BBA/B.COM - IV YEAR

SUBJECT – RESEARCH METHODOLOGY

UNIT - 1	**Introduction of Research Methodology:** Meaning of research, Objectives of research, Motivation in research, Types of research, Research approaches, Significance of research, Research and scientific method, Research process, Criteria of good research, Problems faced by researchers.
UNIT - 2	**Identification and Formulation of Research Problem:** Definition, Objectives of research problem, Selection of problem, Identification of problem, Necessity and techniques of defining research problem, Formulation of research problem, Experimental research design.
UNIT - 3	**Review of Literature:** Searching for the existing literature, Need and significance, Reviewing the selected literature, Developing a theoretical framework, Developing a conceptual framework, Writing about the literature reviewed.
UNIT - 4	**Research andSampling Design:** *Research Design:* Meaning, Need and features of good research design,Important concepts relating to research design, Types of research designs, Basic principles of experimental designs, Developing a research plan. *Sampling Design:* Implications of a sample design, Steps in sampling design, Criteria of Selecting a Sampling Procedure, Characteristics of idealsample design, Different types of Sample Designs, Selectionofrandom sample, Random sample from an infinite Universe, Complex random sampling designs, Sampling V/s non-sampling error.
UNIT - 5	**Testing of Hypotheses and Analysis of Data:** Introduction tohypothesis, Basic concepts concerning testing of hypotheses, Procedure for hypothesis testing, Flow diagram for hypothesis testing, Measuring the power of a hypothesis test, Tests of hypotheses, Important parametric tests, Univariate and bivariate analysis of data, Analysis of variance, Parametric and non-parametric test.

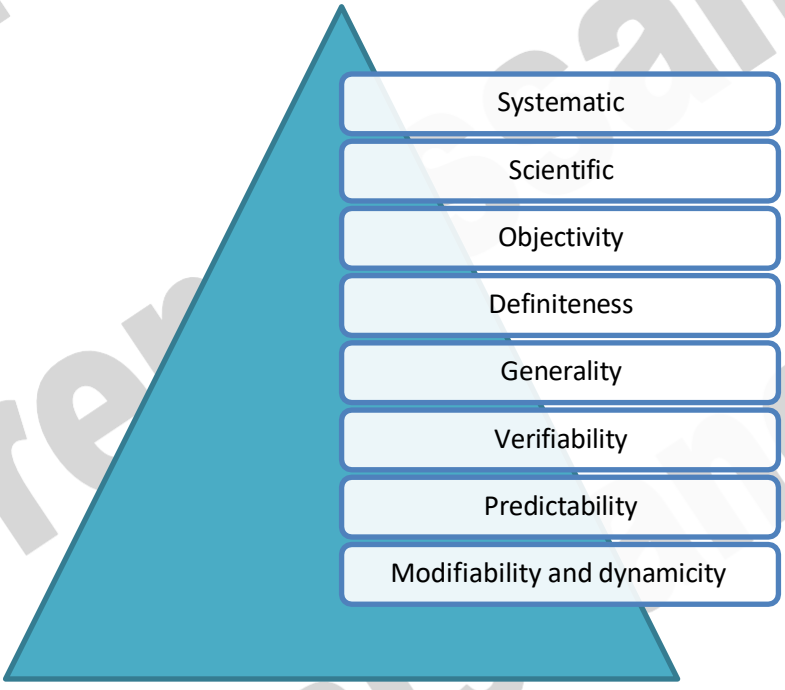
UNIT – 1

MEANING & TYPES OF RESEARCH

Meaning:

1) It is a scientific and systematic search for pertinent information on specific topic.
2) It is an art and as well as science of investigationυ Research may be defined as a 'careful critical enquiry or examination in seeking facts or principles; diligent investigation in order to ascertain something'.
3) Research common sense of the term refers to a search for knowledge. Research is a part and parcel of human knowledge.
4) Gathering and analyzing a body of information or data and **extracting new meaning** from it or **developing unique solutions** to problems or cases.
5) A **report** or **review**, not designed to create new information or insight but to collate and synthesize existing information.
6) A search for **individual facts or data**. May be part of the search for a solution to a larger problem

Nature of research



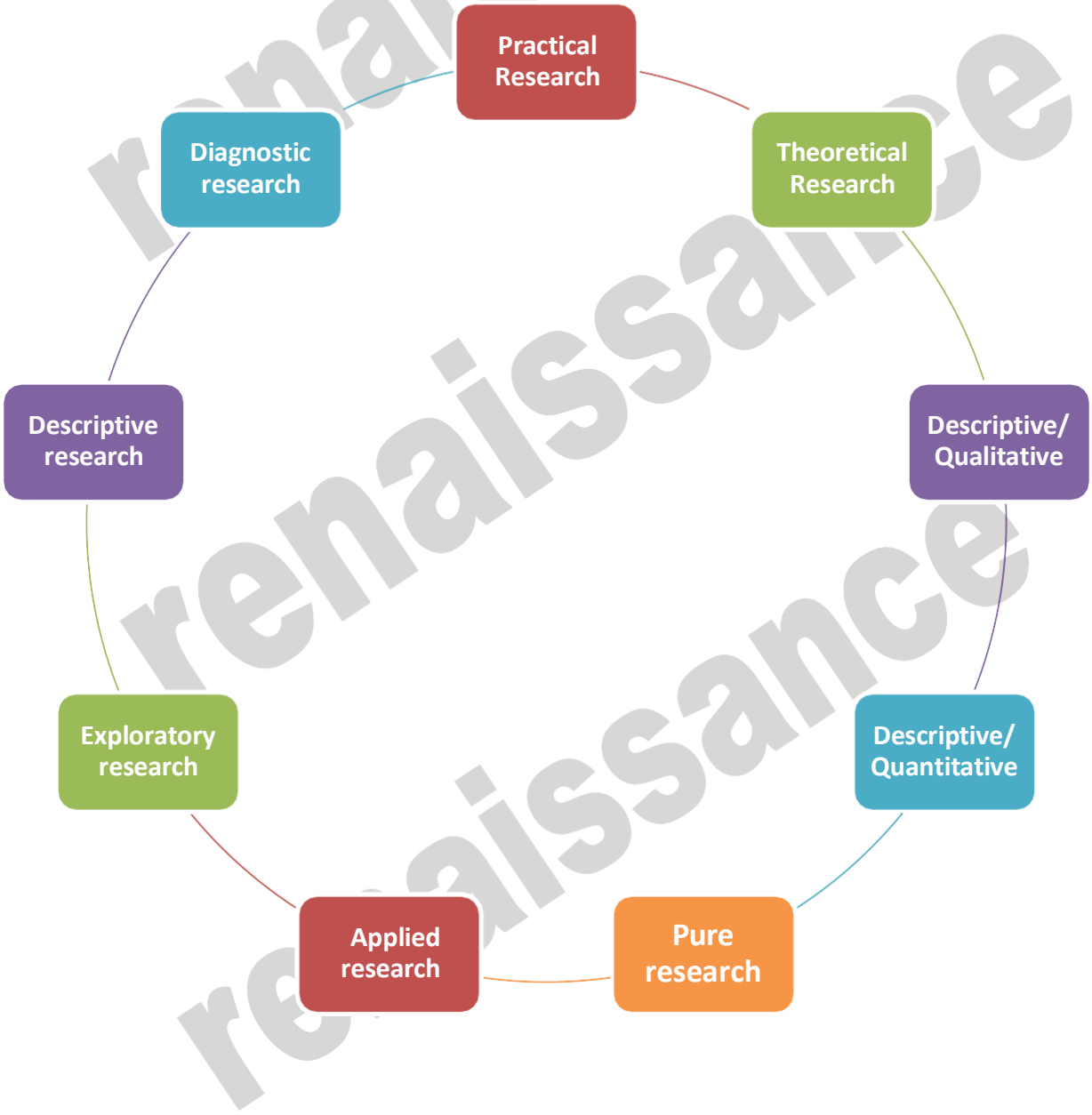
1. **Research is systematic and Scientific**- Research is a scientific and systematic search for pertinent information on a specific topic. • Generally, research has to follow a certain structural process.
2. **Research has objectivity** – Research is quite objective in its approach and is almost free from biases, prejudices and subjectivity.

