree of association between two objects is high if they belong to the same group and low if they belong to different groups.

In the car example that we have been discussing so far, suppose we have the data on engine capacities of about 130 cars. The engine capacities range from a minimum of 799 cc to 2399cc. If we want these 130 cars to be placed in three groups namely small, medium and large cars, cluster analysis would be a recommended technique. The Cluster analysis algorithm places the objects in homogenous groups depending on the characteristics specified by the researcher. In our example, the cars would be placed in groups based on engine capacity. Clustering can be done based on multiple characteristics too. Either hierarchical or non-hierarchical clustering procedures may be adopted. Basically hierarchical methods could be either agglomerative or divisive. The algorithms followed in the hierarchical methods are single, complete and average linkage methods. The other methods are the Centroid and Ward' methods. Alternatively the non-hierarchical clustering popularly follows the k-means algorithm and places objects in cluster groups once the number of clusters are specified. The decision on whether to adopt the hierarchical or non-hierarchical procedure depends on the choice of the researcher and the problem defined.

4.5.1. Common Applications of Cluster Analysis:

- **Marketing:** Marketers commonly use cluster analysis to develop market segments, which allow for better positioning of products and messaging. Company to better position itself; explore new markets, and development products that specific clusters find relevant and valuable.
- **Insurance:** Insurance companies often leverage cluster analysis if there are a high number of claims in a given region. This enables them to learn exactly what is driving this increase in claims.
- **Education:** What are student groups that need special attention? Researchers may measure psychological, aptitude, and achievement characteristics. A cluster analysis then may identify what homogeneous groups exist among students (for example, high achievers in all subjects, or students that excel in certain subjects but fail in others).
- **Medicine:** What are the diagnostic clusters? To answer this question the researcher would devise a diagnostic questionnaire that includes possible symptoms (for

example, in psychology, anxiety, depression etc.). The cluster analysis can then identify groups of patients that have similar symptoms.

4.5.2. Three Types of Cluster Analysis:

There are three primary methods used to perform cluster analysis:

- Hierarchical Cluster: This is the most common method of clustering. It creates a series of models with cluster solutions from 1 (all cases in one cluster) to n (each case is an individual cluster). This approach also works with variables instead of cases. Hierarchical clustering can group variables together in a manner similar to factor analysis. Finally, hierarchical cluster analysis can handle nominal, ordinal, and ratio scale data. But, remember not to mix different levels of measurement into your study.
- K-Means Cluster: This method is used to quickly cluster large datasets. Here, researchers define the number of clusters prior to performing the actual study. This approach

is useful when testing different models with a different assumed number of clusters.

- Two-Step Cluster: This method uses a cluster algorithm to identify groupings by performing pre-clustering first, and then performing hierarchical methods. Two-step clustering is best for handling larger datasets that would otherwise take too long a time to calculate with strictly hierarchical methods. Essentially, two-step cluster analysis is a combination of hierarchical and k-means cluster analysis. It can handle both scale and ordinal data and it automatically selects the number of clusters.

4.5.3. Process of Cluster Analysis:

- Step 1: Build and Distribute a Survey: Your survey should be designed to include multiple measures of propensity to purchase and the preferences for the product at hand. It should be distributed to your population of interest, and your sample size should be large enough to inform statistically-based decisions.

- Step 2: Analyze Response Data: It's considered best practice to perform a factor analysis on your survey to minimize the factors being clustered. If after your factor analysis it's concluded that a handful of questions are measuring the same thing, you should combine these questions prior to performing your cluster analysis.

  After reducing your data by factoring, perform the cluster analysis and decide how many clusters seem appropriate, and record those cluster assignments. You'll now be able to view the means of all of your factors across clusters.

- Step 3: Take informed opinion: Search through your data to identify differences in the means of factors, and name your clusters based on these differences. These differences between clusters are then able to inform your marketing, allowing you to target precise groups of customers with the right message, at the right time, in the right manner.

