BUSINESS RESEARCH METHODS

IV - SEMESTER BBA
Bangalore University

Prepared by,
Dr.S.HARIHARAPUTHRIAN
Prof.Management,
New Horizon College,
Kasturi Nagar,
Bangalore -43.

BUSINESS RESEARCH METHODS

OBJECTIVE

The objective is to create an awareness of the Process of Research, the tools and techniques of research and generation of reports

Unit 1: INTRODUCTION TO RESEARCH **14Hrs**

Meaning – Objectives – Types of Research – Scope of Research – Research Approaches – Research Process – Research Design – Research Methods Vs Research Methodology - Steps in Research – Problem Formulation – Statement of Research Objective – Exploratory – Descriptive – Experimental Research.

Unit 2: METHODS OF DATA COLLECTION **08 Hrs**

Observational and Survey Methods – Field Work Plan - Administration of surveys - Training field investigators - Sampling methods - Sample size.

Unit 3: TOOLS FOR COLLECTION OF DATA **08 Hrs**

Questionnaire Design; Attitude measurement techniques – Motivational Research Techniques – Selection of Appropriate Statistical Techniques

Unit 4: STATISTICAL METHODS **18 Hrs**

Tabulation of data - Analysis of data –Testing of Hypothesis, Advanced techniques – ANOVA, Chi-Square - Discriminant Analysis - Factor analysis, Conjoint analysis - Multidimensional Scaling - Cluster Analysis (Concepts Only).

Unit 5: REPORT WRITING **08 Hrs**

Types of Reports, Business, Technical and Academic Report writing – Methodology Procedure – Contents – Bibliography

SKILL DEVELOPMENT

- Illustrate different types of samples with examples
- Construct a questionnaire for collection of primary data keeping in mind the topic chosen for research
- Narrate your experience using observation technique
- Diagrammatically present the information collected through the questionnaire

BOOKS FOR REFERENCE

1. O.R.Krishnaswamy; Research methodology in Social Sciences, HPH, 2008.
2. R. Divivedi: Research Methods in Behavior Science, Macmillan India Ltd., 2001.
3. J.K. Sachdeva: Business Research Methodology HPH
4. S.N. Murthy, V. Bhojanna: Business Research Methods Excel Books
5. Levin &Rubin: Statistics for Management, Prentice Hall of India, 2002
6. Gupta S; Research Methodology and Statistical Techniques, Deep & Deep Publication (P) Ltd., 2002
7. Thakur D: Research Methodology in Social Sciences, Deep & Deep Publications (P) Ltd.,1998.
8. Tripathi P.C:A Textbook of Research Methodology, Sultan Chand & Sons, 2002.
9. Cooper: Business Research Methods 6th edition, MC Graw Hill,
10. C.R. Kothari, Research Methodology, Vikas Publications
11. Usha Devi N, Santhosh Kumar - Business Research Methodology

Unit I

INTRODUCTION TO RESEARCH

Meaning – Objectives – Types of Research – Scope of Research – Research Approaches – Research Process – Research Design – Research Methods Vs Research Methodology - Steps in Research – Problem Formulation – Statement of Research Objective – Exploratory – Descriptive – Experimental Research.

*****

Meaning and Definition of Research:

Systematic investigative process employed to increase or revise current knowledge by discovering new facts.
It is divided into two general categories:
(1) Basic research is inquiry aimed at increasing scientific knowledge, and
(2) Applied research is effort aimed at using basic research for solving problems or developing new processes, products, or techniques.

Research is work that involves studying something and trying to discover facts about it.investigation, study, inquiry, analysis, examination, probe, exploration, scrutiny, experimentation, delving, groundwork, fact-finding careful, systematic, patient study and investigation in some field of knowledge, undertaken to discover or establish facts or principles

Definition of research

1. The collecting of information about a particular subject
2. To study (something) carefully
3. To collect information about or for (something)
4. Careful or diligent search
8. studious inquiry or examination; especially **:** investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws

What is Research?

Research is a systematic inquiry that investigates hypothesis, suggests new interpretations of data or texts, and poses new questions for future research to explore.

Research consists of:

- Asking a question that nobody has asked before;
- Doing the necessary work to find the answer; and
- Communicating the knowledge you have acquired to a larger audience.

Objectives of Research:

The purpose of research is to discover answers to questions through the application of scientific procedures.

The main aim of research is to find out the truth which is hidden and which has not been discovered as yet.

Though each research study has its own specific purpose, we may think of research objectives as falling into a number of following broad groupings:

- To gain familiarity with a phenomenon or to achieve new insights into it (studies with this object in view are termed as exploratory or formulative research studies);
- To portray accurately the characteristics of a particular individual, situation or a group(studies with this object in view are known as descriptive research studies);
- To determine the frequency with which something occurs or with which it is associated with something else (studies with this object in view are known as diagnostic research studies);
- To test a hypothesis of a causal relationship between variables (such studies are known as hypothesis-testing research studies).

Types of Research :

1. Descriptive Research
2. Exploratory Research
3. Applied Research
4. Fundamental Research
5. Quantitative Research
6. Qualitative Research
7. Conceptual Research
8. Empirical Research
9. Experimental Research
10. Historical Research

Specific to field of study

Graduate students learn about research methods used in their particular field of study. Whatever methods are used, there are many resources to support research, and any number of variations to the basic methods. Choose a method or variation that is manageable in your first years as a new faculty member to be certain to get your research agenda underway. Often the biggest challenge is often to get started, so establish research priorities early on, and create a plan to impliment them.

Quantitative Research

Quantitative research is inquiry into an identified problem, based on testing a theory, measured with numbers, and analyzed using statistical techniques. The goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true. We will explore some of the issues and

challenges associated with quantitative research in this section. Seek the advice of faculty members who have conducted quantitative studies for advice, support and encouragement.

Qualitative Research

A study based upon a qualitative process of inquiry has the goal of understanding a social or human problem from multiple perspectives. Qualitative research is conducted in a natural setting and involves a process of building a complex and holistic picture of the phenomenon of interest. We will explore some of the issues and challenges associated with qualitative research in this section. Look for colleagues who engage in qualitative research to serve as a sounding board for procedures and processes you may use as a new faculty member.

Collaborative Research

There are many ways to collaborate and thereby enrich your work as a faculty member. Multiple perspectives offer a more complete view of an issue under study. Whether presenting or publishing, having several iterations of a study in different settings offers a means of validating findings. Collaboration may take many forms, including crossing disciplines, types of institutions, or engaging the community outside the campus environment. Explore the ways that collaboration may enrich your research. Some of the issues and challenges associated with collaborative research are explored in this section.

Practitioner Research

Simple definitions of practitioner research address the investigator, the setting and the purpose. The investigator is the practitioner, in workplace settings ranging from hospitals, to schools and communities. The general purpose is to better align the practitioner's purpose with their actions. There are those who argue that practitioner research stems from a larger social justice movement within qualitative research. Even when social justice is not the sole motivating principle, an underlying commonality of purpose is the desire to improve upon and develop deeper insights into one's practice. Practitioner research by its nature offers practitioners a voice in the research conversation. Some consider it a bridge of sorts between theory and practice, although practitioners claim a rightful place in the research continuum. Some of the issues and challenges associated with practitioner research are explored in this section.

What are the various types of research?

1. Pure research

a. This is 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.

2. 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.

3. 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.

4. 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.

5. Diagnostic study

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.

6. Evaluation study

a. Form of applied research.
b. Studies the development project.
c. Gives access to social or economical programmes.
d. Studies the quality and also the quantity of an activity.

7. Action research

a. Type of evaluation study.
b. Is a concurrent evaluation study.

SCOPE OF RESEARCH

1. To know the Business Competition – Environment

a. To know the Customers:- Know about customers needs,

b. To know the Product – Design, Price, Expectations

2. To know the Industry Competition:

3. Maturing of Management as a group of disciplines

4 Stakeholders demanding greater influence

5. To know the Global Competition

6. Government Intervention

7. Economical Data Collection

RESEARCH APPROACHES: The four main approaches (Types of research)

Approaches of Research

1. Quantitative Approach

2. Qualitative Approach

3. Pragmatic Approach

4. Advocacy / Participatory Approach

Quantitative research

Quantitative research is generally associated with the positivist/postpositivist paradigm. It usually involves collecting and converting data into numerical form so that statistical calculations can be made and conclusions drawn.

The process

Researchers will have one or more **hypotheses.** These are the questions that they want to address which include predictions about possible relationships between the things they want to investigate **(variables)**. In order to find answers to these questions, the researchers will also have various instruments and materials (e.g. paper or computer tests, observation check lists etc.) and a clearly defined plan of action.

Data is collected by various means following a strict procedure and prepared for **statistical analysis**. Nowadays, this is carried out with the aid of sophisticated statistical computer packages. The analysis enables the researchers to determine to what extent there is a relationship between two or more variables. This could be a simple association (e.g. people who exercise on a daily basis have lower blood pressure) or a causal relationship (e.g. daily exercise actually leads to lower blood pressure). Statistical analysis permits researchers to discover complex causal relationships and to determine to what extent one variable influences another.

The results of statistical analyses are presented in journals in a standard way, the end result being a **P value**. For people who are not familiar with scientific research jargon, the discussion sections at the end of articles in peer reviewed journals usually describe the results of the study and explain the implications of the findings in straightforward terms

Principles

Objectivity is very important in quantitative research. Consequently, researchers take great care to avoid their own presence, behaviour or attitude affecting the results (e.g. by changing the situation being studied or causing participants to behave differently). They also critically examine their methods and conclusions for any possible bias.

Researchers go to great lengths to ensure that they are really measuring what they claim to be measuring. For example, if the study is about whether background music has a positive impact on restlessness in residents in a nursing home, the researchers must be clear about what kind of music to include, the volume of the music, what they mean by restlessness, how to measure restlessness and what is considered a positive impact. This must all be considered, prepared and controlled in advance.

External factors, which might affect the results, must also be controlled for. In the above example, it would be important to make sure that the introduction of the music was not accompanied by other changes (e.g. the person who brings the CD player chatting with the residents after the music session) as it might be the other factor which produces the results (i.e. the social contact and not the music). Some possible contributing factors cannot always be ruled out but should be acknowledged by the researchers.

The main emphasis of quantitative research is on deductive reasoning which tends to move from the general to the specific. This is sometimes referred to as a top down approach. The validity of conclusions is shown to be dependent on one or more premises (prior statements, findings or conditions) being valid. Aristotle's famous example of deductive reasoning was: All men are mortal àSocrates is a man à Socrates is mortal. If the premises of an argument are inaccurate, then the argument is inaccurate. This type of reasoning is often also associated with the fictitious character Sherlock Holmes. However, most studies also include an element of inductive reasoning at some stage of the research (see section on qualitative research for more details).

Researchers rarely have access to all the members of a particular group (e.g. all people with dementia, carers or healthcare professionals). However, they are usually interested in being able to make inferences from their study about these larger groups. For this reason, it is important that the people involved in the study are a representative **sample** of the wider population/group. However, the extent to which generalizations are possible depends to a certain extent on the number of people involved in the study, how they were selected and whether they are representative of the wider group. For example, generalizations about psychiatrists should be based on a study involving psychiatrists and not one based on psychology students. In most cases, random samples are preferred (so that each potential participant has an equal chance of participating) but sometimes researchers

might want to ensure that they include a certain number of people with specific characteristics and this would not be possible using random sampling methods. Generalizability of the results is not limited to groups of people but also to situations. It is presumed that the results of a laboratory experiment reflect the real life situation which the study seeks to clarify.

When looking at results, the **P value** is important. P stands for probability. It measures the likelihood that a particular finding or observed difference is due to chance. The P value is between 0 and 1. The closer the result is to 0, the less likely it is that the observed difference is due to chance. The closer the result is to 1, the greater the likelihood that the finding is due to chance (random variation) and that there is no difference between the groups/variables.

Qualitative research

Qualitative research is the approach usually associated with the social constructivist paradigm which emphasises the socially constructed nature of reality. It is about recording, analysing and attempting to uncover the deeper meaning and significance of human behaviour and experience, including contradictory beliefs, behaviours and emotions. Researchers are interested in gaining a rich and complex understanding of people's experience and not in obtaining information which can be generalized to other larger groups.

The process

The approach adopted by qualitative researchers tends to be inductive which means that they develop a theory or look for a pattern of meaning on the basis of the data that they have collected. This involves a move from the specific to the general and is sometimes called a bottom-up approach. However, most research projects also involve a certain degree of deductive reasoning (see section on quantitative research for more details).

Qualitative researchers do not base their research on pre-determined hypotheses. Nevertheless, they clearly identify a problem or topic that they want to explore and may be guided by a theoretical lens - a kind of overarching theory which provides a framework for their investigation.

The approach to data collection and analysis is methodical but allows for greater flexibility than in quantitative research. Data is collected in textual form on the basis of observation and interaction with the participants e.g. through participant observation, in-depth interviews and focus groups. It is not converted into numerical form and is not statistically analysed.

Data collection may be carried out in several stages rather than once and for all. The researchers may even adapt the process mid-way, deciding to address additional issues or dropping questions which are not appropriate on the basis of what they learn during the process. In some cases, the researchers will interview or observe a set number of people. In other cases, the process of data collection and analysis may continue until the researchers find that no new issues are emerging.

Principles

Researchers will tend to use methods which give participants a certain degree of freedom and permit spontaneity rather than forcing them to select from a set of pre-determined responses (of which none might be appropriate or accurately describe the participant's thoughts, feelings, attitudes or behaviour) and to try to create the right atmosphere to enable people to express themselves. This may mean adopting a less formal and less rigid approach than that used in quantitative research.

It is believed that people are constantly trying to attribute meaning to their experience. Therefore, it would make no sense to limit the study to the researcher's view or understanding of the situation and expect to learn something new about the experience of the participants. Consequently, the methods used may be more open-ended, less narrow and more exploratory (particularly when very little is known about a particular subject). The researchers are free to go beyond the initial response that the participant gives and to ask why, how, in what way etc. In this way, subsequent questions can be tailored to the responses just given.

Qualitative research often involves a smaller number of participants. This may be because the methods used such as in-depth interviews are time and labour intensive but also because a large number of people are not needed for the purposes of statistical analysis or to make generalizations from the results.

The smaller number of people typically involved in qualitative research studies and the greater degree of flexibility does not make the study in any way "less scientific" than a typical quantitative study involving more subjects and carried out in a much more rigid manner. The objectives of the two types of research and their underlying philosophical assumptions are simply different. However, as discussed in the section on "philosophies guiding research", this does not mean that the two approaches cannot be used in the same study.

Pragmatic approach to research (mixed methods)

The pragmatic approach to science involves using the method which appears best suited to the research problem and not getting caught up in philosophical debates about which is the best approach. Pragmatic researchers therefore grant themselves the freedom to use any of the methods, techniques and procedures typically associated with quantitative or qualitative research. They recognise that every method has its limitations and that the different approaches can be complementary.

They may also use different techniques at the same time or one after the other. For example, they might start with face-to-face interviews with several people or have a focus group and then use the findings to construct a questionnaire to measure attitudes in a large scale sample with the aim of carrying out statistical analysis.

Depending on which measures have been used, the data collected is analysed in the appropriate manner. However, it is sometimes possible to transform qualitative data into quantitative data and vice versa although transforming quantitative data into qualitative data is not very common.

Being able to mix different approaches has the advantages of enabling triangulation. Triangulation is a common feature of mixed methods studies. It involves, for example:

- the use of a variety of data sources (data triangulation)
- the use of several different researchers (investigator triangulation)
- the use of multiple perspectives to interpret the results (theory triangulation)
- the use of multiple methods to study a research problem (methodological triangulation)

In some studies, qualitative and quantitative methods are used simultaneously. In others, first one approach is used and then the next, with the second part of the study perhaps expanding on the results of the first. For example, a qualitative study involving in-depth interviews or focus group discussions might serve to obtain information which will then be used to contribute towards the development of an experimental measure or attitude scale, the results of which will be analysed statistically.

Advocacy/participatory approach to research (emancipatory)

To some degree, researchers adopting an advocacy/participatory approach feel that the approaches to research described so far do not respond to the needs or situation of people from marginalised or vulnerable groups. As they aim to bring about positive change in the lives of the research subjects, their approach is sometimes described as emancipatory. It is not a neutral stance. The researchers are likely to have a political agenda and to try to give the groups they are studying a voice. As they want their research to directly or indirectly result in some kind of reform, it is important that they involve the group being studied in the research, preferably at all stages, so as to avoid further marginalizing them.

The researchers may adopt a less neutral position than that which is usually required in scientific research. This might involve interacting informally or even living amongst the research participants (who are sometimes referred to as co-researchers in recognition that the study is not simply about them but also by them). The findings of the research might be reported in more personal terms, often using the precise words of the research participants. Whilst this type of research could by criticised for not being objective, it should be noted that for some groups of people or for certain situations, it is necessary as otherwise the thoughts, feelings or behaviour of the various members of the group could not be accessed or fully understood.

Vulnerable groups are rarely in a position of power within society. For this reason, researchers are sometimes members of the group they are studying or have something in common with the members of the group.

Research Process - 10 Steps

Research process contains a series of closely related activities which has to carry out by a researcher. Research process requires patients. There is no measure that shows your research is the best. It is an art rather than a science. Following are the main steps in social or business research process.

1. Selection of **Research Problem**
2. Extensive Literature **Survey**
3. Making Hypothesis
4. Preparing the Research Design
5. Sampling
6. Data collection
7. Data Analysis
8. Hypothesis Testing
9. Generalization and Interpretation
10. Preparation of Report

1. Selection of Research Problem

The selection of topic for research is a difficult job. When we select a title or research statement, then other activities would be easy to perform. So, for the understanding thoroughly the problem it must have to discuss with colleagues, friend, experts and teachers. The research topic or problem should be practical, relatively important, feasible, ethically and politically acceptable.