3. **Research has definiteness** - Research is characterized by definiteness in its process as well as product. Here the modes and measures for (i) collection and organizing information or data and (ii) testing and verifying the collected information for arriving at the conclusion are all well planned and definite.
4. **Research has Verifiability**– Research lays emphasis on the proper verification of the collected information, data or facts. Here, nothing is accepted and derived unless verified through adequate observation, tests and experimentation.
5. **Research has Generality**– The conclusions or results derived from the scientific method show a marked characteristic of generality. First, it means that inductive reasoning and process is sued in making generalization and of the particular happenings or evens and secondly, the principles, laws and theories established through scientific method are quite universal having generalized application in similar situations.
6. **Research has Predictability**– The results obtained through scientific method are characterized with the ability of predicting the future outcomes of the things or events. In a given situation, under the known circumstances, what would happen to a person, object or phenomenon can be reasonably predicted through the properly derived conclusions or results of a scientific procedure.
7. **Research has modifiability and dynamicity**– The conclusion reached or results obtain through research are never final, absolute and static. They are always open to verification, observation and experimentation.
8. **Research has modifiability and dynamicity** – Consequently, what is true today in terms of the derived fact or reached generalization may be proved wrong tomorrow based on new findings. Therefore, research neither advocates rigidity in the process adopted for discovering the facts nor stands in the way of bringing desired modification and changes in the pre-established principles, laws or theories.

Significance of Research

- Research provides the basic for nearly all government policies in our economic system.
- The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times.
- Research inculcates scientific and inductive thinking and it promotes the development of logical habits of thinking and organization.
- According to Hudson Maxim Significance as, “All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry and inquiry leads to investigation”
- Research is equally important for social scientists in studying social relationships and in seeking answers 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.
- Research has its special significance in solving various operational and planning problems of business and industry. In several ways, operations research, market research and motivational research are vital and their results assist in taking business decisions.
- Research provides the basis for nearly all government policies in our economic system.

Types of RESEARCH:



1) **Practical Research:** The practical approach consists of the empirical study of the topic under research and chiefly consists of hands on approach. This involves first hand research in the form of questionnaires, surveys, interviews, observations and discussion groups.
2) **Theoretical Research:** A non empirical approach to research, this usually involves perusal of mostly published works like researching through archives of public libraries, court rooms and published academic journals.
3) **Descriptive/Qualitative**: This type of research methods involve describing in details specific situation using research tools like interviews, surveys, and Observations. It focuses on gathering of mainly verbal data rather than measurements.
4) **Descriptive/Quantitative**: This type of research methods requires quantifiable data involving numerical and statistical explanations. Quantitative analysis hinges on researchers understanding the assumptions inherent within different statistical models. It generates numerical data or information that can be converted into numbers. The presentation of data is through tables containing data in the form of numbers and statistics.
5) **Pure research**
   a. Also called as the fundamental or the theoretical research.
   b. Is basic and original.
   c. Can lead to the discovery of a new theory.
   d. Can result in the development or refinement of a theory that already exists.
   e. Helps in getting knowledge without thinking formally of implementing it in practice based on the honesty, love and integrity of the researcher for discovering the truth.
6) **Applied research**
   a. Based on the concept of the pure research.
   b. Is problem oriented.
   c. Helps in finding results or solutions for real life problems.
   d. Provides evidence of usefulness to society.
   e. Helps in testing empirical content of a theory.
   f. Utilizes and helps in developing the techniques that can be used for basic research.
   g. Helps in testing the validity of a theory but under some conditions.
   h. Provides data that can lead to the acceleration of the process of generalization.
7) **Exploratory research**
   a. Involves exploring a general aspect.
   b. Includes studying of a problem, about which nothing or a very little is known.
   c. Follows a very formal approach of research.
   d. Helps in exploring new ideas.
   e. Helps in gathering information to study a specific problem very minutely.
   f. Helps in knowing the feasibility in attempting a study.
8) **Descriptive research**
   a. Simplest form of research.
   b. More specific in nature and working than exploratory research.
   c. It involves a mutual effort.
   d. Helps in identifying various features of a problem.
   e. Restricted to the problems that are describable and not arguable and the problems in which valid standards can be developed for standards.
   f. Existing theories can be easily put under test by empirical observations.
   g. Underlines factors that may lead to experimental research.

h. It consumes a lot of time.
i. It is not directed by hypothesis.

9) Diagnostic research

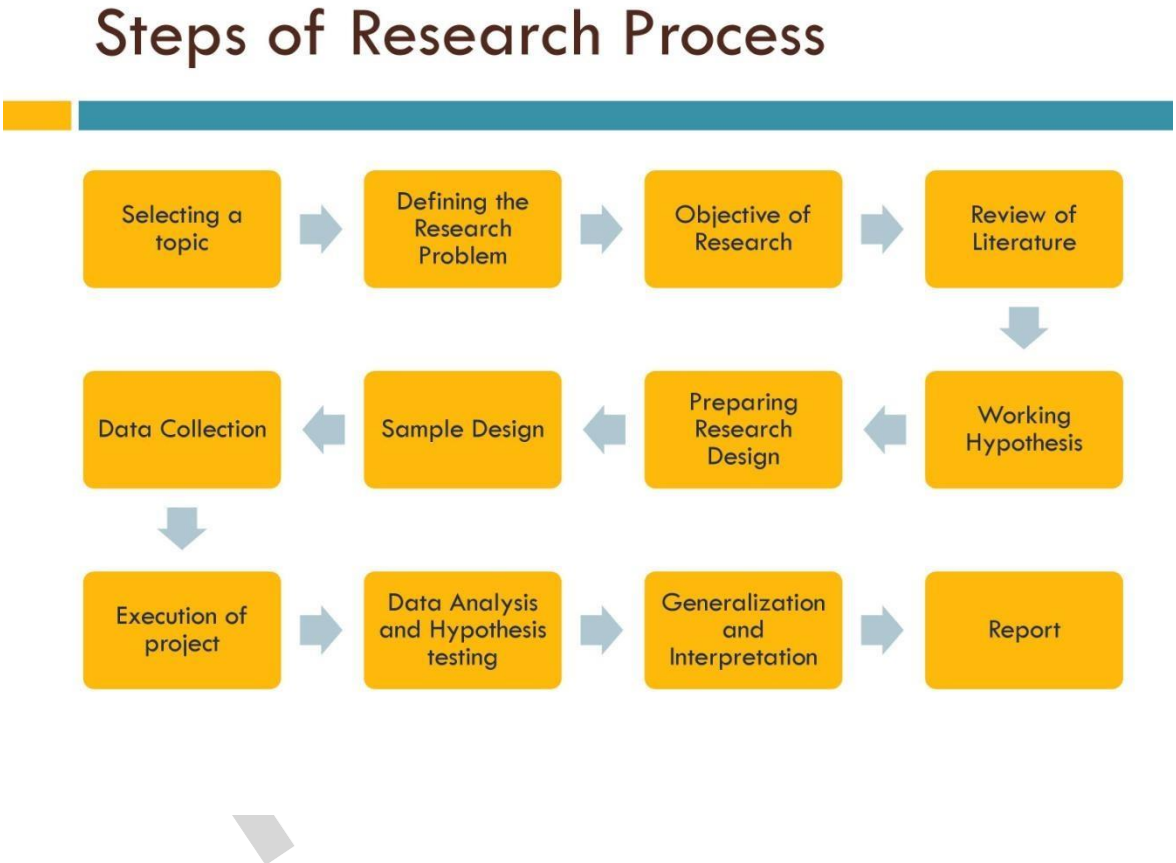
a. Quite similar to the descriptive research.
b. Identifies the causes of the problems and then solutions for these problems.
c. Related to causal relations.
d. It is directed by hypothesis.
e. Can be done only where knowledge is advanced.

Scope of Research


Scope of research
Environmental Level
1. Technological innovations.
2. Competitors Analysis.
3. Industry fears.
4. New Market entry.
5. New product development
Organizational Level
1. HRM.
2. Finance.
3. Production.
4. Organizational Effectiveness and Success.
Marketing Level
1. Product.
2. Price.
3. Place.
4. Promotion.
5. Sales.
6. Customer.