4.6. Report Writing:

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented or written. As a matter of fact even the most

brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others. 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 ofthe research process. But the general opinion is in favor of treating the presentation of researchresults or the writing of report as part and parcel of the research project. Writing of report is the laststep in a research study and requires a set of skills somewhat different from those called for inrespect of the earlier stages of research. This task should be accomplished by the researcher withutmost care; he may seek the assistance and guidance of experts for the purpose

4.6.1. Different Steps in Writing Report

Research reports are the product of slow, painstaking, accurate inductive work. The usual steps involved in writing report are:

1. logical analysis of the subject-matter;

2. preparation of the final outline;
3. preparation of the rough draft;
4. rewriting and polishing;
5. preparation of the final bibliography; and
6. 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.

1. 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

1. **Logical**: The logical development is made on the basis of mental connections and associations between the one thing and another by means of analysis. Logical treatment often consists in developing the material from the simple possible to the most complex structures
2. **Chronological:** Chronological development is based on a connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order.

2. 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.

4. Rewriting and polishing of the rough draft*:*This step happens to be most difficult part of all formal writing. Usually this step requires more time than the writing of the rough draft. The careful revision makes the difference between a mediocre and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation. The researcher should also "see whether or not

the material, as it is presented, has unity and cohesion; does the report stand upright and firm and exhibit a definite pattern, like a marble arch? Or does it resemble an old wall of moldering cement and loose brick." 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.

5. Preparation of the final bibliography:Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally included 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**;

1. The first part may contain the names of books and pamphlets, and
2. 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.

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

Layout of the Research Report

Anybody, who is reading the research report, must necessarily be conveyed enough about the study so that he can place it in its general scientific context, judge the adequacy of its methods and thus form an opinion of how seriously the findings are to be taken. For this purpose there is the need of proper layout of the report. The layout of the report means as to what the research report should contain. A comprehensive layout of the research report should comprise.

1. Preliminary pages;
2. The main text; and
3. The end matter. Let us deal each point above separately.

1. Preliminary Pages

In its preliminary pages the report should carry a title, author names 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, list of charts, list of graphs and illustrations so that the decision-maker or anybody interested in reading the report can easily locate the required information in the report.

2. **Main Text**

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

1. Introduction;
2. Statement of findings and recommendations;
3. The results;
4. The implications drawn from the results; and
5. The summary.

1. **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 Generalizations ability 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.

2. **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.
3. **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 spitted into readily identifiable

sections. All relevant results must find a place in the report. But how one is to decide about what is relevant is the basic question. Quite often guidance comes primarily from the research problem and from the hypotheses, if any, with which the study was concerned. But ultimately the researcher must rely on his own judgment in deciding the outline of his report. “Nevertheless, it is still necessary that he states clearly the problem with which he was concerned, the procedure by which he worked on the problem, the conclusions at which he arrived, and the bases for his conclusions.”

4. **Implications of the results**:Toward the end of the main text, the researcher should again put down the results of his research clearly and precisely. He should, state the implications that flow from the results of the study, for the general reader is interested in the implications for understanding the human behaviour. Such implications may have three aspects as stated below:

a) A statement of the inferences drawn from the present study which may be expected to apply in similar circumstances.

b) The conditions of the present study which may limit the extent of legitimate generalizations of the inferences drawn from the study.

c) The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them.

It is considered a good practice to finish the report with a short conclusion which summarises and recapitulates the main points of the study. The conclusion drawn from the study should be clearly related to the hypotheses that were stated in the introductory section. At the same time, a forecast of the probable future of the subject and an indication of the kind of research which needs to be done in that particular field is useful and desirable.

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

3. End Matter:

At the end of the report, appendices should be enlisted in respect of all technical data such as questionnaires, sample information,

mathematical derivations and the like 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.

Short Questions

1. What is multivariate analysis? Explain
2. Explain the concept of cluster analysis
3. Factor analysis
4. Decision reports
5. Bibliography
6. Footnotes

Essay Questions

1. Discuss the structure of a research report in detail
2. What is data analysis? Explain the use of factor analysis in data analysis with suitable example.
3. Define a Research report and explain its purpose.