2. Literature Review or Extensive Literature Survey

After the selection of research problem, the second step is that of literature mostly connected with the topics. The availability of the literature may bring ease in the research. For this purpose academic journals, conference and govt. reports and library must be studied.

3. Making Hypothesis

The development of hypothesis is a technical work depends on the researcher experience. The hypothesis is to draw the positive & negative cause and effect aspects of a problem. Hypothesis narrows down the area of a research and keep a researcher on the right path.

4. Preparing the Research Design

After the formulation of the problem and creating hypothesis for it, research Design is to prepare by the researcher. It may draw the conceptual structure of the problem. Any type of research design may

be made, depend on the nature and purpose of the study. Daring R. Design the information about sources, skill, time and finance is taken into consideration.

5. **Sampling**

The researcher must design a sample. It is a plan for taking its respondents from a specific areas or universe. The sample may be of two types:

1. Probability Sampling
2. Non-probability Sampling

6. **Data collection**

Data collection is the most important work, is researcher. The collection of information must be containing on facts which is from the following two types of researcher.

Primary Data Collection: Primary data may be from the following.

1. Experiment
2. Questionnaire
3. Observation
4. Interview

Secondary data collection: it has the following categories:

1. Review of literature
2. Official and non-official reports
3. Library approach

7. **Data Analysis**

When data is collected, it is forwarded for analysis which is the most technical job. Data analysis may be divided into two main categories.

Data Processing: it is sub-divided into the following.

Data editing, Data coding, Data classification, Data tabulation, Data presentation, Data measurement

Data Exposition: Date Exposition has the following sub-categories.

Description, Explanation, Narration, Conclusion/Findings, Recommendations/Suggestions

8. **Hypothesis Testing**

Research data is then forwarded to test the hypothesis. Do the hypothesis are related to the facts or not? To find the answer the process of testing hypothesis is undertaken which may result in accepting or rejecting the hypothesis.

9. Generalization and Interpretation

The acceptable hypothesis is possible for researcher to arrival at the process of generalization or to make & theory. Some types of research has no hypothesis for which researcher depends upon on theory which is known as interpretation.

10. Preparation of Report

A researcher should prepare a report for which he has done is his work. He must keep in his mind the following points:

Report Design in Primary Stages

The report should carry a title, brief introduction of the problem and background followed by acknowledgement. There should be a table of contents, grapes and charts.

Main Text of the Report

It should contain objectives, hypothesis, explanations and methodology of the research. It must be divided into chapters and every chapter explains separate title in which summary of the findings should be enlisted. The last section would be clearly of conclusions to show the main theme of the R-study.

Closing the Report

After the preparation of report, the last step in business research process contains of bibliography, references, appendices, index and maps or charts for illustration. For this purpose the information should more clearer.

RESEARCH DESIGN

Meaning:

The **research design** refers to the overall strategy that you choose to integrate the different components of the study in a coherent and logical way, thereby, ensuring you will effectively address the **research** problem; it constitutes the blueprint for the collection, measurement, and analysis of data.

A detailed outline of how an investigation will take place. A research design will typically include how data is to be collected, what instruments will be employed, how the instruments will be used and the intended means for analyzing data collected.

The key features of a research design are:

(i) It is a plan which describes the sources and kinds of information strongly related to the research problem.
(ii) It is a strategy indicating which method will be employed for collecting and examining the data.
(iii) It also consists of the time and cost budgets because most studies are done under these two limitations.
In a nutshell, research design must, at least, contain

(a) a clear statement of the research problem;

(b) processes and methods to be utilized for collecting data;

(c) the population to be researched; and

(d) techniques to be employed in processing and examining data.

Research design is what makes the entire research project work .

More explicitly, the design decisions happen to be in respect of:

- What is the study about?
- Why is the study being made?
- Where will the study be carried out?
- What type of data is required?
- Where can the required data be found?
- What periods of time will the study include?
- What will be the sample design?
- What techniques of data collection will be used?
- How will the data be analyzed?
- In what style will the report be prepared?

Advantages of research design

1. Consumes less time.
2. Ensures project time schedule.

3. Helps researcher to prepare himself to carry out research in a proper and a systematic way.

4. Better documentation of the various activities while the project work is going on.

5. Helps in proper planning of the resources and their procurement in right time.

6. Provides satisfaction and confidence, accompanied with a sense of success from the beginning of the work of the research project.

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. Research design has a significant impact on the reliability of the results obtained. It thus acts as a firm foundation for the entire research.

For example, economical and attractive construction of 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.

The need for research design is as follows:

- It reduces inaccuracy;
- Helps to get maximum efficiency and reliability;
- Eliminates bias and marginal errors;
- Minimizes wastage of time;
- Helpful for collecting research materials;
- Helpful for testing of hypothesis;
- Gives an idea regarding the type of resources required in terms of money, manpower, time, and efforts;
- Provides an overview to other experts;
- Guides the research in the right direction.

Research Methods Vs Research Methodology

Definition of Research Method

Research method pertains to all those methods, which a researcher employs to undertake research process, to solve the given problem. The techniques and procedure, that are applied during the course of studying research problem are known as the research method. It encompasses both qualitative and quantitative method of performing research operations, such as survey, case study, interview, questionnaire, observation, etc.

These are the approaches, which help in collecting data and conducting research, in order to achieve specific objectives such as theory testing or development. All the instruments and behaviour, used at various levels of the research activity such as making observations, data collection, data processing, drawing inferences, decision making, etc. are included in it. Research methods are put into three categories:

- **First Category**: The methods relating to data collection are covered. Such methods are used when the existing data is not sufficient, to reach the solution.
- **Second Category:** Incorporates the processes of analysing data, i.e. to identify patterns and establish a relationship between data and unknowns.
- **Third Category**: Comprise of the methods which are used to check the accuracy of the results obtained.

What are Research Methods?

As mentioned above, **research methods are the methods used for data collection in research**. ***Research methods involve surveys, interviews, case studies, observation, experiments, etc***. It can be said that research methods are mainly used to gather information so that the researcher can find answers to his research problem.

When speaking of research methods whether it is the natural sciences or else the social sciences there is a vast range of methods that can be used. In the natural sciences, the researcher is mostly interested in gaining quantitative data that will allow him to provide specific conclusions. But in the social sciences the research methods mostly provide the researcher with quantitative data. However, this does not mean that in social sciences qualitative data is ignored. On the contrary, a combination of data can be used for social research.

Definition of Research Methodology

Research Methodology, as its name suggest is the study of methods, so as to solve the research problem. It is the science of learning the way research should be performed systematically. It refers to the rigorous analysis of the methods applied in the stream of research, to ensure that the conclusions drawn are valid, reliable and credible too.

The researcher takes an overview of various steps that are chosen by him in understanding the problem at hand, along with the logic behind the methods employed by the researcher during study.

It also clarifies the reason for using a particular method or technique, and not others, so that the results obtained can be assessed either by the researcher himself or any other party.

What is a Research Methodology?

Research methodology explains the overarching theoretical and philosophical frameworks which guide the research. Research methodology works as a framework within which the researcher works. It is even accurate to consider it as the inception of the research. For various researches, the researcher can employ different methodologies. This will allow him to look at the research problem from different angles and use different methods, techniques and even perspectives.

Let us take an example and comprehend the difference between research methods and research methodology. A research that is being conducted on the stigmatization of HIV patients can employ a variety of research methods. They are interviews, observation and even case studies. These allow the researcher to collect data from the participants. This allows him to find answers to his research questions and overall research problem.

When paying attention to the research methodology, it refers to the broader framework that is used by the researcher to conduct the research. This will decide what types of methods the researcher use, the theoretical perspectives, etc. In this sense, the methodology works more as an overall guide to the research.

What are the Difference Between Research Methods and Research Methodology?

Definitions of Research Methods and Research Methodology:

Research methods: Research methods are the methods used for data collection in a research.

Research methodology: Research methodology explains the overarching theoretical and philosophical frameworks which guide the research.

Characteristics of Research Methods and Research Methodology:

Content:

Research Methods: Research methods involve surveys, interviews, case studies, observation, experiments, etc.

Research Methodology: Research methodology involves the theoretical frameworks and learning of the various techniques that can be used in the conduct of research and the conduct of tests, experiments, surveys and critical studies.

Aim:

Research Methods: Research methods aim at finding solutions to research problems.

Research Methodology: Research methodology aims at the employment of the correct procedures to find out solutions.

Relationship:

Research Methods: Research methods are the end of any research.

Research Methodology: Research methodology is the beginning.

Key Differences Between Research Method and Research Methodology

The differences between research method and research methodology can be drawn clearly on the following grounds:

1. The research method is defined as the procedure or technique applied by the researcher to undertake research. On the other hand, research methodology is a system of methods, used scientifically for solving the research problem.

2. The research method is nothing but the behaviour or tool, employed in selecting and building research technique. Conversely, research methodology implies the science of analysing, the manner in which research is conducted appropriately.

3. The research method is concerned with carrying out experiment, test, surveys, interviews, etc. As against this, research methodology is concerned with learning various techniques which can be employed in the performance of experiment, test or survey.

4. Research method covers various investigation techniques. Unlike, research methodology, which consists of complete approach aligned towards the attainment of purpose.

5. Research method intends to discover the solution to the problem at hand. In contrast, research methodology aspires to apply appropriate procedures, with a view to ascertaining solutions.

Conclusion

The scope of research methodology is wider than that of research method, as the latter is the part of the former. For understanding the research problem thoroughly, the researcher should know the research methodology along with the methods.

In a nutshell, research method refers to the technique which can be adopted to explore the nature of the world that surrounds us. On the contrary, research methodology is the foundation, which helps us to understand the determinants influencing the effectiveness of the methods applied.

Formulating the research problem

Once the general topic or problem has been identified, this should then be stated as a clear **research problem**, that is, taken from just a statement about a problematic situation to a clearly defined researchable problem that identifies the issues you are trying to address.

It is not always easy to formulate the research problem simply and clearly. In some areas of scientific research the investigator might spend years exploring, thinking, and researching before they are clear about what research questions they are seeking to answer. Many topics may prove too wide-ranging to provide a researchable problem. Choosing to study, for instance a social issue such as child poverty, does not in itself provide a researchable problem. The problem is too wide-ranging for one researcher to address. Time and resources would make this unfeasible and the results from such a study would consequently lack depth and focus.

Statement of research problem

An adequate statement of the research problem is one of the most important parts of the research. Different researchers are likely to generate a variety of researchable problems from the same situation since there are many research issues that can arise out of a general problem situation. Your research will be able to pursue only one in depth.

For a problem statement to be effective in the planning of applied research it should have the following characteristics (Andrew and Hildebrand 1982).

1. The problem reflects felt needs
2. The problem is non-hypothetical, ie it must be based on factual evidence
3. It should suggest meaningful and testable hypotheses - to avoid answers that are of little or no use to the alleviation of the problem
4. The problems should be relevant and manageable

Formulating the research problem allows you to make clear, both to yourself and the reader, what the purpose of your research is. Subsequent elaboration of method should be oriented to providing information to address that problem. The problem statement is therefore a very important device for keeping you on track with your research. It is also one means by which your research will be evaluated - does the research address the problem as stated.

5 Ways to Formulate the Research Problem

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.

Research Objectives and Hypotheses

Research Objectives are statements of what the researcher intends to do. The objectives flow directly from the problem. They communicate what the researcher plans to do. Structurally, the objectives are seen as small particles which constitute the problem.

The problem may be stated broadly but the objective should be should be stated in more specific and measurable term:

For example in the problem, *" Students' Extent of Exposure to Mass Media"...*
The phrase "exposure to mass media" is broad. But to state the objective as … *" To determine the students' frequency of reading newspaper," frequency of reading newspapers* is specific and measurable.

Research Objectives are usually stated in declarative form and starts with infinitives like:

1. *to find out*
2. *to determine*
3. *to describe*
4. *to compare*
5. *to test*

Some researchers ask research questions instead. Example

"*This study aims to answer the following questions," or*
"this study sought to answer the following questions.'

The two types of Objective:

1. Immediate or general
2. specific

The **immediate or general objective** specifies the activity that will take place and the variables that will be examined.

The **Specific objectives** may be viewed as sub-objectives or small particles of the general objective. The following should be examined in stating the specific objectives:

1. Specific variables
2. Variables expressed in measurable terms
3. Suggestion on the type of analysis to be done

Research Hypothesis A **hypothesis** is defined as an educated guess or a tentative answer to the research question . it is a statement about an expected relationship betwwn two variables.

Types of hypotheses (main categories)

A **null hypothesis** is a negative statement which indicate s no relationship nor correlation between two variables.

Example

Given Objective: "determine whether there is a significant relationship between the extent of mass media exposure and attitude towards land reform among lowland farmers"

The Null Hypothesis: " There is no significant relationship between the extent of mass media exposure and attitude towards land reform among lowland farmers."

An **alternative hypothesis** is also called the **research hypothesis** . It is the positive form of the null hypothesis.

Example

The Null Hypothesis: " There is a significant relationship between mass media exposure and attitude towards land reform among lowland farmers."

Other classifications of hypotheses

Directional vs. Non-directional hypotheses

A **directional** hypothesis states whether the relationship between two variables is direct (positive) or inverse (negative). A positive or direct relationship is present when the value of one variable increases with the increase in the value of another. A negative relationship is present when the value one variable increases as the value of other variable decreases.

Example:
" The higher the level of exposure of farmers to mass media, the more favourable their attitude towards land refor"- - - - positive
" The more time employees spend in meetings, the less productive they are." - - - - negative
" the higher the advertisement expenses of food establishments, the higher their monthly gross sales."- - - - positive

A **non-directional** hypothesis does not specify the direction of relationship between variables. It merely states the presence or absence of a relationship between two variables or that one influences (or does not influence) the other.

Example:
"The women's educational attainment is significantly associated with their extent of participation in decision-making at home."

TYPES OF RESEARCH

1. **Exploratory Research**
2. **Descriptive Research**
3. **Experimental Research.**

Definition of Exploratory Research

Even as children, we have a natural curiosity about the world around us. We ask questions like: Why is the sky blue? Why do birds fly? Questions like these are often the foundation of exploratory research because they reveal our desire to understand the world around us.

Exploratory research (or ER) is an examination into a subject in an attempt to gain further insight. With ER, a researcher starts with a general idea and uses research as a tool to identify issues that could be the focus of future research.

Look at how ER is used in business. For instance, let's say you own a bakery called The Cupcake King. If you wanted to improve your sales, but weren't sure where to start, you might employ ER to find out the areas of your business that need improvement.

It's important to note that the point of exploratory research is not to gain a definitive answer, like you would with a math problem. For instance, you know that no matter how many different ways you look at the math problem 1 + 1, the answer is always 2.

Exploratory Research Methods

You may wonder how you can explore a topic if there is little information about it. There are several methods that are used in exploratory research. Researchers may use primary or secondary research, or a combination of both types of research.

Primary research is data that someone collects personally, usually from a group of people gathered specifically for the study. Primary research is collected through the use of interviews, focus groups, customer surveys, or any way that organizations are able to obtain feedback. For instance, social media and blogs are a great way for business owners to obtain customer feedback.

Secondary research is the analysis and synthesis of primary research that was compiled at a previous date. Secondary research can be gathered from marketing research data, magazines, old reports, or any other source where relevant information has been stored.

Once upon a time, someone had the idea that the world was flat and that if you went too far, you would surely fall off. We now know that is not true. We know this because of ER. When you conduct ER, you are an explorer, like Magellan or Lewis and Clark or even Dora the Explorer! Before explorers set out on a new adventure, they gather primary and secondary research. They look at similar expeditions, talk to others about their expeditions, and gather any data that will be helpful in guiding them on their journey. ER is the initial research conducted so you understand where you need to focus your efforts or where to point your compass.

2. DESCRIPTIVE RESEARCH

Definition

As the name implies, descriptive research methods are used when *the researcher wants to describe specific behavior as it occurs in the environment*. There are a variety of descriptive research methods available, and once again, the nature of the question that needs to be answered drives which method is used. Traditionally, descriptive research involves three main categories: *observation, case studies, and surveys*. Let's take a closer look at each of these individually.

A descriptive study is one in which information is collected without changing the environment (i.e., nothing is manipulated). It is used to obtain information concerning the current status of the phenomena to describe "what exists" with respect to variables or conditions in a situation. The methods involved range from the survey which describes the status quo, the correlation study which investigates the relationship between variables, to developmental studies which seek to determine changes over time.

1. Statement of the problem
2. Identification of information needed to solve the problem
3. Selection or development of instruments for gathering the information
4. Identification of target population and determination of sampling procedure
5. Design of procedure for information collection
6. Collection of information
7. Analysis of information

8. Generalizations and/or predictions

4. Experimental Research.

Research designs are either experimental or non-experimental. Experimental research is conducted mostly in laboratories in the context of basic research. The principle advantage of experimental designs is that it provides the opportunity to identify cause-and-effect relationships. Non-experimental research, e.g., case studies, surveys, correlation studies, is non-manipulative observational research usually conducted in natural settings. While laboratory-controlled experimental studies tend to be higher in internal validity, non-experimental studies tend to be higher in external validity.

One major limitation of experimental research is that studies are typically conducted in contrived or artificial laboratory settings. Results may not generalize or extrapolate to external settings. Two exceptions to this rule are natural experiments and field experiments. Natural experiments document and compare the behaviors of subjects before and after some natural event; e.g., floods, tornadoes, hurricanes. Field experiments involve manipulating conditions in the natural setting for the purpose of determining their influence on behavior. The field experiment is unique insofar as it tends to be moderately high on both external and internal validity.

