SASURIE COLLEGE OF ENGINEERING

DEPARTMENT OF MASTER OF BUSINESS ADMINISTRATION

REGULATION 2021

I YEAR – II SEMESTER

– BA4205

BUSINESS RESEARCH METHODS

BA4205

BUSINESS RESEARCH METHODS

UNIT I INTRODUCTION 9

Business Research – Definition and Significance – the research process – Types of Research – Exploratory and causal Research – Theoretical and empirical Research – Cross –Sectional and time – series Research – Research questions / Problems – Research objectives – Research hypotheses – characteristics – Research in an evolutionary perspective – the role of theory in research.

UNIT II RESEARCH DESIGN AND MEASUREMENT 9

Research design – Definition – types of research design – exploratory and causal research design – Descriptive and experimental design – different types of experimental design – Validity of findings – internal and external validity – Variables in Research – Measurement and scaling – Different scales – Construction of instrument – Validity and Reliability of instrument.

UNIT III DATA COLLECTION 9

Types of data – Primary Vs Secondary data – Methods of primary data collection – Survey Vs Observation – Experiments – Construction of questionaire and instrument – Validation of questionaire – Sampling plan – Sample size – determinants optimal sample size – sampling techniques – Probability Vs Non–probability sampling methods.

UNIT IV DATA PREPARATION AND ANALYSIS 9

Data Preparation – editing – Coding –Data entry – Validity of data – Qualitative Vs Quantitative data analyses – Bivariate and Multivariate statistical techniques – Factor analysis – Discriminant analysis – cluster analysis – multiple regression and correlation – multidimensional scaling – Application of statistical software for data analysis.

UNIT V REPORT DESIGN, WRITING AND ETHICS IN BUSINESS RESEARCH 9

Research report – Different types – Contents of report – need of executive summary – chapterization – contents of chapter – report writing – the role of audience – readability – comprehension – tone – final proof – report format – title of the report – ethics in research – ethical behaviour of research – subjectivity and objectivity in research.

TOTAL: 45 PERIODS

TEXT BOOKS

1. Donald R. Cooper, Pamela S. Schindler and J K Sharma, Business Research methods, 11th Edition, Tata Mc Graw Hill, New Delhi, 2012.
2. Alan Bryman and Emma Bell, Business Research methods, 3rd Edition, Oxford University Press, New Delhi, 2011.
3. Uma Sekaran and Roger Bougie, Research methods for Business, 5th Edition, Wiley India, New Delhi, 2012.
4. William G Zikmund, Barry J Babin, Jon C.Carr, Atanu Adhikari,Mitch Griffin, Business Research methods, A South Asian Perspective, 8th Edition, Cengage Learning, New Delhi,2012.

BUSINESS RESEARCH METHODS

CHAPTER I

INTRODUCTION

1.1.1 Meaning Of Research

It refers to a search for knowledge.
Research is an art of scientific investigation.

The Advanced Learner"s Dictionary of Current English lays down the meaning of research as,"a careful investigation or inquiry specially through search for new facts in any branch of knowledge".

Research is required because of the following reasons:

i) To identify and find solutions to the problems.
ii) To help making decisions.
iii) To develop New concepts
iv) To find alternative strategies.

1.1.2 Definition:

Redman and Mory define research as a," Systematized effort to gain new knowledge". Some people consider research as a movement, a movement from the known to the unknown.

According toClifford woody, research comprises defining and redefining problems, formulating hypothesis or suggested solutions collecting, organizing and evaluating data, making deductions and reaching conclusions; to determine whether they fit the formulating hypothesis.

1.1.3 Characteristics of Research

- Research is directed towards the *solution of a problem*
- Research is based upon *observable experience* or *empirical (practical/experimental)evidence*
- Research demands accurate *observation* and *description (Explanation)*
- Research involves gathering *new data* from primary sources or using *existing data for a new purpose.*
- Research activities are characterized by *carefully designed procedures .*
- Research requires *expertise* i.e., skill necessary to carryout investigation, search the related literature and to understand and analyze the data gathered.
- Research is *objective and logical* – applying every possible test to validate the data collected and conclusions reached.
- Research involves the quest for *answers to unsolved problems* . Research requires courage.
- Research is characterized by patient(longterm) *and unhurried(planned) activity.* Research is carefully recorded and reported.