4. What is the purpose of footnotes? what are the different kinds of footnotes
5. Explain the need and importance of report writing
6. What is factor analysis? Explain the procedure.
7. What is the general structure of a research report
8. What is cluster analysis? Explain its procedure.
9. What are quotations used in a research report? Explain briefly
10. What points have to be kept in mind while revising the draft report
11. What is a Research report? Explain the format of a research report
12. What is ANOVA? Explain its usage in business applications. Also explain the procedure

Unit – V: Business Analytics

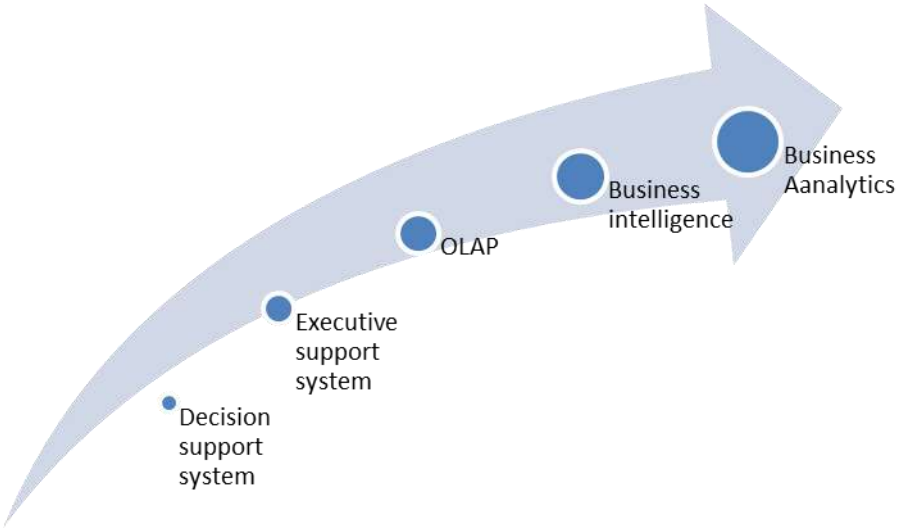
5.1. Evolution of Business Analytics

5.1. Evolution of Business Analytics:

Analytics has been an integral part of a firm's operation since early times. However with the passage of time analytics has become more Complex and refined and wide in its applications. The term ‘business analytics’ has evolved over the period of time.

5.1.1 The decision support system: emerged in the 1970s and was designed to serve the needs of the managers in the organization. These systems are captured in the internal information through the daily transaction reports on regular

reports. Along with this is the systems also capture external data. This data is then analyzed using various modeling software. This feature also allows the users to work directly and they are flexible enough to answer a large number of queries of the users. They are interactive systems that allow the user to change assumptions, include new data and ask more and more questions. The main components of a decision support system are a database, a software system and a user interface.



The growth of the concept of business analytics

5.1.2. The executive support system: evolves to a higher level that supports non-routine decision making. These systems are generally used by senior level managers who often have to handle non-routine strategic decision making. These systems have the ability to filter, spot and track critical data and thus

help the managers make a well informed decision. They became popular during the 1980s decade .These systems were superior as they had good computing and communicating capabilities that was designed to solve not a specific problem but were rather capable of solving a wide array of problems.

5.1.3. Online Analytical Processing (OLAP): The major development in the series of analytical processing was the emergence of online analytical processing (OLAP). It allowed the organisation to perform analysis along multiple dimensions. It is the dynamic synthesis, analysis and consolidation of large value of multidimensional data. It enables users to gain a deeper understanding and knowledge about various aspects of their corporate data through past consistent and interactive access to a wide variety of possible views of the data. The same data is viewed from different angles with each viewpoint having a new interpretation.