Characteristics of Good research
1. Originates with a question or problem.
2. Requires clear articulation of a goal.
3. Follows a specific plan or procedure.
4. Often divides main problem into sub problems.
5. Guided by specific problem, question, or hypothesis.
6. Accepts certain critical assumptions.
7. Requires collection and interpretation of data.
8. Cyclical (helical) in nature.

Research Process



1. Problem identification: The first step in a research process is to identify the problem or opportunity. The problem may be about decrease in sales, increase in competition, expansion of market, etc.

2. Problem definition: The second step in a research process is to define the problem. In this stage, the researcher must understand the problem correctly. He must find out the scope of the problem, the type of information needed, etc. If the problem is not defined properly, then it will result in waste of time, money and resources.

3. Research design: The third step in a research process is to prepare research design.

Research design is a plan for conducting a research. It guides the researcher in data collection. It gives proper direction to the research.

There are three types of research designs:

1. Exploratory research,
2. Descriptive Research and
3. Experimental Research.

All three types are used for marketing research.

4. Determining data needs: The fourth step in a research process is to determine the data needs. The researcher must consider the following issues:

1. Whether to use primary data or secondary data or both.
2. The accuracy and reliability of the data.
3. The availability of accurate and reliable data.
4. The cost and time required to collect the data.

5. Determining data sources: The fifth step in a research process is to determine the data sources. The researcher decides the sources of collecting data. The two main sources are secondary data and primary data. The researcher first collects secondary data. This is because it is easily available and less costly. It is collected by Desk Research. Desk Research can be internal for e.g. collected from company's records or external i.e. acquired from libraries, trade journals, government sources, etc. If the secondary data is not sufficient to solve the marketing problem, then primary data is wheeled.

Collecting primary data is very costly and time consuming. It can be collected by using survey methods, i.e. by doing personal interviews, telephone interviews and mail surveys. It can also be collected by using observational method and experimentation method.

So in this step the researcher decides what source and what method to use for collecting data.

6. Sampling design: The sixth step in a research process is of sampling design. The Researchers has limited time and other resources. So he cannot contact the total population. That is, he cannot collect information from all the people in the market. Therefore, he selects few persons from the population. These handful persons are called sample respondent. They are considered to represent the total population. The researcher collects data from the sample respondents. Sampling helps to save time, efforts and cost. It is used to collect primary data. The researcher has to decide about method of sampling, the size-of-sample, etc.

7. Designing questionnaire: The seventh step in a research process is of designing a questionnaire. In this stage, primary data is collected with the help of a questionnaire. So the researcher has to prepare a questionnaire. A questionnaire is a list of questions. These questions are asked to the respondents for collecting data. The questionnaire must be suitable so that the require data is collected easily, quickly and correctly. It is used for conducting person interview, telephone interviews and mail survey. The researcher must decide about the type of the information required, the type of questioned to be asked, the wordings of the questionnaire, its order, etc.

8. Field staff selection: The eighth step in a research process is of selecting field staff. After preparing the questionnaire, the researcher selects field interviewers. The field interviewers collect information from the respondents. They must be property trained. Students of psychology and statistics are good for this job.

9. Collection and processing of data: The ninth step in a research process is of collection and processing of data. In this stage, the data is collected from the respondents. The questionnaire is used for collecting data. In case of mail surveys, the questionnaire is sent to the respondents by post. In case of telephone interviews, the data is collected through telephone. In case of personal interviews, the data is collected by the field interviewers. The researcher can also use observation method and experimentation method for collecting data. The data collected must be reliable and complete. It must also be collected quickly. Secondary data is also collected. The data collected is raw. It cannot be used directly. It has to be processed and organized neatly. That is, the data must be edited, coded, classified and tabulated. Editing helps to remove the unwanted data. Coding, classification and tabulation make the data ready for analysis and interpretation.

10. Analysis and interpretation of data: The tenth step in a research process is of analysis and interpretation of data. In this stage, the researcher analyzes and interprets the data. That is, he studies the data very careful and draws conclusions from it. These conclusions are then used to solve the marketing problem.

11. Project reporting: The eleventh step in a research process is to prepare a project report.

In this stage, the researcher prepares the final research report. This report contains a title of the report, method used, findings, conclusions and suggestions about how to solve the marketing problem. The language of the report must not be very difficult. The report must be submitted to the marketing executives for recommendations and implementation.

12. Follow up: Finally, the last step in a research process is to do a follow up.

In this stage, the marketing executive makes changes in the product, price, marketing policies, etc. as per the recommendations of the report. Here, the researcher should find out, whether his recommendations are implemented properly or not. He should also figure-out, whether the marketing problem is solved or not.

CRITERIA OF GOOD RESEARCH

Good research is a **systematic, objective and scientific investigation** carried out to discover new facts or verify existing knowledge.

Clearly Defined Research Problem

- The research problem must be **clear, specific and well-formulated**.
- A vague topic leads to confusion in data collection and analysis.
- Clear objectives help guide the entire study in the right direction.

Systematic and Organized Process

- Research must follow a **step-by-step procedure** (problem → literature review → hypothesis → design → data → analysis → conclusion).
- Each step should logically follow the previous one.
- Proper organization increases accuracy and reliability.

Empirical (Evidence-Based)

- Research must be based on **real observations or measurable data**, not personal beliefs.
- Data can be collected through surveys, interviews, experiments, or observations.
- Conclusions must be supported by facts.

Objectivity

- The researcher must remain **neutral and unbiased**.
- Personal opinions should not influence findings.
- Results should be reported honestly, even if they contradict expectations.

Validity

- The research tool must measure **what it is intended to measure**.
- For example, an intelligence test should measure intelligence, not memory or language skill.

Reliability

- The results should be **consistent if the study is repeated** under similar conditions.
- Reliable tools produce stable and dependable results.

Logical and Accurate Analysis

- Data must be analyzed using appropriate statistical or qualitative techniques.
- Interpretation should be logical and directly related to findings.
- Avoid over-generalization.

Replicability

- The study should be explained clearly so that **other researchers can repeat it**.

- This increases trust and credibility of research.

Ethical Standards

- Participants must give **informed consent**.
- Confidentiality and privacy must be protected.
- Plagiarism and data manipulation must be avoided.

Contribution to Knowledge

- Good research should add something new to existing knowledge.
- It may solve practical problems or suggest areas for future research.

PROBLEMS FACED BY RESEARCHERS

Difficulty in Selecting Research Problem

- Choosing a suitable and researchable topic can be confusing.
- Some topics may be too broad, too narrow, or lack sufficient data.

Lack of Financial Resources

- Research requires money for travel, materials, data collection, and publication.
- Limited funding restricts the scope of research.

Time Constraints

- Researchers often have deadlines for submission.
- Balancing research with job or academic work becomes stressful.

Data Collection Problems

- Respondents may refuse to participate or give false information.
- Language barriers and accessibility issues may affect accuracy.

Sampling Issues

- Selecting a truly representative sample is difficult.
- Small or biased samples reduce validity of findings.

Methodological Challenges

- Choosing the correct research design and tools requires expertise.
- Lack of technical knowledge can affect quality.

Problems in Data Analysis

- Statistical analysis can be complex.
- Errors in coding or interpretation may lead to wrong conclusions.

Research Bias

- Personal beliefs of the researcher may influence interpretation.
- Confirmation bias may lead to focusing only on supportive data.

Ethical Difficulties

- Handling sensitive topics can be challenging.
- Ensuring confidentiality in small samples is difficult.

Publication Barriers

- Research papers may be rejected by journals.
- Peer review process can be lengthy and competitive.

Unit II

RESEARCH PROBLEM

Objectives of research problem

The objectives of a research problem refer to the clear and specific goals that a researcher intends to achieve through a study. They provide direction and focus to the entire research process and help in determining what data should be collected and how it should be analyzed. The main objectives of research include exploring a new phenomenon, describing the characteristics of a situation or group, determining the frequency of occurrence of an event, testing hypotheses, and establishing cause-and-effect relationships between variables. Research objectives also aim to provide practical solutions to problems and contribute new knowledge to the existing body of literature. Well-defined objectives are essential because they ensure that the study remains systematic, relevant, and meaningful.