In experimental research, the investigator manipulates conditions for the purpose of determining their effect on behavior. Subjects should be unaware of their membership in an experimental group so that they don't act differently (Hawthorne Effect). In the simplest experimental design, investigators administer a placebo to the control group and a treatment to the experimental group. Experimental designs vary in terms of subjects' assignments to different groups, whether subjects were pre-tested, whether different treatments were administered to different groups, and the number of variables being investigated.

Experiments are typically structured in terms of independent, organism, and dependent variables. The independent variable is a manipulated environmental stimulus dimension, the organism-variable is some dimension (e.g., sex, race) of more or less stable characteristics of the organism, and the dependent variable is a behavioral dimension that reflects the influence of the independent and organism-variables. The general objective in experimental research is to define the relationship between the antecedent (independent and organism) variables and the consequent (dependent) variables.

Unit II

METHODS OF DATA COLLECTION

Observational and Survey Methods – Field Work Plan - Administration of surveys - Training field investigators - Sampling methods - Sample size.

Data Collection

Definition

Data collection is the process of gathering and measuring data, information or any variables of interest in a standardized and established manner that enables the collector to answer or test hypothesis and evaluate outcomes of the particular collection. This is an integral, usually initial, component of any research done in any field of study such as the physical and social sciences, business, humanities and others.

Data collection is concerned with the accurate acquisition of data; although methods may differ depending on the field, the emphasis on ensuring accuracy remains the same. The primary goal of any data collection endeavor is to capture quality data or evidence that easily translates to rich data analysis that may lead to credible and conclusive answers to questions that have been posed.

Data collection is the systematic approach to gathering and measuring information from a variety of sources to get a complete and accurate picture of an area of interest. Data collection enables a person or organization to answer relevant questions, evaluate outcomes and make predictions about future probabilities and trends.

Types of Data: Primary and Secondary data

There are many ways of classifying data.

A common classification is **based upon *who collected the data***.

Primary data: Data collected by the investigator himself/ herself for a specific purpose.

Examples: Data collected by a student for his/her thesis or research project.

(In movies) The hero is directly told by the heroine that he is her "ideal man".

Secondary data: Data collected by someone else for some other purpose (but being utilized by the investigator for another purpose).

Examples: Census data being used to analyze the impact of education on career choice and earning.

Some Advantages of using Primary data:

1. The investigator collects data specific to the problem under study.
2. There is no doubt about the quality of the data collected (for the investigator).
3. If required, it may be possible to obtain additional data during the study period.

Some Disadvantages of using Primary data (for reluctant/ uninterested investigators):

1. The investigator has to contend with all the hassles of data collection-

- deciding why, what, how, when to collect
- getting the data collected (personally or through others)
- getting funding and dealing with funding agencies
- ethical considerations (consent, permissions, etc.)

2. Ensuring the data collected is of a high standard-

- all desired data is obtained accurately, and in the format it is required in
- there is no fake/ cooked up data
- unnecessary/ useless data has not been included

3. Cost of obtaining the data is often the major expense in studies

Some Advantages of using Secondary data:

1. The data's already there- no hassles of data collection
2. It is less expensive
3. The investigator is not personally responsible for the quality of data ("I didn't do it")

Some disadvantages of using Secondary data:

1. The investigator cannot decide what is collected (if specific data about something is required, for instance).
2. One can only hope that the data is of good quality
3. Obtaining additional data (or even clarification) about something is not possible

Data Collection Techniques

Information you gather can come from a range of sources. Likewise, there are a variety of techniques to use when gathering primary data. Listed below are some of the most common data collection techniques.

- Interviews
- Questionnaires and Surveys
- Observations
- Focus Groups
- Ethnographies, Oral History, and Case Studies
- Documents and Records

Overview of Different Techniques Data Collection Techniques

Technique	Key Facts	Example
Interviews	• Interviews can be conducted in person or over the telephone • Interviews can be done formally (structured), semi-structured, or informally • Questions should be focused, clear, and encourage open-ended responses • Interviews are mainly qualitative in nature	One-on-one conversation with parent of at-risk youth who can help you understand the issueto see a sample key informant interview.
Questionnaires and Surveys	• Responses can be analyzed with quantitative methods by assigning numerical values to Likert-type scales • Results are generally easier (than qualitative techniques) to analyze • Pretest/Posttest can be compared and analyzed	Results of a satisfaction survey or opinion survey to see an example of a survey created using the CYFERnetSEARCH Interactive Survey Builder feature.to see a sample survey on middle school youth risk behavior.

Observations	• Allows for the study of the dynamics of a situation, frequency counts of target behaviors, or other behaviors as indicated by needs of the evaluation • Good source for providing additional information about a particular group, can use video to provide documentation • Can produce qualitative (e.g., narrative data) and quantitative data (e.g., frequency counts, mean length of interactions, and instructional time)	Site visits to an after-school program to document the interaction between youth and staff within the program
Focus Groups	• A facilitated group interview with individuals that have something in common • Gathers information about combined perspectives and opinions • Responses are often coded into categories and analyzed thematically	A group of parents of teenagers in an after-school program are invited to informally discuss programs that might benefit and help their children succeed
Ethnographies, Oral History, and Case Studies	• Involves studying a single phenomenon • Examines people in their natural settings • Uses a combination of techniques such as observation, interviews, and surveys • Ethnography is a more holistic approach to evaluation • Researcher can become a confounding variable	Shadowing a family while recording extensive field notes to study the experience and issues associated with youth who have a parent or guardian that has been deployed for an example of an oral history. for an additional example of an oral history.

Documents and Records	• Consists of examining existing data in the form of databases, meeting minutes, reports, attendance logs, financial records, newsletters, etc. • This can be an inexpensive way to gather information but may be an incomplete data source	To understand the primary reasons students miss school, records on student absences are collected and analyzed for an example of a searchable database of aggregate data on youth risk behavior.

FIELD WORK PLAN:

Meaning

Field research or **fieldwork** is the collection of information outside a laboratory, library or workplace setting. The approaches and methods used in field research vary across disciplines. For example, biologists who conduct field research may simply observe animals interacting with their environments, whereas social scientists conducting field research may interview or observe people in their natural environments to learn their languages, folklore, and social structures.

Field research involves a range of well-defined, although variable, methods: informal interviews, direct observation, participation in the life of the group, collective discussions, analyses of personal documents produced within the group, self-analysis, results from activities undertaken off- or on-line, and life-histories. Although the method generally is characterized as qualitative research, it may (and often does) include quantitative dimensions.

Administration of Surveys:

Survey Research and Administration

Survey research is a method in which data is collected from a target population, called the sample, by personal interviews, online surveys, the telephone, or paper questionnaires. Some forms of survey research such as online surveys may be completed in an automated fashion. The professionals at Statistics Solutions provide survey administration help to master's and doctoral candidates in the survey administration phase of their research.

The choice of survey instrument(s) used to gather data for your thesis or dissertation is critical. If you are planning to create your own survey instrument and administer it online (e.g., Survey Monkey, QuestionPro, PsychData or Zoomerang), Statistics Solutions can help you create the survey questions and any subscales so they can be easily analyzed and answer your research questions. Our consultants can then help you validate your instrument and expedite the IRB approval process by helping you avoid the typical university and committee pitfalls.

Sampling methods

Meaning of Sampling:

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample may fairly generalize our results back to the population from which they were chosen. Let's begin by covering some of the key terms in sampling like "population" and "sampling frame." Then, because some types of sampling rely upon quantitative models, we'll talk about some of the statistical terms used in sampling. Finally, the major distinction between probability and Nonprobability sampling methods and work through the major types in each.

Sampling Methods can be classified into one of two categories:

- **Probability Sampling**: Sample has a known probability of being selected
- **Non-probability Sampling**: Sample does not have known probability of being selected as in convenience or voluntary response surveys

Probability Sampling

In probability sampling it is possible to both determine which sampling units belong to which sample and the probability that each sample will be selected. The following sampling methods are examples of **probability sampling**:

1. **Simple Random Sampling (SRS)**
2. **Stratified Sampling**
3. **Cluster Sampling**
4. **Systematic Sampling**
5. **Multistage Sampling (in which some of the methods above are combined in stages)**

Of the five methods listed above, students have the most trouble distinguishing between **stratified sampling** and **cluster sampling**.

Stratified Sampling is possible when it makes sense to partition the population into groups based on a factor that may influence the variable that is being measured. These groups are then called strata. An individual group is called a stratum. With **stratified sampling** one should:

- partition the population into groups (strata)
- obtain a simple random sample from each group (stratum)
- collect data on each sampling unit that was randomly sampled from each group (stratum)

Stratified sampling works best when a heterogeneous population is split into fairly homogeneous groups. Under these conditions, stratification generally produces more precise estimates of the

population percents than estimates that would be found from a simple random sample. **Table 3.2** shows some examples of ways to obtain a stratified sample.

Cluster Sampling is very different from Stratified Sampling. With **cluster sampling** one should

- divide the population into groups (clusters).
- obtain a simple random sample of so many clusters from all possible clusters.
- obtain data on every sampling unit in each of the randomly selected clusters.

It is important to note that, unlike with the strata in stratified sampling, the clusters should be microcosms, rather than subsections, of the population. Each cluster should be heterogeneous. Additionally, the statistical analysis used with cluster sampling is not only different, but also more complicated than that used with stratified sampling.

Non-probability Sampling

The following sampling methods that are listed in your text are types of **non-probability sampling that should be avoided**:

1. **volunteer samples**
2. **haphazard (convenience) samples**

Since **such non-probability sampling methods** are based on human choice rather than random selection, statistical theory cannot explain how they might behave and potential sources of bias are rampant. In your textbook, the two types of non-probability samples listed above are called "sampling disasters."

Sample size:

The Definition of Sample Size

Sample size measures the number of individual samples measured or observations used in a survey or experiment.

Sample size is an important concept in statistics, and refers to the number of individual pieces of data collected in a survey. A survey or statistic's sample size is important in determining the **accuracy and reliability** of a survey's findings.

The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. In practice, the sample size used in a study is determined based on the expense of data collection, and the need to have sufficient statistical power.

What is a “Sample Size”?

A sample size is a **part of the population** chosen for a survey or experiment. For example, you might take a survey of dog owner's brand preferences. You won't want to survey *all* the millions of dog owners in the country (either because it's too expensive or time consuming), so you take a sample size. That may be several thousand owners. The sample size is a *representation* of all dog owner's brand preferences. If you choose your sample wisely, it will be a good representation.

Sample size determination

Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. In practice, the sample size used in a study is determined based on the expense of data collection, and the need to have sufficient statistical power. In complicated studies there may be several different sample sizes involved in the study: for example, in a stratified survey there would be different sample sizes for each stratum. In a census, data are collected on the entire population, hence the sample size is equal to the population size. In experimental design, where a study may be divided into different treatment groups, this may be different sample sizes for each group.

Sample sizes may be chosen in several different ways:

- experience – for example, include those items readily available or convenient to collect. A choice of small sample sizes, though sometimes necessary, can result in wide confidence intervals or risks of errors in statistical hypothesis testing.
- using a target variance for an estimate to be derived from the sample eventually obtained
- using a target for the power of a statistical test to be applied once the sample is collected.
- using a confidence level determines how accurate a result will turn out with lower chances of error.

****

Unit III
TOOLS FOR COLLECTION OF DATA

Questionnaire Design; Attitude measurement techniques – Motivational Research Techniques – Selection of Appropriate Statistical Techniques

TOOLS FOR COLLECTION OF DATA
Introduction:

- Data Collection is an important aspect of any type of research study. Inaccurate data collection can impact the results of a study and ultimately lead to invalid results.

- Data collection methods for impact evaluation vary along a continuum. At the one end of this continuum are quantitative methods and at the other end of the continuum are Qualitative methods for data collection .
- Depending on the nature of the information to be gathered, different instruments are used to conduct the assessment: forms for gathering data from official sources such as police or school records; surveys/interviews to gather information from youth, community residents, and others; and focus groups to elicit free-flowing perspectives.

TYPES OF TOOLS

The various methods of data gathering involve the use of appropriate recording forms. These are called tools or instruments of data collection. They consist of

- Observation schedule
- Interview guide
- Interview schedule
- Mailed questionnaire
- Rating scale
- Checklist
- Document schedule/data sheet
- Schedule for institutions
- Each of the above tools is used for a specific method of data gathering: Observation schedule for observation method, interview schedule and interview guide for interviewing, questionnaire for mail survey, and so on.

Questionnaire Design Process

Definition: Questionnaire is a systematic, data collection technique consists of a series of questions required to be answered by the respondents to identify their attitude, experience, and behavior towards the subject of research.

One of the most critical parts of the survey is the creation of questions that must be framed in such a way that it results in obtaining the desired information from the respondents. There are no scientific principles that assure an ideal questionnaire and in fact, the questionnaire design is the skill which is learned through experience.

The following steps are involved in the questionnaire design process:

1. **Specify the Information Needed:** The first and the foremost step in designing the questionnaire is to specify the information needed from the respondents such that the objective of the survey is fulfilled. The researcher must completely review the components of the problem, particularly the hypothesis, research questions, and the information needed.

2. **Define the Target Respondent:** At the very outset, the researcher must identify the target respondent from whom the information is to be collected. The questions must be designed keeping in mind the type of respondents under study. Such as, the questions that are appropriate for serviceman might not be appropriate for a businessman. The less diversified respondent group shall be selected because the more diversified the group is, the more difficult it will be to design a single questionnaire that is appropriate for the entire group.

3. **Specify the type of Interviewing Method:** The next step is to identify the **way in which the respondents are reached**. In personal interviews, the respondent is presented with a questionnaire and interacts face-to-face with the interviewer. Thus, lengthy, complex and varied questions can be asked using the personal interview method. In telephone interviews, the respondent is required to give answers to the questions over the telephone. Here the respondent cannot see the questionnaire and hence this method restricts the use of small, simple and precise questions.

   The questionnaire can be sent through mail or post. It should be self-explanatory and contain all the important information such that the respondent is able to understand every question and gives a complete response. The electronic questionnaires are sent directly to the mail ids of the respondents and are required to give answers online.

4. **Determine the Content of Individual Questions:** Once the information needed is specified and the interviewing methods are determined, the next step is to decide the content of the question. The researcher must decide on what should be included in the question such that it contribute to the information needed or serve some specific purpose.

   In some situations, the indirect questions which are not directly related to the information needed may be asked. It is useful to ask neutral questions at the beginning of a questionnaire with intent to establish respondent's involvement and rapport. This is mainly done when the subject of a questionnaire is sensitive or controversial. The researcher must try to avoid the use of **double-barreled questions**. A question that talks about two issues simultaneously, such as Is the Real juice tasty and a refreshing health drink?

5. **Overcome Respondent's Inability and Unwillingness to Answer:** The researcher should not presume that the respondent can provide accurate responses to all the questions. He must attempt to overcome the respondent's inability to answer. The questions must be designed in a simple and easy language such that it is easily understood by each respondent. In situations, where the respondent is not at all informed about the topic of interest, then the researcher may ask the **filter questions**, an initial question asked in the questionnaire to identify the prospective respondents to ensure that they fulfil the requirements of the sample.

   Despite being able to answer the question, the respondent is unwilling to devote time in providing information. The researcher must attempt to understand the reason behind such unwillingness and design the questionnaire in such a way that it helps in retaining the respondent's attention.

6. **Decide on the Question Structure:** The researcher must decide on the structure of questions to be included in the questionnaire. The question can be structured or unstructured. The **unstructured questions are the open-ended questions** which are answered by the respondents in their own words. These questions are also called as a **free-response** or **free-answer question**s.

   While, the **structured questions are called as closed-ended questions** that pre-specify the response alternatives. These questions could be a multiple choice question, dichotomous (yes or no) or a scale.

7. **Determine the Question Wording:** The desired question content and structure must be translated into **words which are easily understood** by the respondents. At this step, the researcher must translate the questions in easy words such that the information received from the respondents is similar to what was intended.

   In case the question is written poorly, then the respondent might refuse to answer it or might give a wrong answer. In case, the respondent is reluctant to give answers, then "**nonresponse"** arises which increases the complexity of data analysis. On the other hand, if the wrong information is given, then **" response error"** arises due to which the result is biassed.

8. **Determine the Order of Questions:** At this step, the researcher must decide the **sequence in which the questions are to be asked**. The opening questions are crucial in establishing respondent's involvement and rapport, and therefore, these questions must be interesting, non-threatening and easy. Usually, the **open-ended questions** which ask respondents for their opinions are considered as good opening questions, because people like to express their opinions.

9. **Identify the Form and Layout:** The **format, positioning and spacing** of questions has a significant effect on the results. The layout of a questionnaire is specifically important for the self-administered questionnaires. The questionnaires must be divided into several parts, and each part shall be numbered accurately to clearly define the branches of a question.

10. **Reproduction of Questionnaire:** Here, we talk about the **appearance of the questionnaire,** i.e. the quality of paper on which the questionnaire is either written or printed. In case, the questionnaire is reproduced on a poor-quality paper; then the respondent might feel the research is unimportant due to which the quality of response gets adversely affected.

    Thus, it is recommended to reproduce the questionnaire on a good-quality paper having a professional appearance. In case, the questionnaire has several pages, then it should be presented in the form of a booklet rather than the sheets clipped or stapled together.

11. **Pretesting:** Pretesting means **testing the questionnaires on a few selected respondents** or a small sample of actual respondents with a purpose of improving the questionnaire by identifying and eliminating the potential problems. All the aspects of the questionnaire must be tested such as question content, structure, wording, sequence, form and layout, instructions, and question difficulty. The researcher must ensure that the respondents in the pretest should be similar to those who are to be finally surveyed.