1.1.4 Qualities Of A Good Research

- Systematic
- Logical
- Empirical
- Replicable
- Creative
- Use of multiple methods

1.2 Scope / Significance Of Research

- Identify alternative courses of action
- Helps in economic use of resources
- Helps in project identification
- Solves investment problems
- Solves pricing problems
- Solves allocation problems
- Solves decision making issues in HR Solves various operational and planning problems of business and industry.
- Provides the basis for all government policies in our economic system. Helps social scientists in studying social relationships and in seeking answers to various social problems. For students, research means a careerism or a way to attain a high position in the social structure. For professionals in research, it may mean a source of livelihood.
- For philosophers and thinkers, research means the outlet for new ideas and insights. For literary men and women, research means development of new styles and creative work. For analysts and intellectuals, research means generalizations of new theories.

1.3 Research Process:

Research Process consists of series of actions or steps necessary to effectively carry out research. Research process involves gathering data,use statistyical techniques,Interpretations, and drawing conclusions about the research data. Various steps involved in a research process are not mutually exclusive. Nor are they separate and distinct. They do not necessarily follow each other in any specific order and the researcher has to be constantly anticipating at each step in the research process.

Steps in Research Process:

i. Formulating the research problem
ii. Extensive literature survey
iii. Development of working hypothesis
iv. Preparing the research design
v. Determining sample design
vi. Collecting the data
vii. Execution of the project
viii. Analysis of data
ix. Hypothesis-testing
x. Generalization and Interpretation
xi. Preparation of the Report

1. Formulating the research problem :

The first step in research is defining a research problem." A problem well defined is half solved".Poorly defined problems cause confusion.

Two types of problems:

- Problems which related to state of nature Problems which relate to relationships between variables
- The formulation of a general topic into a specific research problem is the first step in scientific enquiry

Two steps in formulating the research problem:

- Understanding the problem thoroughly
- Rephrasing the same into meaningful terms from an analytical point of view.

2.Extensive literature survey:

Once the problem is formulated, the next step is to write down a brief summary. For this the researcher should undertake extensive literature survey connected with the problem.

- Review Abstracting/Indexing journals Published/Unpublished bibliographies Academic journals Conference proceedings Govt. Reports Books

3.Development of working hypothesis:

- Working Hypothesis is a tentative assumption made in order to draw out and test its logical or empirical assumptions Hypothesis is the focal point of the research, for ex: "students who receive counseling will show a greater increase in creativity than students not receiving counseling" or "car A is performing as well as car B" .
- Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested.

4.Preparing the research design:

- It is a conceptual structure within which the research would be conducted.
- The primary objective of the research design is to collect the relevant data.
- Research Purposes may be grouped into Exploration Description Diagnosis Experimentation many research designs exist.
- The function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money.

5. Determining sample design:

The researcher must decide the way of selecting a "Sample" or the Sample Design.

Sample design is a definite plan determined before any data is actually collected for obtaining a sample from a given population.

Samples can be either probability samples or non probability samples.

- All the items under consideration in any field constitute a "Universe" or "Population"
- A complete enumeration of all the items in the "population" is known as a "census enquiry" Since a complete census enquiry is not possible generally,
- We select a 'sample' – a few items from the "universe" for our study .
- Researcher selects the sample by using 'sampling design' – a definite plan determined before any data is actually collected.

Types of Sampling:

- Deliberate Sampling
- Simple Random Sampling
- Systematic Sampling
- Quota Sampling
- Stratified Sampling
- Cluster/area Sampling
- Multi-stage Sampling
- Sequential Sampling

6.Collecting the data:

The next step is to determine the sources of data to be used. The researcher has to decide whether he has to collect primary data or depend exclusively on secondary data.