Identification & formulation of Research Problem

The main steps in identification & formulation of research problem are:

1. Specify the Research Objectives
A clear statement of objectives will help you develop **effective research.**
It will help the decision makers evaluate your project. **It's critical** that you have manageable objectives. (Two or three clear goals will help to keep your research project focused and relevant.)
2. Review the Environment or Context of the Research Problem
As a marketing researcher, you must work closely with your team. This will help you determine whether the findings of your project will produce enough information to be worth the cost.
In order to do this, you have to identify the environmental variables that will affect the research project.
3. Explore the Nature of the Problem
Research problems range from simple to complex, depending on the number of variables and the nature of their relationship.
If you understand the nature of the **problem as a researcher**, you will be able to better develop a solution for the problem.
To help you understand all dimensions, you might want to consider focus groups of consumers, sales people, managers, or professionals to provide what is sometimes much needed insight.
4. Define the Variable Relationships
Marketing plans often focus on creating a sequence of behaviors that occur over time, as in the adoption of a new package design, or the introduction of a new product.
Such programs create a commitment to follow some behavioral pattern in the future.
Studying such a process involves:

- Determining which variables affect the solution to the problem.
- Determining the degree to which each variable can be controlled.
- Determining the functional relationships between the variables and which variables are critical to the solution of the problem.

During the **problem formulation** stage, you will want to generate and consider as many courses of action and variable relationships as possible.
5. The Consequences of Alternative Courses of Action
There are always consequences to any course of action. Anticipating and communicating the

possible outcomes of various courses of action is a primary responsibility in the research process.

TECHNIQUES OF DEFINING A RESEARCH PROBLEM

Extensive Literature Survey

- Study books, journals, research papers, reports, and previous studies.
- Helps understand what has already been researched.
- Identifies gaps in existing knowledge.
- Prevents duplication of research.

Experience Survey (Discussion with Experts)

- Consult experienced persons, professionals, or subject experts.
- Helps gain practical insights into the problem.
- Clarifies doubts and refines the research idea.

Analysis of Existing Data

- Examine available data, records, and statistics.
- Helps identify trends, issues, and researchable areas.
- Useful in narrowing down the problem.

Pilot Study

- Conduct a small preliminary study before the main research.
- Tests feasibility of the research problem.
- Helps in refining objectives and research design.

Brainstorming

- Generate ideas through discussions or self-thinking.
- Encourages creative thinking.
- Helps in selecting a focused and meaningful problem.

Rephrasing the Research Problem

- State the problem in clear, simple, and specific terms.
- Define key concepts and variables.
- Convert the general problem into operational terms.

EXPERIMENTAL RESEARCH DESIGN

Experimental research design is a scientific method of research used to study the **cause-and-effect relationship** between two or more variables. In this design, the researcher deliberately manipulates one variable and observes its effect on another variable under controlled conditions.

Purpose of Experimental Research:

- The main purpose of experimental research is to determine whether a change in one variable produces a change in another variable.
- It helps in establishing a clear cause-and-effect relationship.
- It is mainly used in science, psychology, education, and medical research.

Key Elements of Experimental Research Design

Independent Variable (IV)

- The independent variable is the variable that is deliberately changed or manipulated by the researcher.
- It is considered the "cause" in the experiment.
- The researcher controls this variable to observe its impact on the dependent variable.

Dependent Variable (DV)

- The dependent variable is the variable that is measured or observed in the experiment.
- It represents the "effect" of the independent variable.
- Any change in this variable is assumed to be caused by manipulation of the independent variable.

Experimental Group

- The experimental group is the group that receives the treatment or experimental condition.
- This group is exposed to the independent variable.
- The results of this group are compared with the control group.

Control Group

- The control group does not receive the experimental treatment.
- It is used as a standard for comparison.
- This helps the researcher determine whether the treatment actually caused any change.

Control of Extraneous Variables

- Extraneous variables are other factors that may influence the outcome of the experiment.
- The researcher must control these variables to avoid confusion in results.
- Proper control increases the validity and accuracy of the study.

Types of Experimental Research Design

Pre-Experimental Design

- This is the simplest form of experimental design.
- It may not include a control group or random assignment.
- It is less reliable because external factors may affect results.

True Experimental Design

- This design includes random assignment of participants and a control group.
- It provides a high level of control over variables.
- It is considered the most accurate method to establish cause-and-effect relationships.

Quasi-Experimental Design

- This design resembles true experimental design but lacks random assignment.
- It is commonly used in social sciences where randomization is not possible.
- It is more practical but slightly less reliable than true experimental design.

Unit III

Introduction

Review of literature is a collective body of works done by earlier scientists and published in the form of books or in the form of articles in journals or published as monograph etc. Every scientific investigation starts with a Review of Literature. In fact, working with the literature is an essential part of the research process which help generate ideas, helps in developing significant questions and is regarded as instrumental in the process of research design. In this unit we will be dealing with the review of literature, how to go about it, what is its importance and how the review should be organised and how to relate it to the present research report.

Purpose

A literature review is part of a report. It provides considerable information on the topic being researched and the various works that had gone on in the field over the years. These materials are gathered by the researcher from many sources such as journals, books, documents etc. The review of such a literature could be a matter of fact presentation of the information or it could be a synthesis of a large number of information and put together subject wise for the purpose of understanding. It can be just a simple summary of the sources, but it usually has an organizational pattern and combines both summary and synthesis. In summary all the information is synthesized and given in a capsule form. It synthesizes and organizes the entire

information in terms of its relevance and appropriateness to the topic of research. It might give a new interpretation of old material or combine new with old interpretations. Or it might trace the intellectual progression of the field, including major debates. And depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent and relevant information.

Developing a theoretical framework

A theoretical framework is a foundational review of existing theories that serves as a roadmap for developing the arguments you will use in your own work.

Theories are developed by researchers to explain phenomena, draw connections, and make predictions. In a theoretical framework, you explain the existing theories that support your research, showing that your paper or dissertation topic is relevant and grounded in established ideas.

In other words, your theoretical framework justifies and contextualizes your later research, and it's a crucial first step for your research paper, thesis, or dissertation. A well-rounded theoretical framework sets you up for success later on in your research and writing process.

To state the assumptions and orientations of the researcher regarding the topic of study.

A theoretical framework illuminates the phenomenon of study and the corresponding assumptions adopted by the researcher. Frameworks can take on different orientations. A theoretical framework drives the question, guides the types of methods for data collection and analysis, informs the discussion of the findings, and reveals the subjectivities of the researcher. A theoretical framework does not rationalize the need for the study, and a theoretical framework can come from different fields.

To create your own theoretical framework, you can follow these three steps:

- Identifying your key concepts
- Evaluating and explaining relevant theories
- Showing how your research fits into existing research

1. Identify your key concepts

The first step is to pick out the key terms from your problem statement and research questions. Concepts often have multiple definitions, so your theoretical framework should also clearly define what you mean by each term.

2. Evaluate and explain relevant theories

By conducting a thorough literature review, you can determine how other researchers have defined these key concepts and drawn connections between them. As you write your theoretical framework, your aim is to compare and critically evaluate the approaches that different authors have taken.

After discussing different models and theories, you can establish the definitions that best fit your research and justify why. You can even combine theories from different fields to build your own unique framework if this better suits your topic.

Make sure to at least briefly mention each of the most important theories related to your key concepts. If there is a well-established theory that you don't want to apply to your own research, explain why it isn't suitable for your purposes.

3. Show how your research fits into existing research

Apart from summarizing and discussing existing theories, your theoretical framework should show how your project will make use of these ideas and take them a step further.

You might aim to do one or more of the following:

- Test whether a theory holds in a specific, previously unexamined context
- Use an existing theory as a basis for interpreting your results
- Critique or challenge a theory
- Combine different theories in a new or unique way

Developing a conceptual framework

A conceptual framework illustrates the expected relationship between your variables. It defines the relevant objectives for your research process and maps out how they come together to draw coherent conclusions. To describe the researcher's understanding of the main concepts under investigation.

A conceptual framework is a representation of the relationship you expect to see between your variables, or the characteristics or properties that you want to study.