5.1.4. **Business Intelligence:** the evolution of a better analytical capability is, the firm's moved to business intelligence IBM researcher Hans Peter Luhn coined the term business intelligence and defines business intelligence as “the ability to understand the interrelationship of presented facts in such a way

as to guide action towards a desired goal" it combined the data with the analytical abilities to support better decision making. It turns raw data into an actionable form. It involves the process of accessing data, formatting it and delivering it inside and outside the organization. It makes use of various data mining techniques to identify patterns and unearth(to discover something in the ground)relationships. It tries to establish a correlation between data and Real world objects and there after predictive probabilities and trends.

Some of the applications of business intelligence are:

- To study the consumer behavior
- To develop and manage customer relationships
- To conduct detailed sales analysis and forecast future sales
- Budgeting and financial planning and forecasting
- To design effective delivery and supply chain
- To develop enhanced decision making under risk and uncertainty
- Increase organization's effectiveness and boost internal productivity
- Business intelligence is an important tool that empowers the employees by giving them access to real time analytical

and operational data. This increases the work efficiency of the employees. With the passage of time business intelligence has become more and more specialized and from its core emerges the concept of business analytics.

5.1.5. Business analytics: business analytics is the use of information technology, modeling software on the data to find solutions to the problems being faced by the business. It is the use of various applications on the large volume of data being generated in the organisation to reveal meaningful relationships and patterns. It has been defined as "the study of data through statistical and operation analysis, the formation of predictive models, applications of Optimisation techniques and the communication of these results to customers, business patterns and executives".

1) According to Liberto and Luo, business analytics is the " process of transforming data into actions through analysis and insights in the context of organisational decision making and problem solving"
2) As per James R Evans " business analytics is the use of data information, technology , statistical analysis quantitative methods, mathematical or Computer Based

models to help managers gain improved insight about their business operations and make a better fact based decisions' Business analytics has found widespread applications in the managerial decision making. Some of the areas where decision making using analytics as improved are are;

- Pricing decision
- Market segmentation
- Positioning decisions
- Location decisions

5.2. Master Data Management

5.2.1. Meaning of Master Data is the data which refers to Core business entities and Organisation uses repeatedly across many business processes and Systems.

Master Data captures the key things that all parts of an organisation must agree on both in meaning and in usage. Master Data includes the business objectives, definitions, classification and terminology that constitute business information, consequently Master Data forms the business for business Processes.

5.2.2. Definition of Master Data Management

Master Data Management Is a collection of best data management practices that coordinate key stakeholders, participants and business clients in incorporating the business applications, information management methods, And data management tools to implement the policies, procedures services, and infrastructure to support the capture integration and subsequent shared use of accurate timely, consistent and complete Master Data.

5.2.3. List out the Characteristics of good Master Data Management System

In order to successfully manage the Master Data support corporate governance and augment business intelligence systems The Master Data Management applications must have the following characteristics.

- A metadata management capability for items such as business entity Matrixes relationship and hierarchies.
- A source system management capability to fully cross - reference business objects and to satisfy seemingly conflicting data ownership requirements.

- A data quality function that can find and eliminate Duplicate data while ensuring correct data attribute superior ship
- A Data quality interface to assist with preventing new errors from entering the System even when data entry to outside the Master Data Management applications itself.
- A Continuing data cleansing function to keep the data up to date.
- An integral set off mechanism to create and position change information to all connected systems
- A comprehensive data security system to control and monitor data access, update rights and maintain change history.

5.2.4. Advantages of Master Data Management

Business concerns how many advantages with the maintenance of Master Data Management, the benefits are as follows.

- Increase information quality

- Improve business productivity and operational efficiency
- Improve interoperability
- Improve governance, regulatory compliance and risk
- Simplify and accelerate application development
- Enable consistent reporting and improve decision making
- Enable comprehensive customer knowledge

5.3. Data warehousing

5.3.1. Meaning of data warehouse: A large storage of data accumulated from a wide range of sources within a company and used to guide management decisions.

a data warehouse also known as an enterprise data warehouse is a system used for reporting and data analysis and is considered a core component of business intelligence

5.3.2. Definition of data warehousing: the process of constructing and using a data warehouse can be defined as data warehousing. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and ad hoc queries and decision making. Data warehousing involves data cleaning data integration and data consolidation.