Attitude Measurement Techniques:

Attitude Scales - Rating Scales to measure data

Scaling Techniques for Measuring Data Gathered from Respondents

The term scaling is applied to the attempts to measure the attitude objectively. Attitude is a resultant of number of external and internal factors. Depending upon the attitude to be measured, appropriate

scales are designed. Scaling is a technique used for measuring qualitative responses of respondents such as those related to their feelings, perception, likes, dislikes, interests and prefer

Types of Scales

1. Nominal Scale
2. Ordinal Scale
3. Interval Scale
4. Ratio Scale

Self Rating Scales

1. Graphic Rating Scale
2. Itemized Rating Scales
    a. Likert Scale
    b. Semantic Differential Scale
    c. Stapel's Scale
    d. Multi Dimensional Scaling
    e. Thurston Scales
    f. Guttman Scales/Scalogram Analysis
    g. The Q Sort technique

Four types of scales are generally used for Marketing Research.

Nominal Scale

This is a very simple scale. It consists of assignment of facts/choices to various alternative categories which are usually exhaustive as well mutually exclusive. These scales are just numerical and are the least restrictive of all the scales. Instances of Nominal Scale are - credit card numbers, bank account numbers, employee id numbers etc. It is simple and widely used when relationship between two variables is to be studied. In a Nominal Scale numbers are no more than labels and are used specifically to identify different categories of responses. Following example illustrates -

What is your gender?
[ ] Male
[ ] Female

Another example is - a survey of retail stores done on two dimensions - way of maintaining stocks and daily turnover.

How do you stock items at present?
[ ] By product category
[ ] At a centralized store
[ ] Department wise
[ ] Single warehouse

Daily turnover of consumer is?
[ ] Between 100 – 200
[ ] Between 200 – 300
[ ] Above 300

A two way classification can be made as follows

Daily/Stock Turnover Method	Product Category	Department wise	Centralized Store	Single Warehouse
100 – 200				
200 – 300				
Above 300				

Mode is frequently used for response category.

1 Ordinal Scale

Ordinal scales are the simplest attitude measuring scale used in Marketing Research. It is more powerful than a nominal scale in that the numbers possess the property of rank order. The ranking of certain product attributes/benefits as deemed important by the respondents is obtained through the scale.

Example 1: Rank the following attributes (1 - 5), on their importance in a microwave oven.

1. Company Name
2. Functions
3. Price
4. Comfort
5. Design

The most important attribute is ranked 1 by the respondents and the least important is ranked 5. Instead of numbers, letters or symbols too can be used to rate in a ordinal scale. Such scale makes no attempt to measure the degree of favourability of different rankings.

Example 2 - If there are 4 different types of fertilizers and if they are ordered on the basis of quality as Grade A, Grade B, Grade C, Grade D is again an Ordinal Scale.

Example 3 - If there are 5 different brands of Talcom Powder and if a respondent ranks them based on say, "Freshness" into Rank 1 having maximum Freshness Rank 2 the second maximum Freshness, and so on, an Ordinal Scale results.

Median and *mode* are meaningful for ordinal scale.

2 Interval Scale

Herein the distance between the various categories unlike in Nominal, or numbers unlike in Ordinal, are equal in case of Interval Scales. The Interval Scales are also termed as Rating Scales. An Interval Scale has an arbitrary Zero point with further numbers placed at equal intervals. A very good example of Interval Scale is a Thermometer.

Illustration 1 - How do you rate your present refrigerator for the following qualities.

Company Name	Less Known	1	2	3	4	5	Well Known
Functions	Few	1	2	3	4	5	Many
Price	Low	1	2	3	4	5	High
Design	Poor	1	2	3	4	5	Good
Overall Satisfaction	Very Dis-Satisfied	1	2	3	4	5	Very Satisfied

Such a scale permits the researcher to say that position 5 on the scale is above position 4 and also the distance from 5 to 4 is same as distance from 4 to 3. Such a scale however does not permit conclusion that position 4 is twice as strong as position 2 because no zero position has been established. The data obtained from the Interval Scale can be used to calculate the Mean scores of each attributes over all respondents. The Standard Deviation (a measure of dispersion) can also be calculated.

Ratio Scale

Ratio Scales are not widely used in Marketing Research unless a base item is made available for comparison. In the above example of Interval scale, a score of 4 in one quality does not necessarily mean that the respondent is twice more satisfied than the respondent who marks 2 on the scale. A Ratio scale has a natural zero point and further numbers are placed at equally appearing intervals. For example scales for measuring physical quantities like - length, weight, etc.

The ratio scales are very common in physical scenarios. Quantified responses forming a ratio scale analytically are the most versatile. Rati scale possess all he characteristics of an internal scale, and the ratios of the numbers on these scales have meaningful interpretations. Data on certain demographic or descriptive attributes, if they are obtained through open-ended questions, will have ratio-scale properties. Consider the following questions :

Q 1) What is your annual income before taxes? _____ $
Q 2) How far is the Theater from your home ? _____ miles

Answers to these questions have a natural, unambiguous starting point, namely zero. Since starting point is not chosen arbitrarily, computing and interpreting ratio makes sense. For example we can say that a respondent with an annual income of $ 40,000 earns twice as much as one with an annual income of $ 20,000.

Self Rating Scales

1. Graphic Rating Scale

The respondents rate the objects by placing a mark at the appropriate position on a line that runs from one extreme of the criterion variable to another. Example

0	1	5	7
(poor quality)	(bad quality)	(neither good nor bad)	(good quality)

BRAND 1

This is also known as continuous rating scale. The customer can occupy any position. Here one attribute is taken ex-quality of any brand of icecream.

poor good

BRAND 2

This line can be vertical or horizontal and scale points may be provided. No other indication is there on the continuous scale. A range is provided. To quantify the responses to question

that “indicate your overall opinion about ice-ream Brand 2 by placing a tick mark at appropriate position on the line”, we measure the physical distance between the left extreme position and the response position on the line.; the greater the distance, the more favourable is the response or attitude towards the brand.

Its limitation is that coding and analysis will require substantial amount of time, since we first have to measure the physical distances on the scale for each respondent.

2. Itemized Rating Scales

These scales are different from continuous rating scales. They have a number of brief descriptions associated with each category. They are widely used in Marketing Research. They essentially take the form of the multiple category questions. The most common are - Likert, Sementic, Staple and Multiple Dimension. Others are - Thurston and Guttman.

a. Likert Scale

It was developed Rensis Likert. Here the respondents are asked to indicate a degree of agreement and disagreement with each of a series of statement. Each scale item has 5 response categories ranging from strongly agree and strongly disagree.

5	4	3	2	1
Strongly agree	Agree	Indifferent	Disagree	Strongly disagree

Each statement is assigned a numerical score ranging from 1 to 5. It can also be scaled as -2 to +2.

-2	-1	0	1	2

For example quality of Mother Diary ice-cream is poor then Not Good is a negative statement and Strongly Agree with this means the quality is not good.

Each degree of agreement is given a numerical score and the respondents total score is computed by summing these scores. This total score of respondent reveals the particular opinion of a person.

Likert Scale are of ordinal type, they enable one to rank attitudes, but not to measure the difference between attitudes. They take about the same amount of efforts to create as Thurston scale and are considered more discriminating and reliable because of the larger range of responses typically given in Likert scale.

A typical Likert scale has 20 - 30 statements. While designing a good Likert Scale, first a large pool of statements relevant to the measurement of attitude has to be generated and then from the pool statements, the statements which are vague and non-discriminating have to be eliminated.

Thus, likert scale is a five point scale ranging from 'strongly agreement'to 'strongly disagreement'. No judging gap is involved in this method.

b. Semantic Differential Scale

This is a seven point scale and the end points of the scale are associated with bipolar labels.

1						
Unpleasant	2	3	4	5	6	7 Pleasant
Submissive						Dominant

Suppose we want to know personality of a particular person. We have options-

1. Unpleasant/Submissive
2. Pleasant/Dominant

Bi-polar means two opposite streams. Individual can score between 1 to 7 or -3 to 3. On the basis of these responses profiles are made. We can analyse for two or three products and by joining these profiles we get profile analysis. It could take any shape depending on the number of variables.

Profile Analysis

----------/----------
----------/----------
--------/----------

Mean and *median* are used for comparison. This scale helps to determine overall similarities and differences among objects.

When Semantic Differential Scale is used to develop an image profile, it provides a good basis for comparing images of two or more items. The big advantage of this scale is its simplicity, while producing results compared with those of the more complex scaling methods. The method is easy and fast to administer, but it is also sensitive to small differences in attitude, highly versatile, reliable and generally valid.

c. Stapel's Scale

It was developed by Jan Stapel. This scale has some distinctive features:-

i. i. Each item has only one word/phrase indicating the dimension it represents.

ii. Each item has ten response categories.

iii. Each item has an even number of categories.

iv. The response categories have numerical labels but no verbal labels.

For example, in the following items, suppose for quality of ice cream, we ask respondents to rank from +5 to -5. Select a plus number for words which best describe the ice cream accurately. Select a minus number for words you think do not describe the ice cream quality accurately. Thus, we can select any number from +5,for words we think are very accurate, to -5,for words we think are very inaccurate. This scale is usually presented vertically.

+5
+4
+3
+2
+1
High Quality
-1
-2
-3
-4
-5

This is a unipolar rating scale.

2 Multi Dimensional Scaling

It consists of a group of analytical techniques which are used to study consumer attitudes related to perceptions and preferences. It is used to study-

I. The major attributes of a given class of products perceivedby the consumers in considering the product and by which they compare the different ranks.

II. To study which brand competes most directly with each other.

III. To find out whether the consumers would like a new brand with a combination of characteristics not found in the market.

IV. What would be the consumers ideal combination of product attributes.

V. What sales and advertising messages are compatible with consumers brand perceptions.

It is a computer based technique. The respondents are asked to place the various brands into different groups like similar, very similar, not similar, and so on. A

goodness of fit is traded off on a large number of attributes. Then a lack of fit index is calculated by computer program. The purpose is to find a reasonably small number of dimensions which will eliminate most of the stress. After the configuration for the consumer's preference has been developed, the next step is to determine the preference with regards to the product under study. These techniques attempt to identify the product attributes that are important to consumers and to measure their relative importance.

This scaling involves a unrealistic assumption that a consumer who compares different brands would perceive the differences on the basis of only one attribute. For example, what are the attributes for joining M.Com course. The responses may be -to do PG, to go into teaching line,to get knowledge, appearing in the NET. There are a number of attributes, you can not base decision on one attribute only. Therefore, when the consumers are choosing between brands, they base their decision on various attributes. In practice, the perceptions of the consumers involve different attributes and any one consumer perceives each brand as a composite of a number of different attributes. This is a shortcoming of this scale.

Whenever we choose from a number of alternatives, go for multi- dimensional scaling. There are many possible uses of such scaling like in market segmentation, product life cycle, vendor evaluations and advertising media selection.

The limitation of this scale is that it is difficult to clearly define the concept of similarities and preferences. Further the distances between the items are seen as different

2 Thurston Scales

These are also known as equal appearing interval scales. They are used to measure the attitude towards a given concept or construct. For this purpose a large number of statements are collected that relate to the concept or construct being measured. The judges rate these statements along an 11 category scale in which each category expresses a different degree of favourableness towards the concept. The items are then ranked according to the mean or median ratings assigned by the judges and are used to construct questionnaire of twenty to thirty items that are chosen more or less evenly across the range of ratings.

The statements are worded in such a way so that a person can agree or disagree with them. The scale is then administered to assemble of respondents whose scores are determined by computing the mean or median value of the items agreed with. A person who disagrees with all the items has a score of zero. So, the advantage of this scale is that it is an interval measurement scale. But it is the time consuming method and labour intensive. They are commonly used in psychology and education research.

3 Guttman Scales/Scalogram Analysis

It is based on the idea that items can be arranged along a continuem in such a way that a person who agrees with an item or finds an item acceptable will also agree with or find acceptable all other items expressing a less extreme position. For example - Children should not be allowed to watch indecent programmes or government should ban these programmes or they are not allowed to air on the television. They all are related to one aspect.

In this scale each score represents a unique set of responses and therefore the total score of every individual is obtained. This scale takes a lot of time and effort in development.

They are very commonly used in political science, anthropology, public opinion, research and psychology.

4

5

6 The Q Sort technique

It is used to discriminate among large number of objects quickly. It uses a rank order procedure and the objects are sorted into piles based on similarity with respect to some criteria. The number of objects to be sorted should be between 60-140 approximately. For example, here we are taking nine brands. On the basis of taste we classify the brands into tasty, moderate and non tasty.

We can classify on the basis of price also-Low, medium, high. Then we can attain the perception of people that whether they prefer low priced brand, high or moderate. We can classify sixty brands or pile it into three piles. So the number of objects is to be placed in three piles-low, medium or high.

Thus, the Q-sort technique is an attempt to classify subjects in terms of their similarity to attribute under study.

Motivation Research: Techniques, Uses and Limitation:

Motivation Research Technique:

There are four techniques of conducting motivation research:

(a) Non-disguised Structured Techniques.

(b) Non-disguised, Non-structured Techniques.

(c) Disguised Non-structured Techniques.

(a) Non-disguised Structured Techniques:

This approach employs a standardized questionnaire to collect data on beliefs, feelings, and attitude from the respondent.

Single Question Method:

(I think it is a good product or I think it is a poor product).

Multiple Questions Method:

(Numbers of questionnaires asked about the attitude) and Physiological Tests (laboratory tests such as galvanic skin response, eye movement etc. measure attitudes of people towards products) are carried out under this approval.

(b) Non-disguised, Non-structured Techniques:

These techniques use a non standardized questionnaire. The techniques are also called depth interview, qualitative interviews, unstructured interviews, or focussed interviews. All these techniques are designed to gather information on various aspects of human behaviour including the "why" component.

(c) Disguised, Non-structured Techniques:

In this approach, the purpose of study is not discussed to respondents unlike above two cases. A list of unstructured questions is used to collect data on consumer's attitudes. This art of using disguised and unstructured method is referred to as "Projective Techniques".

The projective techniques include several tests given to the respondents. They may be asked to give their comments on cartoons, pictures, stories etc. The stimuli used for this purpose are capable of answering the respondent to a variety of reactions. A number of Projective Techniques, are available to the market researchers for the purpose of analysing "why" part of consumer behaviour.

Qualitative Techniques:

(Projective Techniques and Word Association as follows).

The main Projective Techniques are:

1. Word Association Test (W.A.I):

The interviewer calls a series of listed words one by one and the respondents quickly replies the first word that enters his mind. The underlying assumption is that by "free associating" with certain

stimuli (words) the responses are timed so that those answers which the respondent "response out" are identified.

2. Sentence Completion:

Sentence completion test is similar to word association test except that the respondent is required to complete an unfinished sentence.

For example, "I do not use shampoos because……."

"Coffee that is quickly made……. "

3. Story Completion:

In this technique the respondent is asked to complete a story, end of which is missing. This enables a researcher to find out the almost exact version of images and feelings of people towards a company's product. This helps in finalising the advertising and promotional themes for the product in question.

4. Research of Ink-blot Tests (or Research Tests):

Motivation Research employs this famous test. These tests are not in much use in marketing research. The research test expresses in a classic way the rationale behind all projective tests, that is, in filling the missing parts of a vague and incomplete stimulus, the respondent projects himself and his personality into the picture.

A lot of ink is put on the piece of paper and reference is made of company, product, and the respondent is asked to give his view points after interpreting what he sees in the blot before him. The respondent say, "ugly packaging of the product", or "excellent performance of the product". This response will help the seller to finalise his marketing strategies.

5. Psychographic Technique:

This includes galvanic skin response, eye movement and eye blink test etc. which uses various Instruments with the physiological responses.

6. Espionage Technique:

There are two methods in this technique:

(i) Use of Hidden Recorders:

Such as hidden tape recorders, cameras used to watch consumers as they make purchases or consume items.

(ii) Rubbish Research:

This is another method of espionage activity. Here, the researcher shifts through the garbage of individuals or groups and record pattern of consumption, waste, and brand preference. It gives most required estimates of consumption of cigarettes, medicines, liquor, and magazines etc.

(d) Disguised Structured Techniques:

When we are to measure those attitudes which respondents might not readily and accurately express, we can use disguised structured techniques. The disguised structured questionnaire are easy to administer and code.

Respondents are given questions which they are not likely to be able to answer accurately. In such circumstances they are compelled to 'guess at' the answers. The respondent's attitude on the subject is assumed to be revealed to the extent and direction in which these guessing errors are committed.

Uses of Motivation Research:

1. Motivation Research leads to useful insights and provides inspiration to creative person in the advertising and packing world.

2. Knowledge and measurement of the true attitude of customers help in choosing the best selling appeal for the product and the best way to represent the product in the sales talk, and in determining the appropriateness and weight age of various promotional methods.

3. Motivation Research can help in measuring changes in attitudes, thus advertising research.

4. Knowledge and measurement of attitudes provides us with an imaginative market segmentation tool and also enables estimating market potential of each additional segment.

5. Strategies to position the offer of the company in a particular market segment should be based on the findings of motivation research.

Limitations of Motivation Research:

1. Cautions are required to be exercised not only in the application of these techniques but also the resultant data should be analysed and interpreted according to the psychological theory.

2. Originally these techniques were developed to collect data from a single individual over a period of time. It is not free from draw backs while we apply these techniques to gather data from a number of individuals.

3. The designing and administering of these techniques need qualified and experimented researchers. Such personnel are not easily available.

Meaning of statistical techniques:

Mathematical concepts, formulas, models, **techniques** used in **statistical** analysis of random data. In comparison, deterministic **methods** are used where the data is easily reproducible or where its behavior is determined entirely by its initial stage and inputs.