Conceptual frameworks can be written or visual and are generally developed based on a literature review of existing studies about your topic.

Step 1: Choose your research question

Your research question guides your work by determining exactly what you want to find out, giving your research process a clear focus.

However, before you start collecting your data, consider constructing a conceptual framework. This will help you map out which variables you will measure and how you expect them to relate to one another.

Step 2: Select your independent and dependent variables

In order to move forward with your research question and test a cause-and-effect relationship, you must first identify at least two key variables: your independent and dependent variables.

Note that causal relationships often involve several independent variables that affect the dependent variable. For the purpose of this example, we'll work with just one independent variable ("hours of study").

Step 3: Visualize your cause-and-effect relationship

Now that you've figured out your research question and variables, the first step in designing your conceptual framework is visualizing your expected cause-and-effect relationship.

We demonstrate this using basic design components of boxes and arrows. Here, each variable appears in a box. To indicate a causal relationship, each arrow should start from the independent variable (the cause) and point to the dependent variable (the effect).

Step 4: Identify other influencing variables

It's crucial to identify other variables that can influence the relationship between your independent and dependent variables early in your research process.

Some common variables to include are moderating, mediating, and control variables.

Moderating variables

Moderating variable (or moderators) alter the effect that an independent variable has on a dependent variable. In other words, moderators change the "effect" component of the cause-and-effect relationship.

Mediating variables

Now we'll expand the framework by adding a mediating variable. Mediating variables link the independent and dependent variables, allowing the relationship between them to be better explained.

The conceptual framework is created by the researcher(s), includes the presumed relationships among concepts, and addresses needed areas of study discovered in literature reviews. The conceptual framework is informed by literature reviews, experiences, or experiments. It may include emergent ideas that are not yet grounded in the literature. It should be coherent with the paper's theoretical framing. A conceptual framework articulates the phenomenon under study through written descriptions and/or visual representations.

Steps in Writing Literature Review

- First, collect relevant sources such as books, research articles, theses, reports, and credible online materials related to the topic.
- Carefully read and analyze the collected materials to understand key findings and concepts.
- Organize the information according to themes, concepts, or chronological order.

- Compare and contrast the views of different authors.
- Identify research gaps and highlight how the present study will fill those gaps.
- Write the review in a logical and coherent manner.

How to Present Literature Review

- Begin with a general overview of the topic.
- Summarize important studies related to the research problem.
- Mention the name of the author and year of study (if required).
- Explain the methodology and findings of previous studies briefly.
- Show the relevance of past studies to the current research.
- End the review by pointing out the research gap.

Unit IV

RESEARCH DESIGN

A **research design** is a systematic plan to study a scientific problem. The design of a study defines the study type (descriptive, correlational, semi-experimental, experimental, review, meta-analytic) and sub-type (e.g., descriptive-longitudinal case study), research question, hypotheses, independent and dependent variables, experimental design, and, if applicable, data collection methods and a statistical analysis plan. Research design is the framework that has been created to seek answers to research questions.

Confirmatory versus exploratory research

Confirmatory research tests *a priori* hypotheses—outcome predictions that are made before the measurement phase begins. Such *a priori* hypotheses are usually derived from a theory or the results of previous studies. The advantage of confirmatory research is that the result is more meaningful, in the sense that it is much harder to claim that a certain result is statistically significant. The reason for this is that in confirmatory research, one ideally strives to reduce the probability of falsely reporting a non-significant result as significant. This probability is known as α-level or a type I error. Loosely speaking, if you know what you are looking for, you should be very confident when and where you will find it; accordingly, you only accept a result as significant if it is highly unlikely to have been observed by chance.

Exploratory research on the other hand seeks to generate *a posteriori* hypotheses by examining a data-set and looking for potential relations between variables. It is also possible to have an idea about a relation between variables but to lack knowledge of the direction and strength of the relation. If the researcher does not have any specific hypotheses beforehand, the study is exploratory with respect to the variables in question (although it might be confirmatory for others). The advantage of exploratory research is that it is easier to make new discoveries due to the less stringent methodological restrictions. Here, the researcher does not want to miss a potentially interesting relation and therefore aims to minimize the probability of rejecting a *real* effect or relation, this probability is sometimes referred to as β and the associated error is of type II. In other words, if you want to see whether some of your measured variables could be related, you would want to increase your chances of finding a significant result by lowering the threshold of what you deem to be *significant*.

Sometimes, a researcher may conduct exploratory research but report it as if it had been confirmatory this is a questionable research practice bordering fraud.

Need and Importance of Research Design

Research design carries an important influence on the reliability of the results attained. It therefore provides a solid base for the whole research. It is needed due to the fact that it allows for the smooth working of the many research operations. This makes the research as effective as possible by providing maximum information with minimum spending of effort, money and time. For building

of a car, we must have a suitable blueprint made by an expert designer. In a similar fashion, we require a suitable design or plan just before data collection and analysis of the research project. Planning of design must be carried out cautiously as even a small mistake might mess up the purpose of the entire project. The design helps the investigator to organize his ideas, which helps to recognize and fix his faults, if any. In a **good research design**, all the components go together with each other in a coherent way. The theoretical and conceptual framework must with the research goals and purposes. In the same way, the data gathering method must fit with the research purposes, conceptual and theoretical framework and method of data analysis.

A research design is like a successful journey:

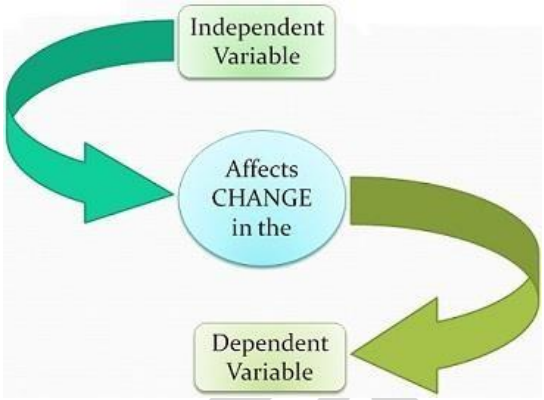
- Broadens your mind
- Provides fascinating & exciting experience
- Gives understanding of world around you
- Provides chance to meet people
- Gives fun and reward, but sometimes, very tedious & monotonous too.

The **importance of research design in research methodology** is due to the following:

- It may result in the preferred kind of study with helpful conclusion.
- It cuts down on inaccuracy.
- Allows you get optimum efficiency and reliability.
- Reduce wastage of time.
- Reduce uncertainty, confusion and practical haphazard related to any research problem.
- Of great help for collection of research material and testing of hypothesis.
- It is a guide for giving research the right path.
- Gets rid of bias and marginal errors.
- Provides an idea concerning the type of resources needed in terms of money, effort, time, and manpower.
- Smooth & efficient sailing (sets boundaries & helps prevent blind search)
- Maximizes reliability of results.
- Provides firm foundation to the endeavor.
- Averts misleading conclusions & thoughtless useless exercise.
- Provides opportunity to anticipate flaws & inadequacies (anticipates problems).
- Incorporates by learning from other people's critical comments & evaluations.

Variables & Types of Variables

When it comes to experiments and data analysis, there are two main types of variables: **dependent variables** and **independent variables**. It's easy to get these mixed up, but the difference between dependent and independent variables is simple. Here is a quick and easy definition of each one, along with some examples.



1) Dependent Variable: This is the output variable you are really interested in monitoring to see if it was affected or not. It can also be called the "measured variable," the "responding variable," the "explained variable," etc. I think it is easy to remember this one because it is *dependent* on the other variables.

2) Independent Variables: These are the individual variables that you believe may have an effect on the dependent variable. They are sometimes called "explanatory variables," "manipulated variables," or "controlled variables."