5.3.3Features of data warehouse:

1. **Subject oriented**: a data warehouse is subject oriented because it provides information around a subject rather than the organisation's on-going operations. These subjects can be product, customers, suppliers, sale, revenues etc. a data warehouse does not focus on the on-going operations, rather it focuses on Modelling and Analysis of data for decision making.
2. **Integrated:** data warehouse is constructed by integrating data from Heterogeneous sources such as a relational databases, flat files etc. This integration enhances the effective Analysis of data.
3. **Time variant:** the data collected in data warehouse in identified with a particular time period. The data in a data warehouse provides information from the historical point of view.
4. **Non-volatile**: non-volatile means the provides a data is not erased when new added to it. A data warehouse is kept separate from the operational database and therefore frequent changes in operational database is not reflected in the data warehouse.

5.3.4. Functions of data warehouse tools.

1. **Data extraction**: involves gathering data from multiple heterogeneous sources.
2. **Data cleaning**: involves finding and correcting the errors in data.
3. **Data transformation**: involves converting the data from Legacy format to warehouse format.
4. **Data loading:** involves storing, summarising, consolidating checking, integrity e building indices and partitions.
5. **Repressing**: involves updating from Data sources to warehouse.

5.3.5. Advantages of data warehousing

Maintenance of data warehouses will yield different benefits to the users, which are as follows.

- Data warehouse maintains data history, even if the source transaction systems do not.
- It integrates data from multiple source systems, enabling a Central View across the Enterprise this benefit is always valuable, but particularly so when the organisation has grown by Merger.

- Data warehouse improve data quality, by providing consistent codes and descriptions weakening. or even fixing bad data.
- it presents The organisations information consistently
- Data warehouse provides a single common data model for all data of interest regardless of the data's source.

5.4. Extraction, transformation, and loading (ETL):

Processes are responsible for the operations taking place in the background of data warehouse architecture. In a high leveldescription of an ETL process, first, the data are extracted from the source data stores that can be on-line transaction processing (OLTP) or legacy systems, files under any format, web pages, various kinds of documents (e.g., spreadsheets and text documents) or even data coming in a streaming fashion. Typically, only the data that are different from the previous execution of an ETL process (newly inserted, updated, and deleted information) should be extracted from the sources. Secondly, the extracted data are propagated to a special-purpose area of the warehouse, called the data staging area (DSA), where

their transformation, homogenization, and cleansing take place. The most frequently used transformations.

5.4. **Data Mining:** Data mining finds it greatest application in retail sector. Data mining is a powerful tool in the hands of a marketing manager where he can make some general observations like spotting the sales trend or perform some specific actions like design one to one marketing and design personalized messages to engage the customers fully with the organisations.

5.5.1 The Advantages of Data Mining:

- **Market Segmentation**: Identify the common characteristics of customers who buy the same products from your company
- **Customer churn:** Predict which customers are likely to leave your company and go to a competitor.
- **Fraud Detection:** Identify which transactions are most likely to be fraudulent

- **Direct Marketing**: Identify which prospects should be included in a mailing list of obtain the highest response rate
- **Interactive Marketing**: Predict what each individual assessing a web site is most likely interested in seeing
- **Market Basket Analysis**: Understand what products are services are commonly purchased together
- **Trend Analysis:** Reveal the differences between a typical customer this month and last

5.6. Meta data

The term Often used in the context of database is ‘meta data’ It means data about data. The objects in a database have attributes that describe the date. Example the date item ‘name’ can be described in terms of the nature of data item, its size, its length etc. All this information about the attributes of the data object is called as metadata for example a field by the name of ‘address’ exists in the database. This field has been divided into three sub fields, namely house number, street name, locality. The information about the individual address details is the data content of the database. At the same time information about

each of these subfields in terms of The Identity, data type, count etc. Is also stored.