Different types of statistical methods:

Some well-known statistical tests and procedures are:

- Analysis of variance (ANOVA)
- Chi-squared test.
- Correlation.
- Factor analysis.
- Mann–Whitney U.
- Mean square weighted deviation (MSWD)
- Pearson product-moment correlation coefficient.
- Regression analysis

What are the statistical tools?

- (When the distribution is skewed **statistical** treatment is more complicated). The primary parameters used are the mean (or average) and the standard deviation and the main **tools** the F-test, the t-test, and regression and correlation analysis

A Quick Guide for Choosing the Appropriate Statistical Test

The most important step in choosing the appropriate statistical procedure is to know what the variables of your study are. What are the independent and dependent variables of your study? How are each of the variables measured? Once you have a better grasp of your variables, you can easily choose the statistical procedure that will best answer your study's questions.

An Example

Five students are asked to design a study that will assess the relationship between using the Wii Fit and weight loss in a group of 150 overweight pre-teens during a month-long period. The weight of the participants is taken at the beginning and at the end of the study. The students come up with five different proposals:

1. Student A proposes to randomly assign half of the group to the Wii intervention. This group will be instructed to do Wii Fit Aerobics training for 30 minutes five times a week. The other half would serve as the control and will not be told to do anything. Student A then wants to determine whether using the Wii Fit would lead to weight loss. Accordingly, she suggests that the participants be classified into one of two groups: no weight loss and weight loss.

2. Student B also proposes to randomly assign half of the group to the Wii intervention. This group will be instructed to do Wii Fit Aerobics training for 30 minutes five times a week. The other half would serve as the control and will not be told to do anything. Student B then wants to determine whether using the Wii Fit would lead to weight loss. Thus, he suggests that weight loss should be defined in terms of the difference in weight prior to and immediately after the study.

3. Student C proposes to randomly assign 30 pre-teens to one of four types of Wii interventions: Aerobics, Yoga, Strength Training, and Balance Games. These four groups will be instructed to do the specific Wii Fit activity for 30 minutes five times a week. The last group of 30 pre-teens would serve as the control and will not be told to do anything. Student C then wants to determine whether using the Wii Fit would lead to weight loss. Thus, she suggests that weight loss should be defined in terms of the difference in weight prior to and immediately after the study.

4. Student D proposes to simply ask the 150 overweight teens to record the number of minutes per day they spend using the Wii Fit. He then suggests that the participants be classified into one of two groups: no weight loss and weight loss.

5. Student E also proposes to simply ask the 150 overweight teens to record the number of minutes per day they spend using the Wii Fit. He then suggests that weight loss should be defined in terms of the difference in weight prior to and immediately after the study.

First Point: Variables of the Study

What are the variables of the study? Using the Wii Fit would be the independent variable of the study while weight loss would be the dependent variable of the study.

Second Point: Definition or Measurement of the Variables

From the example above it is obvious that there are several ways to define or measure the independent and dependent variables of a study. But there are two main questions to consider:

1. Is the independent variable measured categorically or continuously?

2. Is the dependent variable measured categorically or continuously?

Student A. Student A defined Wii Fit use in terms of using the Wii Fit or not using the Wii Fit. Accordingly, Student A defined Wii Fit use in terms of categories. She defined weight loss in terms of no weight loss or weight loss. Thus, her definition of weight loss was categorical.

Student B. Student B also defined Wii Fit use in terms of using the Wii Fit or not using the Wii Fit. Thus, he defined Wii Fit use in terms of categories. But Student B defined weight loss in terms of the difference between weight prior to the study and weight immediately after the study. Weight loss, therefore, was defined continuously.

Student C. Student C defined Wii Fit use in terms of the type of Wii Fit activity. Thus, she defined Wii Fit use in terms of five categories. She also defined weight loss in terms of the difference

between weight prior to the study and weight immediately after the study. Weight loss, therefore, was defined continuously.

Student D. Student D defined Wii Fit use in terms of the number of minutes per day spent using the Wii Fit. As such, Wii Fit use was defined continuously. He defined weight loss in terms of no weight loss or weight loss. Thus, his definition of weight loss was categorical.

Student E. Student E defined Wii Fit use in terms of the number of minutes per day spent using the Wii Fit. As such, Wii Fit use was defined continuously. Student E also defined weight loss in terms of the difference between weight prior to the study and weight immediately after the study. Weight loss, therefore, was defined continuously.

Third Point: Choosing the Appropriate Statistical Procedure

Given that independent and dependent variables can be classified as categorical or continuous, the grid below can be used to classify the more common statistical procedures.

Independent Variable	

Student A. Student A could thus choose to perform either a cross-tabulation analysis or a logistic regression procedure. These tests are useful when the independent and dependent variables are measured categorically.

Student B. Student B would need to conduct an independent t-test procedure since his independent variable would be defined in terms of categories and his dependent variable would be measured continuously. An independent t-test procedure is used only when the independent variable has two categories.

Student C. Student C would need to conduct a one-way ANOVA since her independent variable would be defined in terms of categories and her dependent variable would be measured continuously. One-way ANOVAs are used when the independent variable has three or more categories.

Student D. Student D would use a logistic regression procedure to analyze his data since his independent variable would be measured continuously and his dependent variable would be measured categorically. If Student D defined his dependent variable in terms of three or more categories that could be ranked (e.g., weight gain, no weight loss, weight loss), then he would use an ordinal regression procedure.

Student E. Student E could choose to perform either a Pearson correlation procedure or a linear regression procedure since both of her variables would be defined continuously. Usually, a correlation test is conducted when there is only one independent variable and one independent variable. If Student E wanted to study the relationship between several independent variables (e.g., number of hours spent sleeping, number of calories consumed per day) and weight loss, then she would use a linear regression procedure.

****

Unit IV

STATISTICAL METHODS

Tabulation of data - Analysis of data –Testing of Hypothesis, Advanced techniques – ANOVA, Chi-Square - Discriminant Analysis - Factor analysis, Conjoint analysis - Multidimensional Scaling - Cluster Analysis (Concepts Only).

Statistical Methods:

Methods of collecting, summarizing, analyzing, and interpreting variable numerical data. Statistical methods can be contrasted with deterministic methods, which are appropriate where observations are exactly reproducable or are assumed to be so. While statistical methods are widely used in the life sciences, in economics, and in agricultural science, they also have an important role in the physical sciences in the study of measurement errors, of random phenomena such as radioactivity or meteorological events, and in obtaining approximate results where deterministic solutions are hard to apply.

Data collection involves deciding what to observe in order to obtain information relevant to the questions whose answers are required, and then making the observations. Sampling involves choice of a sufficient number of observations representing an appropriate population. Experiments with variable outcomes should be conducted according to principles of experimental design.

Data summarization is the calculation of appropriate statistics (def. 2) and the display of such information in the form of tables, graphs, or charts. Data may also be adjusted to make different samples more comparable, using ratios, compensating factors, etc.

Statistical analysis relates observed statistical data to theoretical models, such as probability

distributions or models used in regression analysis. By estimating parameters in the proposed model and testing hypotheses about rival models, one can assess the value of the information collected and the extent to which the information can be applied to similar situations. *Statistical prediction* is the application of the model thought to be most appropriate, using the estimated values of the parameters.

Tabulation of Data and Types of Tabulation:

Tabulation is the systematic arrangement of the statistical data in columns or rows. It involves the orderly and systematic presentation of numerical data in a form designed to explain the problem under consideration. Tabulation helps in drawing the inference from the statistical figures.

Tabulation prepares the ground for analysis and interpretation. Therefore a suitable method must be decided carefully taking into account the scope and objects of the investigation, because it is very important part of the statistical methods.

Types of Tabulation

In general, the tabulation is classified in two parts, that is a simple tabulation, and a complex tabulation.
Simple tabulation, gives information regarding one or more independent questions. Complex tabulation gives information regarding two mutually dependent questions.

ONE-WAY TABLE	
DIVISION	POPULATION (Millions)
Karachi	10.875968
Hyderabad	14.186954
Sukkur	12.994401

This table gives us information regarding one characteristic information about the population in different divisions of Sindh.

All questions that can be answered in ONE WAY TABLE are independent of each other. It is therefore an example of a simple tabulation, since the information obtained in it is regarding one independent question, that is the number of persons in various divisions of Sindh in millions.

Two-Way Table

These types of table give information regarding two mutually dependent questions. For example, question is, how many millions of the persons are in the Divisions; the One-Way Table will give the answer. But if we want to know that in the population number, who are in the majority, male, or female. The Two-Way Tables will answer the question by giving the column for female and male. Thus the table showing the real picture of divisions sex wise is as under:

TWO-WAY TABLE			
DIVISION	POPULATION (Millions)		
	Male	Female	Total
Karachi			

Hyderabad Sukkur			

Three-Way Table

Three-Way Table gives information regarding three mutually dependent and inter-related questions. For example, from one-way table, we get information about population, and from two-way table, we get information about the number of male and female available in various divisions. Now we can extend the same table to a three way table, by putting a question, "How many male and female are literate?" Thus the collected statistical data will show the following, three mutually dependent and inter-related questions:

1. Population in various division.
2. Their sex-wise distribution.
3. Their position of literacy.

	THREE-WAY TABLE								
DIVISION	POPULATION (Millions)								
	Male			Female			Total		
	Literate	Illiterate	Total	Literate	Illiterate	Total	Literate	Illiterate	Total
Karachi Hyderabad Sukkur									

This table gives information concerning the literacy of both male and female in various divisions of Sindh. From the table we can explain the sex which has more education in relation to division, and also, we can say whether literacy is low in rural areas than in urban areas.

Higher Order Tables

Higher order tables are those which provide information about a large number of inter related questions. Higher order tables may be of four-way, five-way, six-way etc. Such kinds of tables are called manifold tables.

Advantages of Tabulation

1 The large mass of confusing data is easily reduced to reasonable form, that is understandable to kind.
2.The data once arranged in a suitable form, gives the condition of the situation at a glance, or gives a bird eye view.
3 From the table it is easy to draw some reasonable conclusion or inferences.
4. Tables gave grounds for analysis of the data.
5. Errors, and omission if any are always detected in tabulation.
Therefore the importance of a carefully drawn table is vital for the preparation of data for analysis and interpretation.

ANALYSIS OF DATA:

The process of evaluating data using analytical and logical reasoning to examine each component of the data provided. This form of analysis is just one of the many steps that must be completed when conducting a research experiment. Data from various sources is gathered, reviewed, and then analyzed to form some sort of finding or conclusion. There are a variety of specific data analysis

method, some of which include data mining, text analytics, business intelligence, and data visualizations.
Analysing data to summarise it and look for patterns is an important part of every evaluation.
The options for doing this have been grouped into two categories - quantitative data (number) and qualitative data (text, images).

Options

Numeric analysis
Analysing numeric data such as cost, frequency, physical characteristics.

- Correlation: a statistical measure ranging from +1.0 to -1.0 that indicates how strongly two or more variables are related. A positive correlation (+1.0 to 0) indicates that two variables will either increase or decrease together, while a negative correlation (0 to -1.0) indicates that as one variable increases, the other will decrease.

- Crosstabulations: using contingency tables of two or more dimensions to indicate the relationship between nominal (categorical) variables. In a simple crosstabulation, one variable occupies the horizontal axis and another the vertical. The frequencies of each are added in the intersecting squares and displayed as percentages of the whole, illustrating relationships in the data.

- Data mining: computer-driven automated techniques that run through large amounts of text or data to find new patterns and information.

- Exploratory Techniques: taking a 'first look' at a dataset by summarising its main characteristics, often by using visual methods.

- Frequency tables: a visual way of summarizing nominal and ordinal data by displaying the count of observations (times a value of a variable occurred) in a table.

- Measures of central tendency: a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. The mean (the average value), median (the middle value) and mode (the most frequent value) are all measures of central tendency. Each measure is useful for different conditions.

- Measures of dispersion: a summary measure that provides information about how much variation there is in the data, including the range, inter-quartile range and the standard deviation.

- Multivariate descriptive: providing simple summaries of (large amounts of) information (or data) with two or more related variables.

    - Multiple regression

    - Factor analysis

- Cluster analysis
- Structural equation modelling

- Non-Parametric inferential statistics: methods for inferring conclusions about a population from a sample's data that are flexible and do not follow a normal distribution (ie, the distribution does not parallel a bell curve), including ranking: the chi-square test, binomial test and Spearman's rank correlation coefficient.

- Parametric inferential statistics: methods for inferring conclusions about a population from a sample's data that follows certain parameters: the data will be normal (ie, the distribution parallels the bell curve); numbers can be added, subtracted, multiplied and divided; variances are equal when comparing two or more groups; and the sample should be large and randomly selected.

- Summary statistics: providing a quick summary of data which is particularly useful for comparing one project to another, before and after.

- Time series analysis: observing well-defined data items obtained through repeated measurements over time.

Textual analysis

Analyzing words, either spoken or written, including questionnaire responses, interviews, and documents.

- Content analysis: reducing large amounts of unstructured textual content into manageable data relevant to the (evaluation) research questions.

- Thematic coding: recording or identifying passages of text or images that are linked by a common theme or idea allowing the indexation of text into categories.

- Framework matrices: a method for summarising and analysing qualitative data in a two-by-two matrix table. It allows for sorting data across case and by theme.

- Timelines and time-ordered matrices: aids analysis by allowing for visualisation of key events, sequences and results.

ANALYSIS OF DATA

A Definition of Data Analysis

The process of evaluating data using analytical and logical reasoning to examine each component of the data provided. This form of analysis is just one of the many steps that must be completed when conducting a research experiment. Data from various sources is gathered, reviewed, and then analyzed to form some sort of finding or conclusion. There are a variety of specific data analysis method, some of which include data mining, text analytics, business intelligence, and data visualizations.

Data analysis is a primary component of data mining and Business Intelligence (BI) and is key to gaining the insight that drives business decisions. Organizations and enterprises analyze data from a multitude of sources using Big Data management solutions and customer experience management solutions that utilize data analysis to transform data into actionable insights.
Data analysis for quantitative studies, on the other hand, involves critical analysis and interpretation of figures and numbers, and attempts to find rationale behind the emergence of main findings. Comparisons of primary research findings to the findings of the literature review are critically important for both types of studies – qualitative and quantitative.

Data analysis methods in the absence of primary data collection can involve discussing common patterns, as well as, controversies within secondary data directly related to the research area.

Data Analysis Model

There are seven steps that remain consistent across organizations and their data analysis processes:

- **Decide on the objectives** – Determine objectives for data science teams to develop a quantifiable way to determine whether the business is progressing toward its goals; identify metrics or performance indicators early

- **Identify business levers** – Identify goals, metrics, and levers early in data analysis projects to give scope and focus to data analysis; this means the business should be willing to make changes to improve its key metrics and reach its goals as well

- **Data collection** – Gather as much data from diverse sources as possible in order to build better models and gain more actionable insights

- **Data cleaning** – Improve data quality to generate the right results and avoid making incorrect conclusions; automate the process but involve employees to oversee the data cleaning and ensure accuracy

- **Grow a data science team** – Include on your science team individuals with advanced degrees in statistics who will focus on data modeling and predictions, as well as infrastructure engineers, software developers, and ETL experts; then, give the team the large-scale data analysis platforms they need to automate data collection and analysis

- **Optimize and repeat** – Perfect your data analysis model so you can repeat the process to generate accurate predictions, reach goals, and monitor and report consistently

Types of Data Analysis:

1. Univariate Data Analysis

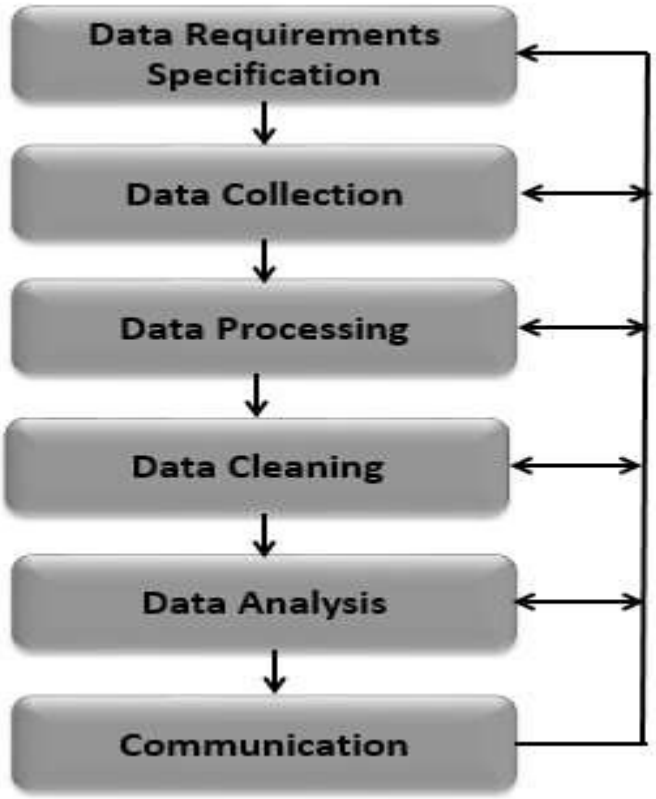
2. Bivariate Data Analysis

3. Multivariate Data Analysis

Data Analysis is a process of collecting, transforming, cleaning, and modeling data with the goal of discovering the required information. The results so obtained are communicated, suggesting conclusions, and supporting decision-making. Data visualization is at times used to portray the data for the ease of discovering the useful patterns in the data. The terms Data Modeling and Data Analysis mean the same.