Independent variable (IV)	Dependent variable (DV)
Also called: • Exposure variable • Control variable • Explanatory variable • Manipulated variable	Also called: • Outcome variable • Controlled variable • Explained variable • Response variable

Example #1: Golf Balls

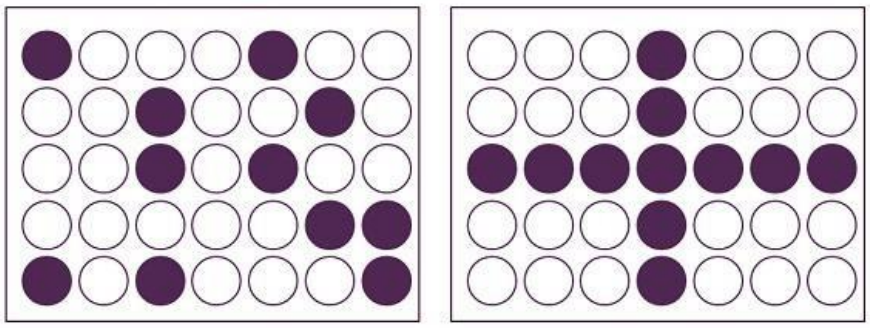
Here's a simple situation: Suppose you want to test golf ball flight distances, so you set up a simple experiment in which various golf balls are placed into a mechanical chute and fired into the air. The variable you really care about, the "output" or ***dependent variable*** is golf ball distance. ***Independent variables*** are the variables you are going to test to see how they affect distance. In this case, they are going to be things like air temperature, golf ball brand, and color of the golf ball. In the end, if you do a fancy regression analysis on all your data, you are going to end up with a formula that looks something like this: golf ball distance = 50 feet + air temperature factor + golf ball brand factor + golf ball color factor. See how all the independent variables (air temp, brand, color) have an effect on the dependent variable (distance)?

Example #2: Ice Cubes

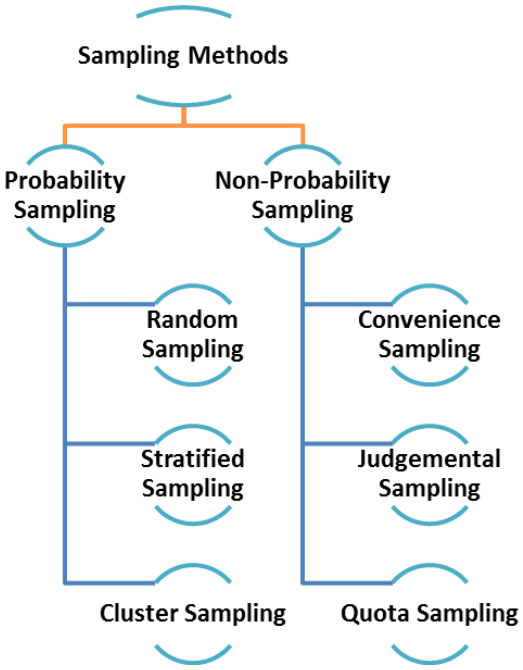
Here's another simple example: Imagine that you have a bunch of ice cubes and you want to test how long it takes them to melt in various situations. You have an experiment with 1,000 equally shaped ice cubes. Some of them are made of frozen cranberry juice and some of them are frozen lemonade. You are going to set some of them on a metal sheet and others are going to be placed on a wooden plank. Air temperature, wind, and every other condition you can think of will remain constant. So, in this case, your ***dependent variable*** is ice cube melting time. Your two ***independent variables*** are: juice type (cranberry or lemonade) and melting surface (metal or wood). I'm not sure why anyone would care to do such an experiment, but hopefully the difference between the dependent and independent variables are clear now.

METHODS OF SAMPLING

Samples can be either probability or non-probability samples. With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Probability samples are those based on simple random sampling, systematic sampling, stratified sampling, cluster/area sampling whereas non-probability samples are those based on convenience sampling, judgement sampling and quota sampling techniques.



Probability Sampling Vs Non-Probability Sampling



Deliberate sampling:

Deliberate sampling is also known as purposive or non-probability sampling. This sampling method involves purposive or deliberate selection of particular units of the universe for constituting a sample which represents the universe. When population elements are selected for inclusion in the sample based on the ease of access, it can be called convenience sampling. If a researcher wishes to secure data from, say, gasoline buyers, he may select a fixed number of petrol stations and may conduct interviews at these stations. This would be an example of convenience sample of gasoline buyers. At times such a procedure may give very biased results particularly when the population is not homogeneous. On the other hand, in judgement sampling, the researcher's judgement is used for selecting items which he considers as representative of the population. For example, a judgement sample of college students might be taken to secure reactions to a new method of teaching. Judgement sampling is used quite frequently in qualitative research where the desire happens to be to develop hypotheses rather than to generalise to larger populations.

Simple random sampling:

This type of sampling is also known as chance sampling or probability sampling where each and every item in the population has an equal chance of inclusion in the sample and each one of the possible samples, in case of finite universe, has the same probability of being selected. For example, if we have to select a sample of 300items from a universe of 15,000 items, then we can put the names or numbers of all the15,000 items on slips of paper and conduct a lottery. Using the random number tables is another method of random sampling. To select the sample, each item is assigned a number from 1 to 15,000. Then, 300 five digits random numbers are selected from the table. To do this we select some random starting point and then a systematic pattern is used in proceeding through the table. We might start in the 4th row, second column and proceed down the column to the bottom of the table and then move to the top of the next column to the right. When a number exceeds the limit of the numbers in the frame, in our case over 15,000, it is simply

passed over and the next number selected that does fall within the relevant range. Since the numbers were placed in the table in a completely random fashion, the resulting sample is random. This procedure gives each item an equal probability of being selected. Incase of infinite population, the selection of each item in a random sample is controlled by the same probability and that successive selections are independent of one another.

Systematic sampling:

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

Stratified sampling:

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

Quota sampling:

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

Cluster sampling and area sampling:

Cluster sampling involves grouping the population and then selecting the groups or the clusters rather than individual elements for inclusion in the sample. Suppose some departmental store wishes to sample its credit card holders. It has issued its cards to 15,000 customers. The sample size is to be kept say 450. For cluster sampling this list of 15,000 card holders could be formed into 100 clusters of 150 cardholders each. Three clusters might then be selected for the sample randomly. The sample size must often be larger than the simple random sample to ensure the

same level of accuracy because is cluster sampling procedural potential for order bias and other sources of error are usually accentuated. The clustering approach can, however, make the sampling procedure relatively easier and increase the efficiency of field work, specially in the case of personal interviews.

Area sampling

It is quite close to cluster sampling and is often talked about when the total geographical area of interest happens to be big one. Under area sampling we first divide thetotalareaintoanumberof smallernon-overlappingareas,generallycalledgeographicalclusters, then a number of these smaller areas are randomly selected, and all units in these small areas are included in the sample. Area sampling is specially helpful where we do not havethelistofthepopulationconcerned.Italsomakesthefield interviewingmoreefficient since interviewer can do many interviews at each location.

Multi-stage sampling:

This is a further development of the idea of cluster sampling. Thistechniqueismeantforbiginquiries extendingtoaconsiderablylargegeographicalarea entire country. Under multi-stage sampling the first stage may be to select large primary sampling units such as states, then districts, then towns and finally certain families within towns. If the technique of random-sampling is applied at all stages, the sampling procedures described as multi-stage random sampling.

Sequential sampling:

This is somewhat a complex sample design where the ultimate sizeofthesampleisnotfixedinadvance butisdeterminedaccordingtomathematicaldecisionson the basis of information yielded as survey progresses. This design is usually adopted under acceptance sampling plan in the context of statistical quality control.

STEPS IN SAMPLE DESIGN

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

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.

Sampling unit:

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

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.

Size of sample:

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

Parameters of interest:

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

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.

Sampling procedure:

Finally, the researcher must decide the type of sample he will use i.e., he must decide about the the items for the sample. Infact, this technique or procedure stands for the sample design technique to be used in selecting 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 which, for a given sample size and for a given cost, has a smaller sampling error.

CRITERIA OF SELECTING A SAMPLING PROCEDURE

In this context one must remember that two costs are involved in a sampling analysis viz., the cost of collecting the data and the cost of an incorrect inference resulting from the data. Researcher must keep in view the two causes of incorrect inferences viz., systematic bias and sampling error. A

systematic bias results from errors in the sampling procedures, and it cannot be reduced or eliminated by increasing the sample size. At best the causes responsible for these errors can be detected and corrected. Usually a systematic bias is the result of one or more of the following factors:

1. Inappropriate sampling frame:

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

2. Defective measuring device:

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

3. Non-respondents:

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

4. Indeterminancy principle:

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

5. Natural bias in the reporting of data:

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

CHARACTERISTICS OF A GOOD SAMPLE DESIGN

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

(a) Sample design must result in a truly representative sample.