5.7. Data marts

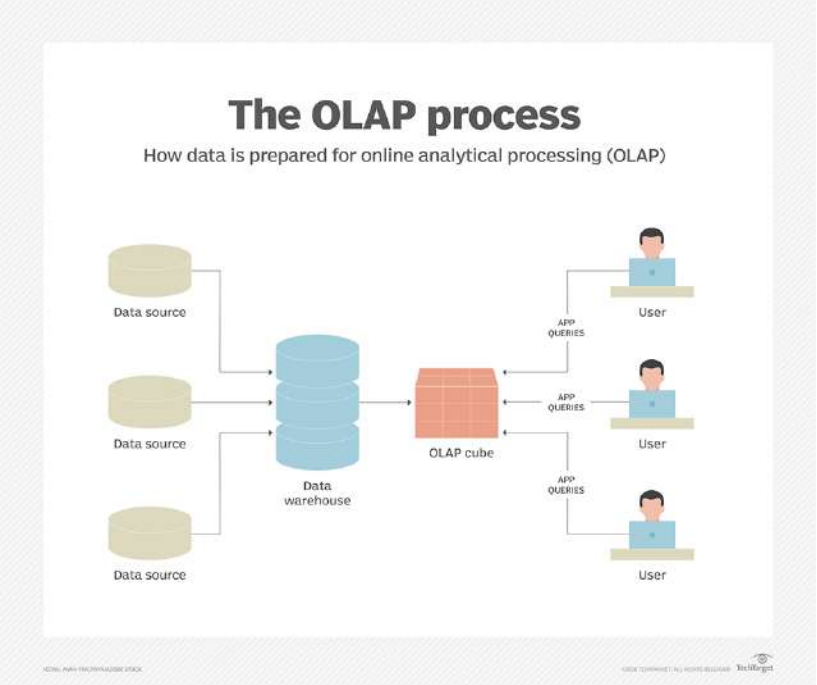
A date mart is a subset of data warehouse that contains some of the data that is contained in a data warehouse. it is a subject oriented database and it generally caters to the needs of the middle level managers in an organisation. General different department within the Organisation may create these departments in Order to have a subset of data specific to their department. The data in specific to some user type which makes it easy for the user to construct analysis on this.

5.8. Data integration

It is the process of combining data from disparate sources to present a Unified of view of data. Generally it involves the process of Discovery, cleansing monitoring and transformation of data. The objective of data integration is to consolidate the data from various sources while still maintaining the integrity and quality of data. Generally for any business organisation data integration ensures consistent, trusted and reliable real time date.

5.9. Online Analytical Processing (OLAP):

5.9.1 Define OLAP: OLAP stands for online analytical processing, which is characterized by relatively low volume of transactions. Queries are often very complex and involve aggregations. For OLAP systems a response time is an effectiveness measure. OLAP applications are widely used by data mining techniques. In OLAP database there is aggregated, historical data stored in multi-dimensional schemas.



5.9.2. Characteristics of OLAP

According to business intelligence Limited OLAP Can be defined in 5 words Past Analysis of Share Multidimensional Information (FASMI)

1. **Fast:** past means that the system targeted to deliver most responses to use within about 5 second, with the simplest analysis taking no more than one second and very few taking more than 20 seconds.
2. **Analysis**: analysis means that the system can cope with any business logic and statistical analysis that it relevant for the application and other user keep it easy enough for the target user.
3. **Share:** share means that the system implements the entire security requirement for confidentiality and if multiple write access is needed, concurrent update location at an appropriated level not all applications need uses to write data back but for the growing number that do, The system should be able to handle multiple updates in a timely secure manner.
4. **Multidimensional:** multidimensional is the key requirement. OLAP system must provide a multidimensional conceptual view of the data including Full support for hierarchies, as This Certainly the most logical way to analyse business and Organisation.

5. **Information:** information is all of the data and derived information needed. Wherever it is and however much is relevant for the application. We are measuring the capacity of various product in terms of how much input data they can handle, not how many Giga bytes they take to store it.