Data Analysis Process consists of the following phases that are iterative in nature −

- Data Requirements Specification
- Data Collection
- Data Processing
- Data Cleaning
- Data Analysis
- Communication



Data Requirements Specification

The data required for analysis is based on a question or an experiment. Based on the requirements of those directing the analysis, the data necessary as inputs to the analysis is identified (e.g.,

Population of people). Specific variables regarding a population (e.g., Age and Income) may be specified and obtained. Data may be numerical or categorical.

Data Collection

Data Collection is the process of gathering information on targeted variables identified as data requirements. The emphasis is on ensuring accurate and honest collection of data. Data Collection ensures that data gathered is accurate such that the related decisions are valid. Data Collection provides both a baseline to measure and a target to improve.

Data is collected from various sources ranging from organizational databases to the information in web pages. The data thus obtained, may not be structured and may contain irrelevant information. Hence, the collected data is required to be subjected to Data Processing and Data Cleaning.

Data Processing

The data that is collected must be processed or organized for analysis. This includes structuring the data as required for the relevant Analysis Tools. For example, the data might have to be placed into rows and columns in a table within a Spreadsheet or Statistical Application. A Data Model might have to be created.

Data Cleaning

The processed and organized data may be incomplete, contain duplicates, or contain errors. Data Cleaning is the process of preventing and correcting these errors. There are several types of Data Cleaning that depend on the type of data. For example, while cleaning the financial data, certain totals might be compared against reliable published numbers or defined thresholds. Likewise, quantitative data methods can be used for outlier detection that would be subsequently excluded in analysis.

Data Analysis

Data that is processed, organized and cleaned would be ready for the analysis. Various data analysis techniques are available to understand, interpret, and derive conclusions based on the requirements. Data Visualization may also be used to examine the data in graphical format, to obtain additional insight regarding the messages within the data.

Statistical Data Models such as Correlation, Regression Analysis can be used to identify the relations among the data variables. These models that are descriptive of the data are helpful in simplifying analysis and communicate results.

The process might require additional Data Cleaning or additional Data Collection, and hence these activities are iterative in nature.

Communication

The results of the data analysis are to be reported in a format as required by the users to support their decisions and further action. The feedback from the users might result in additional analysis.

The data analysts can choose data visualization techniques, such as tables and charts, which help in communicating the message clearly and efficiently to the users. The analysis tools provide facility to highlight the required information with color codes and formatting in tables and charts.

5 Most Important Methods For Statistical Data Analysis

1. Mean

The arithmetic mean, more commonly known as "the average," is the sum of a list of numbers divided by the number of items on the list. The mean is useful in determining the overall trend of a data set or providing a rapid snapshot of your data. Another advantage of the mean is that it's very easy and quick to calculate.

2. Standard Deviation

The standard deviation, often represented with the Greek letter sigma, is the measure of a spread of data around the mean. A high standard deviation signifies that data is spread more widely from the mean, where a low standard deviation signals that more data align with the mean. In a portfolio of data analysis methods, the standard deviation is useful for quickly determining dispersion of data points.

3. Regression

Regression models the relationships between dependent and explanatory variables, which are usually charted on a scatterplot. The regression line also designates whether those relationships are strong or weak. Regression is commonly taught in high school or college statistics courses with applications for science or business in determining trends over time.

4. Sample Size Determination

When measuring a large data set or population, like a workforce, you don't always need to collect information from every member of that population – a sample does the job just as well. The trick is to determine the right size for a sample to be accurate. Using proportion and standard deviation methods, you are able to accurately determine the right sample size you need to make your data collection statistically significant.

5. Hypothesis Testing

Also commonly called *t* testing, hypothesis testing assesses if a certain premise is actually true for your data set or population. In data analysis and statistics, you consider the result of a hypothesis test *statistically significant* if the results couldn't have happened by random chance. Hypothesis tests are used in everything from science and research to business and economic

TESTING OF HYPOTHESIS:

Definition of Hypothesis:

Hypothesis

A hypothesis is an idea which is suggested as a possible explanation for a particular situation or condition, but which has not yet been proved to be correct.

A **statistical hypothesis**, sometimes called **confirmatory data analysis**, is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables.[1] A **statistical hypothesis test** is a method of statistical inference. Commonly, two statistical data sets are compared, or a data set obtained by sampling is compared against a synthetic data set from an idealized model. A hypothesis is proposed for the statistical relationship between the two data sets, and this is compared as an alternative to an idealized null hypothesis that proposes no relationship between two data sets. The comparison is deemed *statistically significant* if the relationship between the data sets would be an unlikely realization of the null hypothesis according to a threshold probability—the significance level. Hypothesis tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance. The process of distinguishing between the null hypothesis and the alternative hypothesis is aided by identifying two conceptual types of errors (type 1 & type 2), and by specifying parametric limits on e.g. how much type 1 error will be permitted.

What is a hypothesis test?

A hypothesis test is a statistical test that is used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population.

A hypothesis test examines two opposing hypotheses about a population: the null hypothesis and the alternative hypothesis. The null hypothesis is the statement being tested. Usually the null hypothesis is a statement of "no effect" or "no difference". The alternative hypothesis is the statement you want to be able to conclude is true.

Based on the sample data, the test determines whether to reject the null hypothesis. You use a p-value, to make the determination. If the p-value is less than or equal to the level of significance, which is a cut-off point that you define, then you can reject the null hypothesis.

A common misconception is that statistical hypothesis tests are designed to select the more likely of two hypotheses. Instead, a test will remain with the null hypothesis until there is enough evidence (data) to support the alternative hypothesis.

Examples of questions you can answer with a hypothesis test include:

- Does the mean height of undergraduate women differ from 66 inches?
- Is the standard deviation of their height equal less than 5 inches?
- Do male and female undergraduates differ in height?

Types of Hypothesis

Null and Alternative hypothesis

A hypothesis test examines two opposing hypotheses about a population: the null hypothesis and the alternative hypothesis. How you set up these hypotheses depends on what you are trying to show.

- **Null hypothesis (H_0)**
  The null hypothesis states that a population parameter is equal to a value. The null hypothesis is often an initial claim that researchers specify using previous research or knowledge.
- **Alternative Hypothesis (H_1)**
  The alternative hypothesis states that the population parameter is different than the value of the population parameter in the null hypothesis. The alternative hypothesis is what you might believe to be true or hope to prove true.

What are type I and type II errors?

When you do a hypothesis test, two types of errors are possible: type I and type II. The risks of these two errors are inversely related and determined by the level of significance and the power for the test. Therefore, you should determine which error has more severe consequences for your situation before you define their risks.

No hypothesis test is 100% certain. Because the test is based on probabilities, there is always a chance of drawing an incorrect conclusion.

Type I error

> When the null hypothesis is true and you reject it, you make a type I error. The probability of making a type I error is α, which is the level of significance you set for your hypothesis test. An α of 0.05 indicates that you are willing to accept a 5% chance that you are wrong when you reject the null hypothesis. To lower this risk, you must use a lower value for α. However, using a lower value for alpha means that you will be less likely to detect a true difference if one really exists.

Type II error

> When the null hypothesis is false and you fail to reject it, you make a type II error. The probability of making a type II error is β, which depends on the power of the test. You can decrease your risk of committing a type II error by ensuring your test has enough power. You can do this by ensuring your sample size is large enough to detect a practical difference when one truly exists.

The probability of rejecting the null hypothesis when it is false is equal to 1–β. This value is the power of the test.

	Null Hypothesis	
Decision	True	False
Fail to reject	Correct Decision (probability = 1 - α)	**Type II Error** - fail to reject the null when it is false (probability = β)
Reject	**Type I Error** - rejecting the null when it is true (probability = α)	Correct Decision (probability = 1 - β)

Example of type I and type II error

To understand the interrelationship between type I and type II error, and to determine which error has more severe consequences for your situation, consider the following example.

A medical researcher wants to compare the effectiveness of two medications. The null and alternative hypotheses are:

- Null hypothesis (H_0): $\mu_1 = \mu_2$

The two medications are equally effective.

- Alternative hypothesis (H_1): $\mu_1 \neq \mu_2$

The two medications are not equally effective.

A type I error occurs if the researcher rejects the null hypothesis and concludes that the two medications are different when, in fact, they are not. If the medications have the same effectiveness, the researcher may not consider this error too severe because the patients still benefit from the same level of effectiveness regardless of which medicine they take. However, if a type II error occurs, the researcher fails to reject the null hypothesis when it should be rejected. That is, the researcher concludes that the medications are the same when, in fact, they are different. This error is potentially life-threatening if the less-effective medication is sold to the public instead of the more effective one.

As you conduct your hypothesis tests, consider the risks of making type I and type II errors. If the consequences of making one type of error are more severe or costly than making the other type of error, then choose a level of significance and a power for the test that will reflect the relative severity of those consequences.

PARAMETRIC TESTS USED FOR TESTING OF HYPOTHESIS:

1. **Z-Test**
2. **T-Test**
3. **X^2 - Test**

1. **Z-Test**

A **z-test** is a statistical **test** used to determine whether two population means are different when the variances are known and the sample size is large. The **test**statistic is assumed to have a normal distribution, and nuisance parameters such as standard deviation should be known for an accurate **z-test** to be performed.

ANOVA, Chi-square

ANOVA (Analysis of Variance)

ANOVA is a statistical technique that assesses potential differences in a scale-level dependent variable by a nominal-level variable having 2 or more categories. For example, an ANOVA can examine potential differences in IQ scores by Country (US vs. Canada vs. Italy vs. Spain). The ANOVA, developed by Ronald Fisher in 1918, extends the *t* and the *z* test which have the problem of only allowing the nominal level variable to have two categories. This test is also called the Fisher analysis of variance.

General Purpose of ANOVA

Researchers and students use ANOVA in many ways. The use of ANOVA depends on the research design. Commonly, ANOVAs are used in three ways: one-way ANOVA, two-way ANOVA, and N-way ANOVA.

One-Way ANOVA

A one-way ANOVA has just one independent variable. For example, difference in IQ can be assessed by Country, and County can have 2, 20, or more different categories to compare.

Two-Way ANOVA

A two-way ANOVA refers to an ANOVA using two independent variables. Expanding the example above, a 2-way ANOVA can examine differences in IQ scores (the dependent variable) by Country (independent variable 1) and Gender (independent variable 2). Two-way ANOVA can be used to examine the interaction between the two independent variables. Interactions indicate that differences are not uniform across all categories of the independent variables. For example, females may have higher IQ scores overall compared to males, but this difference could be greater (or less) in European countries compared to North American countries. Two-way ANOVAs are also called factorial ANOVAs.

N-Way ANOVA

A researcher can also use more than two independent variables, and this is an n-way ANOVA (with n being the number of independent variables you have). For example, potential differences in IQ scores can be examined by Country, Gender, Age group, Ethnicity, etc, simultaneously.

Assumptions of ANOVA

Assumptions in the ANOVA Model

The one-way ANOVA model with effects coding is written

$$Y_{it}=\mu+\tau_i+\epsilon_{it}, i=1,2,...,v t=1,2,...,r_i$$

$$\epsilon_{it}\sim_{iid} N(0,\sigma_2)$$

The v treatments are indexed by i and the number of replicates receiving the i-th treatment is ri. We have so far focused our analysis on the treatment means. In particular, we constructed the ANOVA table to find a test statistic to test the null hypothesis that all treatment means are equal. We also found ways to test for pairwise differences between treatment means, and tests for estimable contrasts (linear combinations) of treatment means.

All of these tests were constructed under the assumption that $\epsilon it\sim iidN(0,\sigma 2)$. We can decompose this assumption into three parts:

Independence of Errors

We assume that each trial is independent of all other trials, except for the effect τi of the treatment on the mean. Statistical independence of two trials means that knowing the result of one trial doesn't change the distribution of the other trial. The most common causes of dependence in experimental data are confounding factors - something measured or unmeasured that affects the experiment. Randomization is a critical technique in experimental design because it can minimize the effect of any confounder.

Normality of Errors

We assume in the ANOVA model that the error terms are normally-distributed with zero mean. If the data are not normally-distributed, but instead come from some other distribution(exponential or binomial, for example), then we may not be able to trust our *p*-values, which were built by assuming normality.

Equal Error Variance Across Treatments

The next assumption is that all error terms have the same variance$\sigma 2$. It is common to see that different treatments seem to result in different means AND ALSO different variances.

Chi-square Test:

- A **chi-squared test**, also written as **test**, is any statistical hypothesis test wherein the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true. Without other qualification, 'chi-squared test' often is used as short for *Pearson's* chi-squared test. The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.

- In the standard applications of the test, the observations are classified into mutually exclusive classes, and there is some theory, or say null hypothesis, which gives the probability that any observation falls into the corresponding class. The purpose of the test is to evaluate how likely it is between the observations and the null hypothesis.

- Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance. Test statistics that follow a chi-squared distribution arise from an assumption of independent normally distributed data, which is valid in many cases due to the central limit theorem. A chi-squared test can be used to attempt rejection of the null hypothesis that the data are independent.

- Also considered a chi-squared test is a test in which this is *asymptotically* true, meaning that the sampling distribution (if the null hypothesis is true) can be made to approximate a chi-squared distribution as closely as desired by making the sample size large enough.

There are **two types of chi-square tests**. Both use the chi-square statistic and distribution for different purposes:

- A **chi-square goodness of fit test** determines if a sample data matches a population. For more details on this type, see: *Goodness of Fit Test*.
- A **chi-square test for independence** compares two variables in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of categorical variables differ from each another.
  - A **very small chi square test statistic** means that your observed data fits your expected data extremely well. In other words, there is a relationship.
  - A **very large chi square test statistic** means that the data does not fit very well. In other words, there isn't a relationship.

Assumptions of Chi-square Test

As with parametric tests, the non-parametric tests, including the χ^2 assume the data were obtained through random selection. However, it is not uncommon to find inferential statistics used when data are from convenience samples rather than random samples. (To have confidence in the results when the random sampling assumption is violated, several replication studies should be performed with essentially the same result obtained). Each non-parametric test has its own specific assumptions as well. **The assumptions of the Chi-square include:**

1. The data in the cells should be frequencies, or counts of cases rather than percentages or some other transformation of the data.

2. The levels (or categories) of the variables are mutually exclusive. That is, a particular subject fits into one and only one level of each of the variables.

3. Each subject may contribute data to one and only one cell in the χ^2. If, for example, the same subjects are tested over time such that the comparisons are of the same subjects at Time 1, Time 2, Time 3, etc., then χ^2 may not be used.

4. The study groups must be independent. This means that a different test must be used if the two groups are related. For example, a different test must be used if the researcher's data consists of paired samples, such as in studies in which a parent is paired with his or her child.

5. There are 2 variables, and both are measured as categories, usually at the nominal level. However, data may be ordinal data. Interval or ratio data that have been collapsed into ordinal categories may also be used. While Chi-square has no rule about limiting the number of cells (by limiting the number of categories for each variable), a very large number of cells (over 20) can make it difficult to meet assumption #6 below, and to interpret the meaning of the results.

6. The value of the cell *expecteds* should be 5 or more in at least 80% of the cells, and no cell should have an expected of less than one (3). This assumption is most likely to be met if the sample size equals at least the number of cells multiplied by 5. Essentially, this assumption specifies the number of cases (sample size) needed to use the χ^2 for any number of cells in that χ^2. This requirement will be fully explained in the example of the calculation of the statistic in the case study example.

Discriminant Analysis:

Discriminant analysis is a technique that is used by the researcher to analyze the research data when the criterion or the dependent variable is categorical and the predictor or the independent variable is interval in nature. The term categorical variable means that the dependent variable is divided into a number of categories. For example, three brands of computers, Computer A, Computer B and Computer C can be the categorical dependent variable.

The objective of discriminant analysis is to develop discriminant functions that are nothing but the linear combination of independent variables that will discriminate between the categories of the dependent variable in a perfect manner. It enables the researcher to examine whether significant differences exist among the groups, in terms of the predictor variables. It also evaluates the accuracy of the classification.

Discriminant analysis is described by the number of categories that is possessed by thse dependent variable.

As in statistics, everything is assumed up until infinity, so in this case, when the dependent variable has two categories, then the type used is two-group discriminant analysis. If the dependent

variable has three or more than three categories, then the type used is multiple discriminant analysis. The major distinction to the types of discriminant analysis is that for a two group, it is possible to derive only one discriminant function. On the other hand, in the case of multiple discriminant analysis, more than one discriminant function can be computed.

There are many examples that can explain when discriminant analysis fits. It can be used to know whether heavy, medium and light users of soft drinks are different in terms of their consumption of frozen foods. In the field of psychology, it can be used to differentiate between the price sensitive and non price sensitive buyers of groceries in terms of their psychological attributes or characteristics. In the field of business, it can be used to understand the characteristics or the attributes of a customer possessing store loyalty and a customer who does not have store loyalty.

For a researcher, it is important to understand the relationship of discriminant analysis with Regression and Analysis of Variance (ANOVA) which has many similarities and differences. Often we can find similarities and differences with the people we come across. Similarly, there are some similarities and differences with discriminant analysis along with two other procedures. The similarity is that the number of dependent variables is one in discriminant analysis and in the other two procedures, the number of independent variables are multiple in discriminant analysis. The difference is categorical or binary in discriminant analysis, but metric in the other two procedures. The nature of the independent variables is categorical in Analysis of Variance (ANOVA), but metric in regression and discriminant analysis.

Discriminant analysis is a statistical method that is used by researchers to help them understand the relationship between a "dependent variable" and one or more "independent variables." A dependent variable is the variable that a researcher is trying to explain or predict from the values of the independent variables. Discriminant analysis is similar to regression analysis and analysis of variance (ANOVA). The principal difference between discriminant analysis and the other two methods is with regard to the nature of the dependent variable.