(b) Sample design must be such which results in a small sampling error.

(c) Sample design must be viable in the context of funds available for the research study.

(d) Sample design must be such so that systematic bias can be controlled in a better way.

(e) Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

Comparison Table

Feature	Sampling Error	Non-Sampling Error
Cause	Due to studying a sample instead of the population	Due to mistakes or biases in data collection, processing, or measurement
Occurrence	Only in sample surveys	Can occur in both sample surveys and censuses
Reduction	Increase sample size, proper random sampling	Careful design, training, supervision, data checking
Nature	Random	Systematic or random
Example	Sample mean differs from population mean	Wrong answers, data entry errors, non-response

---

Conclusion

- **Sampling error** is the difference caused by observing only a sample instead of the entire population, while **non-sampling error** is caused by mistakes or biases in research procedures.
- Both types of errors affect the accuracy of research, but non-sampling errors are usually harder to detect and control.

Unit V

Hypothesis, Types & Formulation of Hypothesis

Introduction and Definition

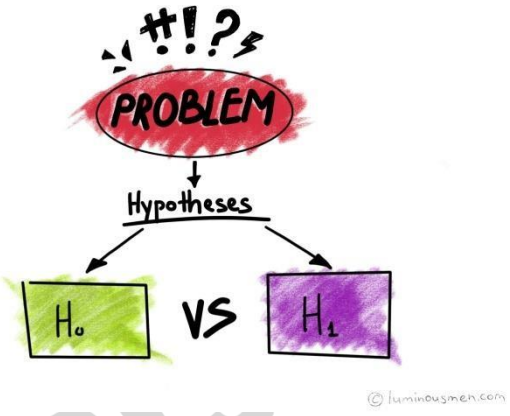
Hypothesis and the theories are generally responsible for the movement of knowledge from the unknown to the known. Hypotheses play a very important and a critical role in the assertion of a particular thing, as they are able to describe certain facts and are also able to explain the various relationships between these facts. As a result of this, hypotheses help a great deal in the investigation operations or activities.

On the institution of the problem to be answered in the process of the research, the researcher forms various tentative or possible solutions to these problems these proposed answers or the solutions are referred to as the hypothesis. But a very critical and essential point to be kept in mind here is that these propositions are not at all verified in nature.

So Hypothesis can be referred to as the interpretation of certain facts which is just a possible solution or a tentative answer to a problem and is completely or partly unverified in nature. Then afterwards on its establishment, it ceases to be a hypothesis and then finally becomes a theory or a principle. The word 'Hypothesis' has come from the Greek word hypo (means under)and tithenas (means to place) together these words indicate towards the support they provide to each other on the placement of the hypothesis under the evidence, which acts as a foundation.

According to George A Luniberg, hypothesis can be defined as a 'tentative generalization, the validity of which remains to be tested. In this elementary stage, the hypothesis may be very hunch, guess, imaginative data, which becomes the basis for an action or an investigation.'

A very vital point that should be kept in mind about the hypotheses is that these are not theories these only have some linkage to the theory but hypothesis is not that much elaborated as the theory is. But it can be said that the hypothesis is derived from the theory.



Role and Functions of the hypothesis

1. Helps in the testing of the theories.
2. Serves as a great platform in the investigation activities.
3. Provides guidance to the research work or study.
4. Hypothesis sometimes suggests theories.
5. Helps in knowing the needs of the data.

6. Explains social phenomena.
7. Develops the theory.
8. Also acts as a bridge between the theory and the investigation.
9. Provides a relationship between phenomena in such a way that it leads to the empirical testing of the relationship.
10. Helps in knowing the most suitable technique of analysis.
11. Helps in the determination of the most suitable type of research.
12. Provides knowledge about the required sources of data.
13. Research becomes focused under the direction of the hypothesis.
14. Is very helpful in carrying out an enquiry of a certain activity.
15. Helps in reaching conclusions, if it is correctly drawn.

Sources of hypothesis

1. Observations made in routine activities.
2. Theories based on the scientific approach.
3. Analogies.
4. Knowledge obtained from the functional executives.
5. Results of the research and development department.
6. Experience of the investigator.

Characteristics of hypothesis

1. Should be very specific in nature.
2. Concept of the hypothesis should be clear.
3. Should be empirically testable.
4. Should be related to the devices and the techniques that are available.
5. Should relate to the body of the theory.
6. Should recognize the specific variables and their relation

Problems faced during hypothesis formulation

Formulating a hypothesis is not at all an easy process and is faced with a large number of difficulties. According to Goode and Hatt, the various difficulties faced during the formulation of the hypothesis generally include the lack of the knowledge about the scientific approach of the method involved, as sometimes it becomes impossible to gather the complete information about a particular scientific method. One other major difficulty in the formulation of the hypothesis is the lack of clear theoretical background. Because of this problem of unclear and indefinite background of theory one is not able to arrive to a conclusion easily.

But with time answers to all such problems are available and these difficulties that arise during the hypothesis formulation can be easily removed by having complete and accurate information about the concepts of the subjects involved. Also the hypothesis should not be very long and should be timely in nature.

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

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data are actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city's 200 drugstores in a certain way constitutes a sample design.

Hypothesis Testing

Hypothesis can be referred to as the interpretation of certain facts which is just a possible solution or a tentative answer to a problem and is completely or partly unverified in nature. Then afterwards on its establishment, it ceases to be a hypothesis and then finally becomes a theory or a principle. The word 'Hypothesis' has come from the Greek word hypo (means under)and tithenas (means to place) together these words indicate towards the support they provide to each other on the placement of the hypothesis under the evidence, which acts as a foundation.

Step 1: State the Null Hypothesis.

The null hypothesis can be thought of as the opposite of the "guess" the research made (in this example the biologist thinks the plant height will be different for the fertilizers). So the null would be that there will be no difference among the groups of plants. Specifically in more statistical language the null for an ANOVA is that the means are the same. We state the Null hypothesis as:

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

for *k* levels of an experimental treatment.

Step 2: State the Alternative Hypothesis.

H_1 : treatment level means not all equal

The reason we state the alternative hypothesis this way is that if the Null is rejected, there are many possibilities.For example, $\mu_1 \neq \mu_2 = \cdots = \mu_k$ is one possibility, as is $\mu 1 = \mu 2 \neq \mu 3 = \cdots = \mu k$. Many

people make the mistake of stating the Alternative Hypothesis as: $\mu 1 \neq \mu 2 \neq \cdots \neq \mu k$ which says that every mean differs from every other mean. This is a possibility, but only one of many possibilities. To cover all alternative outcomes, we resort to a verbal statement of 'not all equal' and then follow up with mean comparisons to find out where differences among means exist. In our example, this means that fertilizer 1 may result in plants that are really tall, but fertilizers 2, 3 and the plants with no fertilizers don't differ from one another. A simpler way of thinking about this is that at least one mean is different from all others.

Step 3: Set α (Significance level)

If we look at what can happen in a hypothesis test, we can construct the following contingency table:

	In Reality	
Decision	H_0 is TRUE	H_0 is FALSE
Accept H_0	OK	Type II Error β = probability of Type II Error
Reject H_0	Type I Error α = probability of Type I Error	OK

You should be familiar with type I and type II errors from your introductory course. It is important to note that we want to set α before the experiment (*a-priori*) because the Type I error is the more 'grevious' error to make. The typical value of α is 0.05, establishing a 95% confidence level. **For this course we will assume α =0.05.**

Step 4: Collect Data

Remember the importance of recognizing whether data is collected through an experimental design or observational.

Step 5: Calculate a test statistic.

For categorical treatment level means, we use an *F* statistic, named after R.A. Fisher. We will explore the mechanics of computing the *F* statistic beginning in Lesson 2. The *F* value we get from the data is labeled $F_{calculated}$.

Step 6: Construct Acceptance / Rejection regions.