- ✯ Primary data which are collected a fresh and for the first time,and thus happen to be original in character.
- ✯ The secondary data,on the other hand,are those which have already been collected by someone else and which have been passed through the statistical process.
- ✯ Primary data can be collected thru *experiment* or *survey.*
- ✯ In experiment, he observes some quantitative measurements (data), with which the *hypothesis is tested*
- ✯ In Survey, data can be collected by the following methods:
- ✯ Observation
- ✯ Personal Interview
- ✯ Telephone Interview
- ✯ Mailing Questionnaires
- ✯ Through Schedules

7.Execution of the project:

- ☺ The research study must be executed in a systematic manner to ensure that adequate and dependable data are collected. Should be rigorously methodological. If the data are to be collected through Interviews, arrangements should be made for proper selection and training of the interviewers.

8.Analysis of data:

After the data have been collected, the researcher turns to the task of analyzing them.

- ✓ Analysis of Data Requires that the data be necessarily condensed into manageable groups and tables for further analyses.
- ✓ Should classify the new data into some purposeful and usable categories Coding is done at this stage.
- ✓ Tabulation – classified data are put into tables Analysis, after tabulation is based on the computation of various percentages, coefficients, etc.

9.Hypothesis-testing:

After analyzing the data the researcher has to test the hypothesis,various tests such as chi square test,T test,F test,have been developed statistician for the purpose. The hypothesis may be tested through the use of one or more of such tests,depending upon the nature and object of research enquiry.

- ❖ Do the data support the hypothesis or they contrary? Chi Square test, t-test, f-test are normally used.
- ❖ Hypothesis testing will result in either accepting the hypothesis or in rejecting it .

10.Generalization and Interpretation:

If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalization, ie. To build a theory.

If the researcher had no hypothesis to start with,he might seek to explain his findings on the basis of some theory. It is known as Interpretation.

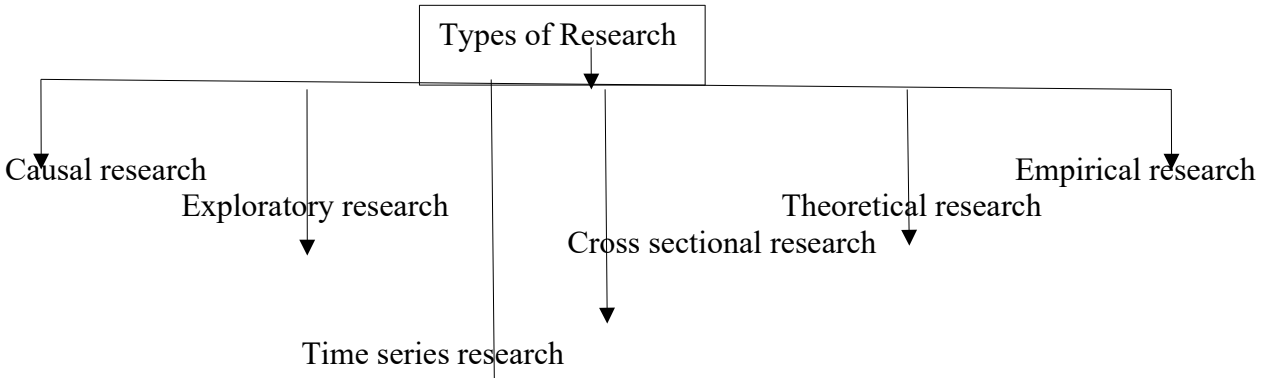
11.Preparation of Report/Thesis:
The researcher should follow the mqin principles of writing a report.

The ***layout of the report*** is as follows:

i. ***The prefatory part*** -Title page, Certification, Acknowledgments ,Preface, Contents page
ii. ***The Main Body/Text*** – Introduction, Summary of Findings, Main Report, conclusion
iii. ***The Supplemental Part :-*** References, or Bibliography Appendices Index

1.4 Types Of Research:

Research comprises of the search for knowledge through objective and systematic method of finding a solution to a problem.



1.4.1 Exploratory research:

- Exploratory research is research conducted for a problem that has not been clearly defined. It often occurs before we know enough to make conceptual distinctions or posit an explanatory relationship.