5.10. Online Transaction Processing (OLTP):

5.10.1 Define OLTP: OLTP Stands for online transaction processing which is characterized by a large number of start online transaction (read, insert, update and delete) The main purposes for OLTP systems is put on very fast query processing, maintaining Data integrity in multi access environment and an effectiveness measured by number of transactions per second. In OLTP database there is a detailed and current data and schema used to store transactional databases in the entity model.

5.10.2 Kinds of processing in OLTP: generally the data in an OLTP can be processed using any of the 3 methods of batch processing, online processing or online validation with batch processing

1. Batch Processing.
2. Online Transaction Processing and
3. Online Entry With Batch Processing

5.10.3. A Comparison of OLTP and OLAP

Parameters	OLTP	OLAP
Purpose	Designed for real time business operations.	Designed for analysis of business measures by category and attributes.
Source	OLTP and its transactions are the sources of data.	Different OLTP databases become the source of data for OLAP.
Design	DB design is application oriented. Example: Database	DB design is subject oriented. Example: Database design

	design changes with industry like Retail, Airline, Banking, etc.	changes with subjects like sales, marketing, purchasing, etc.
Process	It is an online transactional system. It manages database modification.	OLAP is an online analysis and data retrieving process.
Characteristic	It is characterized by large numbers of short online transactions.	It is characterized by a large volume of data.
Method	OLTP uses traditional DBMS.	OLAP uses the data warehouse.
Functionality	OLTP is an online database modifying system.	OLAP is an online database query management system.
Response time	Its response time is in millisecond.	Response time in seconds to minutes.
Usefulness	It helps to control and run fundamental business tasks.	It helps with planning, problem-solving, and decision support.
Productivity	It helps to Increase user's self-service and productivity	Help to Increase productivity of the business analysts.

Short Questions

1. Data integration
2. Scope of business analytics.
3. Data warehousing
4. Business analytics

Essay Questions

1. What is data mining? Explain in detail
2. Explain the importance and evolution of Business analytics
3. Discuss the concept of data mining
4. Briefly explain the master data management in detail.
5. Explain the concept of transformation and uploading of data.
6. Difference between of OLTP and OLAP.

References:

1. Bhattacharya D. K., "Research Methodology", Excel Books, New Delhi.
2. Cooper, "Business Research Methods", Tata McGraw Hill, New Delhi, 2010.
3. C.R. Kothari, "Research Methodology: Methods and Techniques", New Age International Publishers, New Delhi, 2006.
4. Gupta S.P. "Statistical Methods", Sultan Chand, New Delhi, 2010.

5. K.V. Rao, “Research Methodology in Commerce and Management”, Sterling Publishers, New Delhi, 2012.
6. Sriwastava, S. C.: Foundation of Social Research and Economics Techniques, Himalaya Publishing House, 1990.
7. Chou, Ya-Lun : Statistical Analysis with Business and Economics Applications, 2nd Eds., New York, Hold Rinchart and Wrintston, 1974.
8. Clover, Vernon t and Balsely, Howerd L: Business Research Methods, Colombus O. Grid, Inc, 1974.
9. Emary C. Willima: Business Research Methods, Illinois : Richard D. Irwin Inc. Homewood, 1976.
10. Sharma H.D. and Mukherji S. P. : Research Methods in Economics and Business, New York : The Macmillan Company, 1992.
11. Gerber R. and Verdoom, P.J.: Research Methods in Economics and Business, New York, the Macmillan Company, 1992.
12. Courtis J.K. (ed.) Research and Methodology in Accounting & Financial Management, 1980.
13. Menden HYall and Varacity: Reinmuth J.E.: Statistics for Management and Economics (2nd Edition), 1982.

14. Krishnaswami O.R.: Methodology of Research in Social Sciences, Himalaya Publishing House, 1993.

15. Beckerian, D.A. (1993). In search of the typical eyewitness. *American psychologist*, 48, 574-576.