As with all other test statistics, a threshold (critical) value of *F* is established. This *F* value can be obtained from statistical tables, and is referred to as $F_{critical}$ or F α . As a reminder, this critical value is the minimum value for the test statistic (in this case the *F* test) for us to be able to reject the null.

The *F* distribution, F α , and the location of Acceptance / Rejection regions are shown in the graph below:

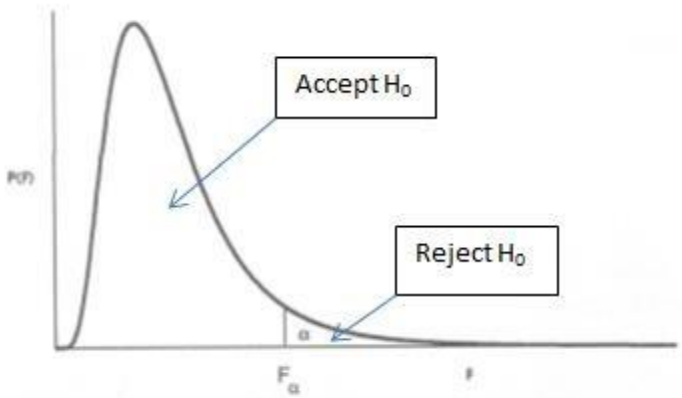


Figure K.1: The F distribution

Step 7: Based on steps 5 and 6, draw a conclusion about H_0.

If the $F_{calculated}$ from the data is larger than the Fα, then you are in the Rejection region and you can reject the Null Hypothesis with (1-α) level of confidence.

Note that modern statistical software condenses step 6 and 7 by providing a *p*-value. The *p*-value here is the probability of getting an $F_{calculated}$ even greater than what you observe. If by chance, the $F_{calculated}$ = F α , then the *p*-value would exactly equal to α. With larger $F_{calculated}$ values, we move further into the rejection region and the *p*-value becomes less than α. So the decision rule is as follows:

If the *p*-value obtained from the ANOVA is less than α, then Reject H_0 and Accept H_A.

Errors In Hypothesis Testing

Type I Error (False Positive Error)

1) A type I error occurs when the null hypothesis is true, but is rejected. Let me say this again, a *type I error occurs when the null hypothesis is actually **true**, but was rejected as **false** by the testing*.

2) A type I error, or false positive, is asserting something as true when it is actually false. This false positive error is basically a "false alarm" – a result that indicates a given condition has been fulfilled when it actually has not been fulfilled (i.e., erroneously a positive result has been assumed).

Type II Error (False Negative)

1) A type II error occurs when the null hypothesis is false, but erroneously fails to be rejected. Let me say this again, a *type II error occurs when the null hypothesis is actually **false**, but was accepted as **true** by the testing.*

2) A type II error, or false negative, is where a test result indicates that a condition failed, while it actually was successful. A Type II error is committed when we fail to believe a true condition.

A tabular relationship between truthfulness/falseness of the null hypothesis and outcomes of the test can be seen in the table below:

	Null Hypothesis is true	Null hypothesis is false
Reject null hypothesis	Type I Error**False Positive**	**Correct Outcome**True Positive
Fail to reject null hypothesis	**Correct outcome**True Negative	Type II Error**False Negative**

Let's look at some business related examples. In these examples I have reworded the null hypothesis, so be careful on the cost assessment.

Null Hypothesis	Type I Error / False Positive	Type II Error / False Negative
Medicine A cures Disease B	(H_0 **true**, but rejected as **false**)Medicine A **cures** Disease B, but is **rejected as false**	(H_0 **false**, but accepted as **true**)Medicine A **does not cure** Disease B, but is **accepted as true**
Cost Assessment	Lost opportunity cost for rejecting an effective drug that could cure Disease B	Unexpected side effects (maybe even death) for using a drug that is not effective

Let's try one more.

Null Hypothesis	Type I Error / False Positive	Type II Error / False Negative
Display Ad A is effective in driving c onversions	(H_0 **true**, but rejected as **false**)Display Ad A **is effective** in driving conversions, but is **rejected as false**	(H_0 **false**, but accepted as **true**)Display Ad A is **not effective** in driving conversions, but is **accepted as true**
Cost Assessment	Lost opportunity cost for rejecting an effective Display	Lost sales for promoting an ineffective Display Ad A to

The cost ramifications in the medicine example are quite substantial, so additional testing would likely be justified in order to minimize the impact of the type II error (using an ineffective drug) in our example. However, the cost ramifications in the Display Ad example are quite small, for both the type I and type II errors, so additional investment in addressing the type I and type II errors is probably not worthwhile

Parametric & Non Parametric Tests

1) **Parametric Test:** If the information about the population is completely known by means of its parameters then statistical test is called parametric test∗ Eg: t- test, f-test, z-test, ANOVA are Parametric Tests.
2) **Non parametric test:** If there is no knowledge about the population or parameters, but still it is required to test the hypothesis of the population. Then it is called non-parametric test∗ E.g.: Mann-Whitney, rank sum test, Kruskal-Wallis test.

Parametric Tests	Nonparametric Tests
Independent-Samples T Test	Mann-Whitney Test
Paired-Samples T Test	Wilcoxon Signed-Rank Test
One-way ANOVA	Kruskal-Wallis Test
One-way Repeated Measures ANOVA	Friedman's ANOVA

t-TEST

- The t test tells you how significant the differences between groups are; In other words it lets you know if those differences (measured in means/averages) could have happened by chance.
- **A very simple example:** Let's say you have a cold and you try a naturopathic remedy. Your cold lasts a couple of days. The next time you have a cold, you buy an over-the-counter pharmaceutical and the cold lasts a week. You survey your friends and they all tell you that their colds were of a shorter duration (an average of 3 days) when they took the homeopathic remedy. What you *really* want to know is, are these results repeatable? A t test can tell you by comparing the means of the two groups and letting you know the probability of those results happening by chance.

There are **three main types of t-test:**

- An Independent Samples t-test compares the means for two groups.
- A Paired sample t-test compares means from the same group at different times (say, one year apart).
- A One sample t-test tests the mean of a single group against a known mean.

F-TEST

An "F Test" is a catch-all term for any test that uses the F-distribution. In most cases, when people talk about the F-Test, what they are actually talking about is The *F-Test to Compare Two Variances.*

Analysis of Variance (ANOVA)

Purpose

1) The reason for doing an ANOVA is to see if there is any difference between groups on some variable.

2) For example, you might have data on student performance in non-assessed tutorial exercises as well as their final grading. You are interested in seeing if tutorial performance is related to final grade. ANOVA allows you to break up the group according to the grade and then see if performance is different across these grades.

ANOVA is available for both parametric (score data) and non-parametric (ranking/ordering) data.

Types of ANOVA

One-way between groups

The example given above is called a **one-way between groups model.**

You are looking at the differences between the groups.

There is only one grouping (final grade) which you are using to define the groups.

This is the simplest version of ANOVA.

This type of ANOVA can also be used to compare variables between different groups - tutorial performance from different intakes.

One-way repeated measures

A one way repeated measures ANOVA is used when you have a single group on which you have measured something a few times.

For example, you may have a test of understanding of Classes. You give this test at the beginning of the topic, at the end of the topic and then at the end of the subject.

You would use a one-way repeated measures ANOVA to see if student performance on the test changed over time.

Two-way between groups

A two-way between groups ANOVA is used to look at complex groupings.

For example, the grades by tutorial analysis could be extended to see if overseas students performed differently to local students. What you would have from this form of ANOVA is:

The effect of final grade

The effect of overseas versus local

The interaction between final grade and overseas/local

Each of the **main effects** are one-way tests. The **interaction effect** is simply asking "is there any significant difference in performance when you take final grade and overseas/local acting together".

Comparison Table

Feature	Univariate Analysis	Bivariate Analysis
Number of variables	One	Two
Purpose	Describe and summarize data	Study relationship or association
Techniques	Mean, Median, Mode, SD, Histograms	Correlation, Regression, Scatter Plots
Focus	Distribution of single variable	Relationship between variables
Example	Average income of a group	Relationship between income and education

Conclusion

- **Univariate analysis** focuses on describing a single variable, while **bivariate analysis** examines the relationship between two variables.
- Both are essential steps in research to summarize data and draw meaningful conclusions.