Exploratory research helps determine the best research design, data collection method and selection of subjects.

Exploratory research or formulative research: The objective of exploratory research is to gather preliminary information that will help define problems and suggest hypotheses.

Exploratory research often relies on such as reviewing available literature and/or data, or qualitative approaches such as informal discussions with consumers, employees, management or competitors, and more formal approaches through in-depth interviews, focus groups, projective methods, case studies or pilot studies. The Internet allows for research methods that are more interactive in nature.

- **Advantages of Exploratory research:**
  - It increase a researcher's understanding of a subject.
  - Offers flexibility of sources employed.
  - It can save a great deal of Time and money by diminishing dead ends early.

Disadvantages of Exploratory Research:

- It usually costs a lot.
- It may be unsuccessful.
- It is not typically generalisable to population at large.
- It can be quite informal,relying on secondary research such as reviewing available literature or data.

1.4.2 Descriptive Research:

Descriptive research: The objective of descriptive research is to describe things, such as the market potential for a product or the demographics and attitudes of consumers who buy the product.

The next step is **descriptive research**, defined as attempts to explore and explain while providing additional information about a topic. This is where research is trying to describe what is happening in more detail, filling in the missing parts and expanding our understanding.

This is also where as much information is collected as possible instead of making guesses or elaborate models to predict the future - the 'what' and 'how,' rather than the 'why.'

Advantages of Descriptive Research:

- It provides multifaceted approach for data collection.
- It provides an insight into life experiences in a way that other research methods can not.
- It is less expensive and time consuming.
- It collects a large amount of data for detailed study.

Disadvantages of Descriptive Research:

- It requires more skills.
- It doesn't identify the cause behind a phenomenon.
- Response rate is low in this research.
- Confidentiality is the primary weakness of descriptive research.
- It also presents the possibility for error and subjectivity.

1.4.3 Applied research:

Applied research in administration is often exploratory because there is need for flexibility in approaching the problem. In addition there are often data limitations and a need to make a decision within a short time period. Qualitative research methods such as case study or field research are often used in exploratory research.

There are three types of objectives in a marketing research project:

- Exploratory research or formulative research
- Descriptive research
- Causal research (also referred to as explanatory research)

1.4.4 Causal research: The objective of causal research is to test hypotheses about cause-and-effect relationships. If the objective is to determine which variable might be causing a certain behavior, i.e. whether there is a cause and effect relationship between variables, causal research must be undertaken. In order to determine causality, it is important to hold the variable that is assumed to cause the change in the other variable(s) constant and then measure the changes in the other variable(s). This type of research is very complex and the researcher can never be completely certain that there are not other factors influencing the causal relationship, especially when dealing with people's attitudes and motivations. There are often much deeper psychological considerations, that even the respondent may not be aware of this is not true.

There are two research methods for exploring the cause and effect relationship between variables:

1. Experimentation, and
2. Simulation

Features of Cause and effect method of research:

- ✓ It is primarily possible in areas of physical sciences,with the help of hypothesis,may also be carried out in social sciences.
- ✓ It is used to obtain evidence of Cause and Effect relationships.

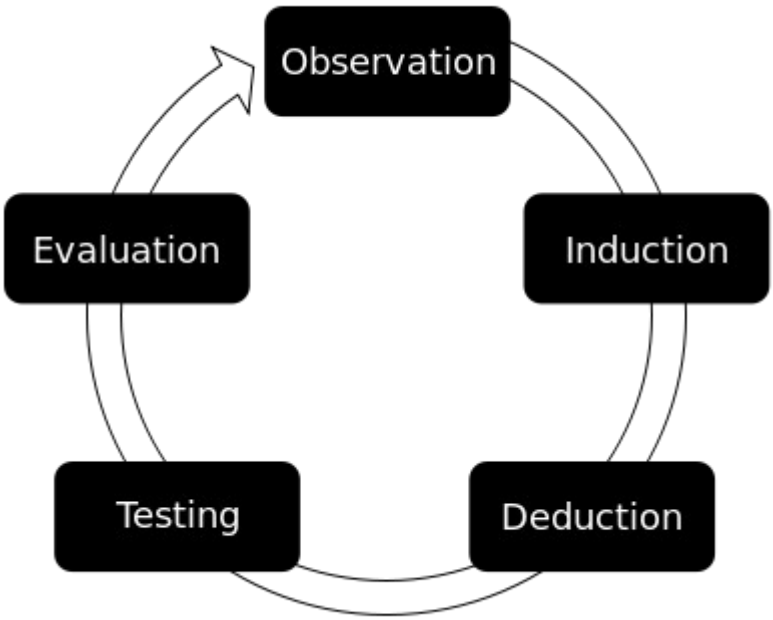
  Advantages of Causal Research:
  - Helps to control over extraneous variables which is usually lesser in other research methods