Discriminant analysis requires the researcher to have measures of the dependent variable and all of the independent variables for a large number of cases. In regression analysis and ANOVA, the dependent variable must be a "continuous variable." A numeric variable indicates the degree to which a subject possesses some characteristic, so that the higher the value of the variable, the greater the level of the characteristic. A good example of a continuous variable is a person's income.

The steps involved in conducting discriminant analysis are as follows:

The problem is formulated before conducting.

- The discriminant function coefficients are estimated.
- The next step is the determination of the significance of these discriminant functions.

• One must interpret the results obtained.
• The last and the most important step is to assess the validity.

FACTOR ANALYSIS

Factor Analysis is extensively used in business research. The purpose of factor analysis in business research is to reduce the number of variables by using lesser number of surrogate variables (factors) while retaining the variability.

The primary objective is to capture some psychological states of customers/ respondents that cannot be measured directly. This is usually implemented by constructing a scale that measures attitudes, perceptions, motivations, etc of targeted customers of a specific line of business.

Initially, the targeted customers/ respondents are subjected to answer a detailed questionnaire with different lifestyle statements along with numeric or verbal scales. Through this, what really are measured are customer's views regarding few dimensions (factors) concerning the success of a business. These dimensions are psychological states that cannot be measured directly. So, factor analysis is used to assess these dimensions (factors) indirectly.

So, factor analysis is primarily used to simplify a data set before subjecting it to multivariate analysis - multiple regression, etc.

Factor analysis is a way to take a mass of data and shrinking it to a smaller data set that is more manageable and more understandable. It's a way to find hidden patterns, show how those patterns overlap and show what characteristics are seen in multiple patterns. It is also used to create a set of variables for similar items in the set (these sets of variables are called dimensions). It can be a very useful tool for complex sets of data involving psychological studies, socioeconomic status and other involved concepts. A "factor" is a set of observed variablesthat have similar response patterns; They are associated with a hidden variable (called a confounding variable) that isn't directly measured. Factors are listed according to factor loadings, or how much variation in the data they can explain.

Assumptions of Factor analysis.

Following are the assumptions that factor analysis entails:

The variables included can be condensed into one or more underlying factors

As a data reduction technique, factor analysis assumes that a research involves multiple dependent variables

The sample is homogeneous in the sense that the sub groups within the sample will not have different patterns of scores on the variables included in the analysis

The factors are related to the score of each variable in linear manner

CONJOINT ANALYSIS:

Conjoint analysis' is a survey based statistical technique used in market research that helps determine how people value different attributes (feature, function, benefits) that make up an individual product or service.

Conjoint analysis is also popularly called trade off analysis as buyers have to let go of certain product features that they consider lucrative to make a more practical purchase. For example a large number of people planning to buy a new smart phone might think that however much they want an iPhone 6, they will have to be content with a less expensive phone.

Thus we see that consumers are put in a situation where they are forced to evaluate the merit of the phone attributes such as configuration, OS, price, brand, etc. Thus, broadly conjoint analysis checks the compromises users make while selecting products or services.

Operational Development

The process of conjoint analysis is described in a simplified manner in the following steps:

1. Recognizing the product attributes: configuration, brand, price, etc in the above case.
2. Selecting the importance degree of these attributes.
3. Creating virtual products by fusing several degrees of these attributes.
4. Collecting responses through a survey.
5. Analyzing the data statistically.
6. Market simulation of the product.

Conjoint Analysis Usage

- Conjoint studies aid in **advertising**. By research the company can selected the most desirable attributes to be marketed. For example, after conjoint analysis, the company determines that brand and hardware configuration of the phone is most important to its users. It would then design advertisements that well promote these attributes and that do not focus on price which is a secondary concern in this case.

- The most common usage is **new product development**. Conjoint analysis identifies opportunities by fusing attributes to generate new products and services that are not yet in the market.

- The method is also good for **test marketing** as it provides information of the degrees of importance of each attribute. Prior to releasing the product full-fledged, it is feasible to foresee the success or failure of a product.

- Conjoint analysis is also applicable in situations where **segmentation** needs to be done. Certain clusters of users give preference to one set of attributes, whereas a different set would be more important to few others. Conjoint uncovers this pattern so that the company can target users accordingly.

Limitations of Conjoint Analysis

For certain kind of products, consumers do their evaluation built on intangible attributes or image. These products mostly comprise of luxury items where the emotional factor rather than the rational side dominates. In cases like these, the logic of conjoint analysis does not apply.

With an exception to this situation, conjoint is quite inexpensive as compared to other similar methods such as concept testing and hence is hugely popular. In a nutshell, it is a versatile and powerful tool to predict consumer choices, foresee their purchase decisions and hence design and launch products accordingly.

MULTIDIMENSIONAL SCALING:

The main function of multidimensional Scaling (MDS) is to re-project the objects (sites) in reduced dimension ordination space. As stated above, the axes scores cannot be used in subsequent analyses due to a lack of independence.

The real advantage of MDS over other ordination techniques is that it will generally represent the ordering relationships amongst objects within a given set of dimensions better than eigen-based ordinations. Whilst the eigen-based analyses rotates all of the original axes such that each axis maximizes the variance preserved by that axes, MDS constructs a configuration of objects specifically for only a nominated number of new axes. Therefore, when a 2 dimensional ordination is desired, the configuration from an MDS is optimized for just two dimensions whereas the eigen-based object configurations are independent of how many dimensions you intend to plot.

Principle components analysis (PCA) can be performed by either spectral (eigen) decomposition of an association matrix or single value decomposition of the original data matrix. Either way, it yields a *rigid rotation* of axes in that the positions of points relative to one another (euclidean distances) are maintained during rotation. The rotated axes are referred to as principle components.

Obviously, the higher the degree of associations between the variables (species), the more strongly focused and oriented the data cloud is and therefore the more 'successful' the PCA in terms of being able to represent many variables by only a few new variables.

The resulting principle components (axes) and scores (new variable values) are independent of one another and therefore the 'reduced data' can be used in subsequent statistical analyses. For example, principle component 1 (as one measure of the community) could be regressed against an environmental variable (such as altitude or rainfall) in a regular linear modelling procedure. Similarly, multiple correlated predictor variables (that otherwise violate multi-collinearity) could be reduced to one or two orthogonal predictors to use in linear modelling. Principle components analysis is therefore often an intermediate step to used as either responses or predictors in linear modeling.

MDS is mathematically and conceptually relatively simple, yet computationally very intense. Rather than have elegant mathematical solutions, MDS works on bruit grunt.

CLUSTER ANALYSIS:

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It is often necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Besides the term *clustering*, there are a number of terms with similar meanings, including *automatic classification*, *numerical taxonomy*, *botryology*(from Greek βότρυς "grape") and *typological analysis*. The subtle differences are often in the usage of the results: while in data

mining, the resulting groups are the matter of interest, in automatic classification the resulting discriminative power is of interest.

Cluster analysis is a data exploration (mining) tool for dividing a multivariate dataset into "natural" clusters (groups). We use the methods to explore whether previously undefined clusters (groups) may exist in the dataset. For instance, a marketing department may wish to use survey results to sort its customers into categories (perhaps those likely to be most receptive to buying a product, those most likely to be against buying a product, and so forth).

Cluster Analysis is used when we believe that the sample units come from an unknown number of distinct populations or sub-populations. We also assume that the sample units come from a number of distinct populations, but there is no apriori definition of those populations. Our objective is to describe those populations using the observed data.

Cluster Analysis, until relatively recently, has had very little interest. This has changed because of the interest in the bioinformatics and genome research. To explore Cluster Analysis in our lesson here, we will use an ecological example.

Applications of Cluster Analysis

- Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.
- Clustering can also help marketers discover distinct groups in their customer base. And they can characterize their customer groups based on the purchasing patterns.
- In the field of biology, it can be used to derive plant and animal taxonomies, categorize genes with similar functionalities and gain insight into structures inherent to populations.
- Clustering also helps in identification of areas of similar land use in an earth observation database. It also helps in the identification of groups of houses in a city according to house type, value, and geographic location.
- Clustering also helps in classifying documents on the web for information discovery.
- Clustering is also used in outlier detection applications such as detection of credit card fraud.
- As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

Requirements of Clustering in Data Mining

The following points throw light on why clustering is required in data mining −

- **Scalability** − We need highly scalable clustering algorithms to deal with large databases.

- **Ability to deal with different kinds of attributes** − Algorithms should be capable to be applied on any kind of data such as interval-based (numerical) data, categorical, and binary data.

- **Discovery of clusters with attribute shape** − The clustering algorithm should be capable of detecting clusters of arbitrary shape. They should not be bounded to only distance measures that tend to find spherical cluster of small sizes.

- **High dimensionality** − The clustering algorithm should not only be able to handle low-dimensional data but also the high dimensional space.

- **Ability to deal with noisy data** − Databases contain noisy, missing or erroneous data. Some algorithms are sensitive to such data and may lead to poor quality clusters.

- **Interpretability** − The clustering results should be interpretable, comprehensible, and usable.

Clustering Methods

Clustering methods can be classified into the following categories −

- Partitioning Method

- Hierarchical Method

- Density-based Method

- Grid-Based Method

- Model-Based Method

- Constraint-based Method

Partitioning Method

Suppose we are given a database of ‘n’ objects and the partitioning method constructs ‘k’ partition of data. Each partition will represent a cluster and $k \leq n$. It means that it will classify the data into k groups, which satisfy the following requirements −

- Each group contains at least one object.

- Each object must belong to exactly one group.

Points to remember −

- For a given number of partitions (say k), the partitioning method will create an initial partitioning.

- Then it uses the iterative relocation technique to improve the partitioning by moving objects from one group to other.

Hierarchical Methods

This method creates a hierarchical decomposition of the given set of data objects. We can classify hierarchical methods on the basis of how the hierarchical decomposition is formed. There are two approaches here −

- Agglomerative Approach
- Divisive Approach

Agglomerative Approach

This approach is also known as the bottom-up approach. In this, we start with each object forming a separate group. It keeps on merging the objects or groups that are close to one another. It keep on doing so until all of the groups are merged into one or until the termination condition holds.

Divisive Approach

This approach is also known as the top-down approach. In this, we start with all of the objects in the same cluster. In the continuous iteration, a cluster is split up into smaller clusters. It is down until each object in one cluster or the termination condition holds. This method is rigid, i.e., once a merging or splitting is done, it can never be undone.

Approaches to Improve Quality of Hierarchical Clustering

Here are the two approaches that are used to improve the quality of hierarchical clustering −

- Perform careful analysis of object linkages at each hierarchical partitioning.
- Integrate hierarchical agglomeration by first using a hierarchical agglomerative algorithm to group objects into micro-clusters, and then performing macro-clustering on the micro-clusters.

Density-based Method

This method is based on the notion of density. The basic idea is to continue growing the given cluster as long as the density in the neighborhood exceeds some threshold, i.e., for each data point within a given cluster, the radius of a given cluster has to contain at least a minimum number of points.

Grid-based Method

In this, the objects together form a grid. The object space is quantized into finite number of cells that form a grid structure.

Advantage

- The major advantage of this method is fast processing time.
- It is dependent only on the number of cells in each dimension in the quantized space.

Model-based methods

In this method, a model is hypothesized for each cluster to find the best fit of data for a given model. This method locates the clusters by clustering the density function. It reflects spatial distribution of the data points.

This method also provides a way to automatically determine the number of clusters based on standard statistics, taking outlier or noise into account. It therefore yields robust clustering methods.

Constraint-based Method

In this method, the clustering is performed by the incorporation of user or application-oriented constraints. A constraint refers to the user expectation or the properties of desired clustering results. Constraints provide us with an interactive way of communication with the clustering process. Constraints can be specified by the user or the application requirement.

Unit V

REPORT WRITING

Types of Reports, Business, Technical and Academic Report writing – Methodology Procedure – Contents – Bibliography

A **report** is **written** for a clear purpose and to a particular audience. Specific information and evidence are presented, analysed and applied to a particular problem or issue. ... When you are asked to **write** a **report** you will usually be given a **report** brief which provides you with instructions and guidelines.

Report Writing is a well-organized writing process that needs enough skill, research, and details.

This report writing process is usually time-consuming and often requires a detailed research for which the report writer may have to visit places, meet people and find a solid conclusion at the end of the day to write a good report.

Use of report writing is to explain a matter or issue and inform higher authorities to help them take a right decision or action in regards to the matter or issue.

There is a little room for creativity in the report writing structure as the main focus is to enlighten the readers about a matter or issue and make the whole thing self-explanatory for easy understanding.

Report Writing is the primary tool of media personnel through which they pass on specific information about an incident or topic.

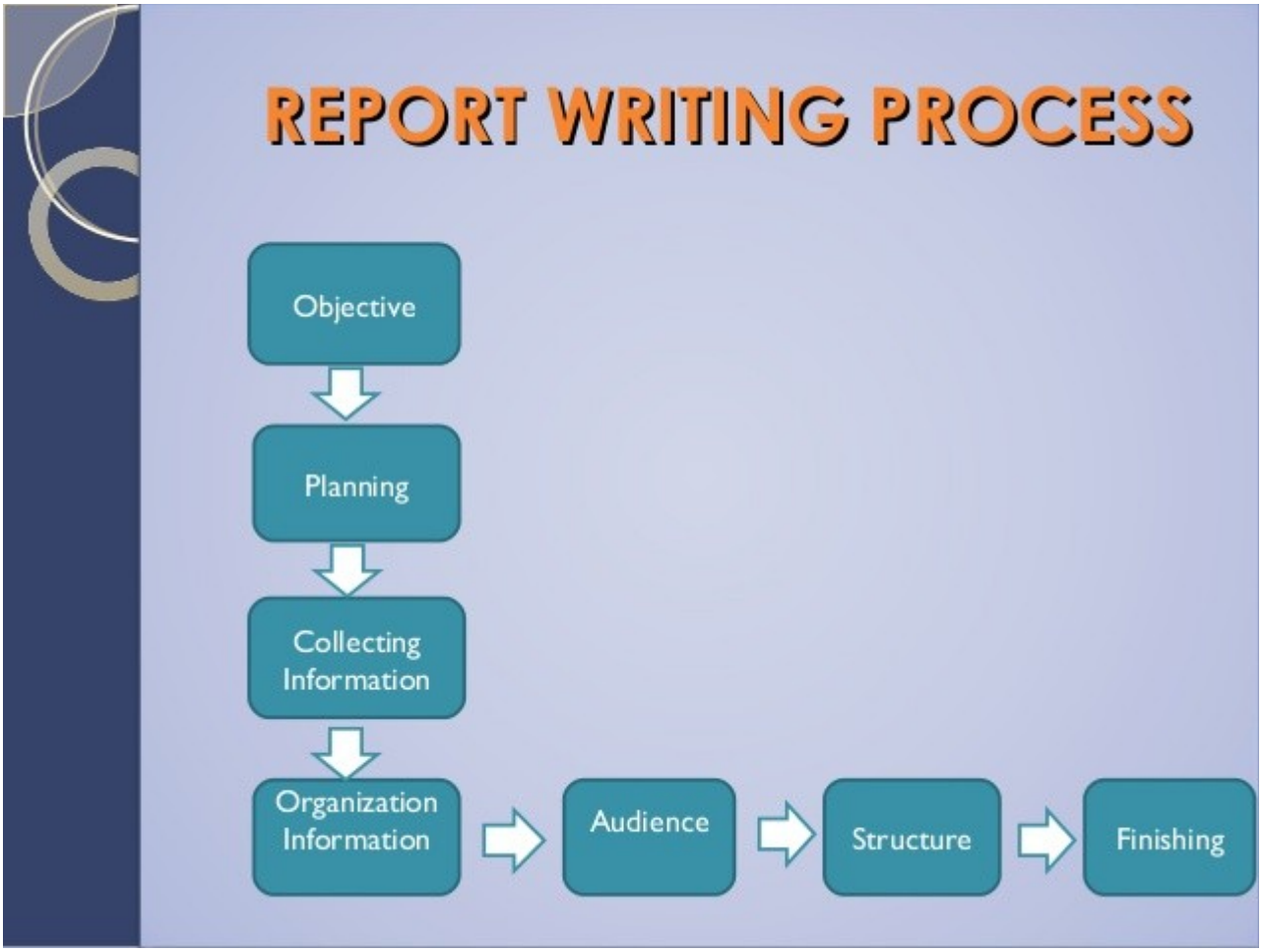
Apart from the media usage, report writing is required in various sectors like corporate, government, politics etc.

To offer a better understanding, learning about the essence of this form of writing – 'The Report' is very important.

What is a report?

A report is a methodical, well planned document which outlines and evaluates a subject or problem, and which may include:

- The record of an order of events
- Explanation of the implication of these events or facts
- Evaluation of the facts or results of research presented
- Discussion of the consequences of a conclusion or course of action
- Conclusions
- References



Reports must always be:

- Correct
- Crisp
- Clear
- Well-Structured

Report Writing Format

Every organization has a predefined report writing format and hence a writer can't remain stuck to a specific format for different reports.

While writing a formal report, the freelance writer must keep in mind that the target readers need don't have to do unnecessary research to take a decision or action after reading the report.

Types of Report Writing

Everything must be in details. There are many types of report writing for organizations that are used for various purposes. To format your report to serve varied needs, have a look at the primary report writing types mentioned below.

Formal Report Writing

Formal Report Writing is pretty complex and time-consuming. Usually, it demands an immense research, explanation, references, links, lists and many other things to make the primary point clear enough for the readers.

This type of report writing is usually preferred for an important incident, issue or matter by big organizations. Formal report writing is generally long and expensive.

Formal Report writing has an internationally accepted pattern that includes various components that are mentioned below.

Cover – The cover of a report is something that drives a reader's attention first. The report's cover leaves a huge impression on the reader's mind and he/she can get an idea regarding the report's topic or quality through the cover design.

There is a saying – ''don't judge a book by its cover'' but it is also true – what seems well, sells well and hence the report writer must take care of the cover of the report with has a major role to depict the report brief.

Title Of The Report – This component includes the report's title and the name of the writer. Apart from these things, the title can also have a date and the name of the organization for which the report has been prepared.

The cover also has these things but putting them in details in the title section is mandatory.

Table Of Contents – This section includes headings and subheadings of the primary text written. This is a very important portion of report writing. It helps your readers to reach desired sections in your report in a hassle-free manner.

Summary – Here you basically provide the synopsis of the whole report's primary text and you can also call it an informative summary.

Many times, it is referred as 'executive summary'. You can also use descriptive summary which is a simple table of contents. The format is always decided by the organization.

Introduction – This is the most important section of the main text. The main text always includes three components – introduction, discussion, and conclusion.

Here, you introduce the main text of your report in the most intriguing a detailed manner so that all types of readers can get your point without much effort.

Discussion – In the discussion section, a report writer discusses the main story of the report. According to your reader's convenience, you decide the order of the report's results.

You can also do a result to theory comparison here along with the analysis, evaluation and interpretation of the data included.

Conclusion/Recommendation – You can present the summary of the discussion section here. Here, you mention your findings and recommend the elements to your readers as per your overall evaluation.

Appendix – In the appendix portion, you can attach the graphs, lists, survey and suchlike stuff that are related to your report and helps your readers to understand the report comprehensively.

Informal Report Writing

This type of report writing is comparatively easier and less time-consuming than the formal report writing. Here, you need to perform lesser research and it also includes lesser components.

The basic components of informal report writing include – Introduction, Discussion & Reference/Recommendations. Different organizations include more components to this type as per their requirements.

Informal Report Writing can also be divided into few types – Credit Report, Feasibility Report, Progress Report, Sales Activity Report, *Financial Report*, Personal Evaluation and Literary Report.

All the above-mentioned information in regards to Report Writing and its types must have helped you, but to understand the process in a better manner, here is a seven step process for report writing that can help any report writer to master the report writing technique.

Best Process For Report Writing

To bring up an effective report, the right process has to be followed. Here are seven steps to undertake best report writing process.

Decide the Objective:

Like any other process, report writing sets its base on the purpose, why a report is being created. With a clear objective ahead, it helps report writer to stay focused and produce quality report that is easier to engage the reader.

Understand Your Audience:

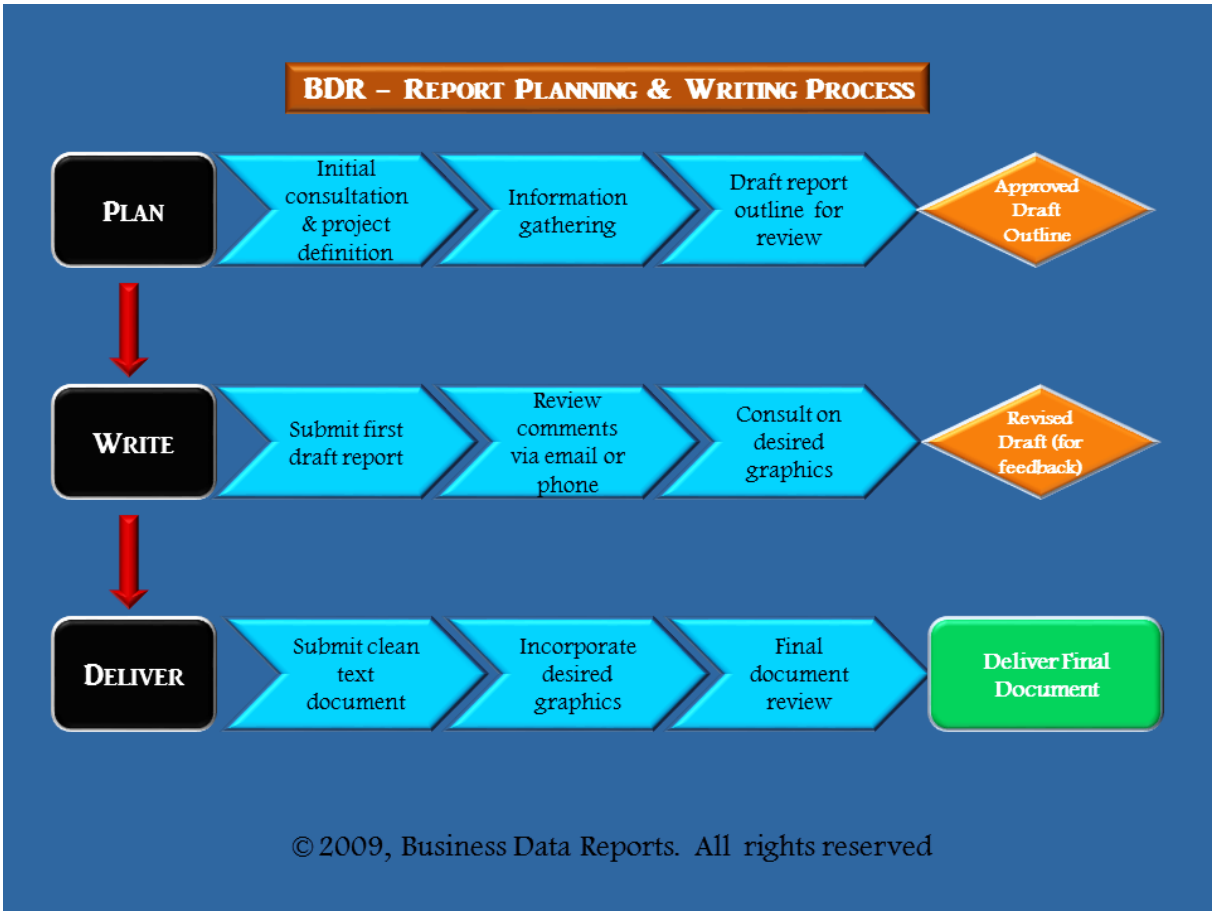
Right understanding of the audience definitely leads to a quality report. For example, an annual financial report for stakeholders is completely different from a financial review.

And with an understanding, a writer can alter the use of language, data incorporated and supporting material that can uplift the indulgence for the set audience.

Having a personal touch as per the audiences' preference can help produce ideas based on their choice. With an understanding writer can present the report that suits their preference.

Report Format and Types

For a report to be an effective communication tool, it must follow a particular format or type. Deciding on parameters like, written report or presentation; what type – formal, informal, financial, annual, technical, fact-finding or problem-solving report; design templates if any available.



Collect the Facts and Data

Adding figures, facts and data adds credibility to the report and strengthens the argument. Adding data or facts brings along a crucial responsibility to cite or mention the sources, like interviews, articles, sayings, articles, etc.

Structure the Report:

A report typically has four elements, the executive summary (this is written after the report is finished), introduction (this includes the structure of the report and table of contents), body (main text and report is occupied in this portion), conclusion (this is a binding portion that brings all the elements of the report into a systematical end).For a better understanding on the structuring of a report, *read The Structure of Reports by UniLearning*.

Readability

The readability part is definitely a crucial aspect as it becomes a must to make the report enjoyable and accessible to read. A great navigation is the best way to make the reader take a uniformed path through the information flow.

Adding proper formatting (h1, h2, h3…), graphics/visuals, break up of long text into shorter for better read and giving the text in report writing structure with bullet points for better understanding.

Edit

The initial draft of report writing is never perfect (at least 90% of times). This calls for edition and revision of the content.

Best practice can be keeping aside the report document for few days and then once more start working over it again or ask a fellow member to review or proofread it for you.

Also while you edit or get it edited, make sure you keep writing samples handy. These samples can act as a guideline.

Example & Sample of Reports

After attaining all the understanding about writing reports, it's time to look at some actual examples of report writing.

Report Writing Example

To grab a better and practical understanding, let's have a look at one of the examples of report writing format here.

Report Writing Samples

There are many more report writing samples available over the web that can give a better idea of how to bring up a perfect report.

Though, writing a report is often considered as a complicated process, but if you lack the expertise in writing your report then outsourcing the writing process can be the best available option.

These hired professionals will offer you a report brief before they submit you the final version.

They will help you with writing reports to perfection by offering a draft and a final copy that is as per your need. A sample report before the final copy lets you achieve an effective report.

They are well-versed to write reports and essential elements that are important for a good report like, how to format your report, how to present the content, etc.

This the sample report mentioned above, aspiring as well as established report writers can make the best out of report writing.

Final Words –

Report writing is the best mean to end a project and document the knowledge or information for the next upcoming project.

These above seven steps offer a systematic and logical way to process the text into a report and develop a report that faultlessly solves its purpose.

Technical Writing

Technical writing is closely related to the process, or "how to," essay. Its function is to clearly explain the steps to accomplishing a task so anyone can do it. As such, its audience is the average layperson. Many technical writing assignments are reminiscent of instruction manuals; this is because manual writing is one type of technical writing. Other types would include business letters, memos, product descriptions, warning labels and, to some extent, editorial letters.

Academic Writing:

Academic writing is by definition more complex because most academic writing is tied to a specific discipline or field, which means it can become jargon-laden. This occurs even in the business and computer sectors, two areas known for their emphasis on technical writing. Scholarly articles and textbooks in either of those disciplines will contain as much jargon and as many complex ideas as the most challenging essay in literature, art or music. This is to be expected because all fields have both specialized knowledge and specialized terminology.

Key Difference – Academic vs. Technical Writing

Academic and technical writing are two forms of writing between which a key difference can be identified. Most people assume that a technical writer is, in fact, an academic writer as well. This, however, is a false assumption. Although both academic writing and technical writing require excellent writing skills, the **key differences** between these two types of writing are **the audience and purpose of writing**. Academic writing is a form of writing that is used in academic disciplines. On the other hand, technical writing is a form of writing that is mostly used in technical disciplines. As you can see, the contexts of the two forms of writing differ from one another. Also, the target audience for academic writing is mostly scholars, but in the case of technical writing this is

not the case. Even a lay person can be the target audience. Through this article let us examine the differences between academic and technical writing.

What is Academic Writing?

Academic writing is **a form of writing that is used in academic disciplines**. This includes both the natural sciences as well as the social sciences. Scholars use academic writing for many reasons. They can use it to present the findings of a new research that they conducted or even to present a new point of view. The target audience of academic writing is usually scholars that belong to a particular discipline.

For academic writing, the writer uses a special jargon. If you go through journal articles, research papers, dissertations, you will notice that not only the jargon but even the style of writing is quite different to what we see every day because the style is very impersonal. You can also notice inter-textuality, or else the quoting of previous works to support or oppose certain arguments. Developing the ability to write academic articles is not an easy task, it requires an extensive knowledge of the subject as well as excellent writing skills.

Preparation of reports:

For preparations or drafting of reports three steps are involved.

1) First Draft:

Comprehensiveness or fullness of facts.

Precision or Accuracy of Facts

Coherence or logic of facts, and

Movement or transition of facts and ideas.

2nd. The researcher should give the first draft, at this stage, a shape so that it can be readable, clear and lucid. Considerable trimming or editing will have to be done to make the writing precise, Concise and brief. Finally, at the second draft stage, critical evaluation will have to be made of all that has been written-facts, findings, conclusions and recommendations.

3rd. And final The final stage in drafting is the preparations of final report. It concentrates mainly on the finish and final touches, i.e., on documentation and polish to make the report weighty, authoritative, convincing and attractive. Documentation indicated the references to the sources, other previous and current work and view, additional data and discussion and suggested further reading on the specific problem as handled by the researcher. In other words, it indicated the thoroughness of the investigation and on the other a guide to further work

Characteristics of good report

1) Attractive

2) Clear Topic

3) Balanced Language

4) No repetition of facts

5) Statement of scientific facts

6) Practicability

7) Description of the difficulties and the shortcomings

CONTENTS OF A RESEARCH REPORT

A research report ordinarily includes the following sections:

ABSTRACT. The abstract is often required to be no more than a given maximum number of words, usually between 100 and 150. It should describe the most important aspects of the study, including the problem investigated, the type of subjects (sample) and data collection method involved, the analytical procedures used, and the major results and conclusions.

INTRODUCTION. This section includes discussions concerning the practical and/or theoretical importance of the topic as well as a description of the research problem. It often starts by introducing the reader to the topic and making a case for the practical significance of the issues being investigated and/or the contribution that the study could make to our understanding of the phenomenon. The statement of research problem(s) is intended to indicate what the general purpose of the study is. This is often done through broadly stated questions or statements regarding whether and how the research variables are (or are expected to be) related to (or affected by) one another. Finally, the introduction section should include the working definitions of those terms used in the study that do not have a commonly known meaning or for which several meanings may be used.

REVIEW OF THE LITERATURE AND THE RESEARCH MODEL. Some authors present the material included here as a separate section under its own heading (as it is shown here), while others present it as part of a longer INTRODUCTION section. In either case, the review of the related literature describes and analyzes the published studies that are directly related to, and/or have some relevance to, the topic and research questions at hand. Related literature should be integrated with, and weaved into, the material in this section and not be simply cataloged. The review could conclude with a brief summary of the literature and its implications.

The study's theoretical/conceptual model and its hypotheses are developed based on the researcher's logical reasoning as well as the implications of his/her literature review. Note that the study's hypotheses should be stated in a language consistent with its proposed conceptual framework and the literature review; they should not be stated in the null and alternate hypotheses forms. Also, a well-developed hypothesis is testable; that is, it can be confirmed or disconfirmed through the collection and analysis of data.

METHODOLOGY. The methodology section includes a description of the research sample (subjects), data collection method, measurement instruments, and data analysis procedures. The description of sample/subjects includes not only the sample size and statistics regarding the subjects but also a definition and description of the population from which the sample was selected. This section also describes the method used in selecting the sample or samples. In the case of questionnaire surveys, information on response rates also should be provided.

The description of instruments should identify and briefly describe all instruments used to collect data pertinent to the study, be they tests, questionnaires, interview or observation forms, or unobtrusive data such as absenteeism reports or productivity figures. When possible, information on validity and reliability of the measures used should be reported. Also, sources should be cited for measurement instruments/procedures (e.g., scales) developed by other researchers. The method section is usually concluded with a few statements about the analysis procedures utilized to test the study's hypotheses.

RESULTS & DISCUSSION. Some authors use a single section to both present and discuss the data analysis results. Others deal with the two issues in two separate sections. In either case, the statistical techniquesthat were applied to the data must be mentioned and the results of each analysis summarized, tabulated, and then discussed. For each research hypothesis, the statistical test of significance selected and applied to the data is briefly described, followed by a statement indicating whether the hypothesis was supported or not supported. Tables and figures are used to present analyses results in summary and/or graph form and to add clarity to the presentation. Good tables and figures are uncluttered, self-explanatory, and non-redundant.

In addition to simply presenting the results in a straightforward manner, the author also has to provide the readers with his/her interpretation of the results, implications of the findings, conclusions and recommendations. Each result is discussed in terms of the original hypothesis to which it relates and in terms of its agreement or disagreement with results obtained by other researchers in similar/related studies. If the results are consistent with the theoretical model, researcher's expectations, and/or findings of other researchers, explanations must be provided as to what the results mean and what their theoretical and practical implications are. When the results do not support the hypotheses and/or contradict previous findings, not only their meaning, but also possible reasons for the discrepancies must be discussed.

Often during a study apparent and/or interesting relationships will be noticed that were not hypothesized by the researcher. These unforeseen results should be acknowledged and discussed. Such results often form the basis for future studies specifically designed to examine the issue more carefully.

Finally, the researcher should address the study's limitations and make recommendations for future research. It is notable that in the discussion portion of this section the researcher is often permitted

more freedom to express opinions and reasonable speculations/assertions that may be rather indirectly and implicitly based on data analysis results.

SUMMARY AND CONCLUSIONS. This section is very similar to the abstract section except that it appears at the end of the report (preceding the REFERENCE section). It summarizes the study's findings in an easy to understand manner. It also explains the practical implications of those findings, and points to recommended directions for future research in that area.

REFERENCES. The references section, or bibliography, lists all the sources, alphabetically by authors' last names, that were directly used in writing the report. Every source cited in the paper must be included in the references, and every entry listed in the references must appear in the paper. Style manuals, such as the APA (American Psychological Association) manual, will give you the correct procedure for all in-text and reference citations. This form is usually different for books, journal articles, and magazine articles. It is recommended that you use the APA style. It is important that whatever form is used be followed consistently.

APPENDIXES. Appendixes include information and data pertinent to the study that either are not important enough to be included in the main body of the report or are too lengthy. Appendixes contain such entries as materials especially developed for the study (e.g., tests, questionnaires, and cover letters), coding scheme, print out of raw data, and the computer print-out of statistical analyses.

Bibliography :

A bibliography can be defined as,

"A list of reference materials (involving any kind of content ; text, music, paintings, video etc.) elucidating the type, nature and other detailed information on the basis of name, date, place and genre of the materials."
OR
"A complete categorical compilation of any type of content based on it's creator(s), editors and time (of production, distribution)."

Bibliography, also known as **works cited, reference list** is basically an orderly study and referencing of books and source materials used in academic research. It might or might not include any information on the literary analysis or criticism of the materials cited.

Importance and Use of Bibliographies:

A mandatory requirement of copyright laws and academic conventions is that whenever a research paper is written, there should be a section at the end of it where you acknowledge the sources used.

So, bibliography means listing all the sources which you have consulted while writing your essay or research article. The sources may be in the form of printed and online books, websites, web

documents, web blogs, newspaper articles, journals, pod casts, wikis, unpublished material, maps etc.

Citation ensures that the information contained in the research paper is based on logic, truth and facts. Absence of references or bibliography indicates that the paper may be a piece of plagiarism.

*****