- Involves manipulating the independent variable to observe the effect on the dependent variable.

Disadvantages of Casual Research:

- ❖ Elimination of extraneous variables is not always possible.
- ❖ Experimental situation may not relate to the real world.
- ❖ It may be unethical or impossible to randomly assign people to groups.
- ❖ Possibility for error and subjectivity also exists.

1.4.5 Empirical research is a way of gaining knowledge by means of direct and indirect observation or experience. Empirical evidence (the record of one's direct observations or experiences) can be analyzed quantitatively or qualitatively.



Empirical research normally starts with some *a priori* theory, which the researcher develops to try to explain and/or predict what happens in the real world. The purpose of the research is to test the theory and possibly refine it. In some cases, research is conducted to develop theory (the *grounded theory* approach), but this is much harder (and more risky) to do.

Advantages of Empirical Research:

- ✓ Helps to understand and respond to dynamics of situations.
- ✓ Meet accepted professional standards of research.
- ✓ Integrate professional knowledge with empirical data to inform instructional development decisions.
- ✓ Establish relationship between intervention and behavioural response.

Disadvantages of Empirical Research:

- o They cannot gain access to the typ[es of firms necessary for their studies.
- o They do not possess the requisite skills necessary to design the empirically based studies.to gather and analyse the data efficiently.

1.4.6 Theoritical Research:

Theories are formulated to explain, predict, and understand phenomena and, in many cases, to challenge and extend existing knowledge, within the limits of the critical bounding assumptions. The **theoretical** framework is the structure that can hold or support a **theory** of a **research** study.

Advantages of Theoritical Research:

- ❖ It provides a common body of understanding that practitioners automatically employ in attacking technical tasks.
- ❖ It is valuable as without the theory there would be nothing to apply.

Disadvantages of Theoritical Research:

- ❖ It is based on existing work or literature,reliability of such sources affects the conclusion of Theoritical research.
- ❖ The new ideas developed from this research are not tested through collecting evidence in form of collected primary data.

1.4.7 Applied research:

It is a form of systematic inquiry involving the practical application of science. It accesses and uses some part of the research communities' (the academia's) accumulated theories, knowledge, methods, and techniques, for a specific, often state-, business-, or client-driven purpose. Applied research is contrasted with pure research (basic research) in discussion about research ideals, methodologies, programs, and projects.

Applied research deals with solving practical problems and generally employs empirical methodologies. Because applied research resides in the messy real world, strict research protocols may need to be relaxed. For example, it may be impossible to use a random sample. Thus, transparency in the methodology is crucial.

Advantages of Applied Research:

- ❖ Social benefits to society
- ❖ Economic gains.
- ❖ Has the potential to solve real problems.
- ❖ Can create Innovation.
- ❖ It is purposeful.

Disadvantages of Applied Research:

- ✓ Needs to be carried out quickly in order to solve problems.
- ✓ Assume that all problems can be solved.
- ✓ It deals with real people, methods may have to be adopted.

1.4.8 Time series Research:

A time series design collects data on the same variable at regular intervals.(Weeks,Months,Years,etc) in the form of aggregate measures of a population.

Measurements are taken on each variable over two or more distinct time periods.

It is useful for:

- ✓ Establishing a baseline measure
- ✓ Describing changes over time.
- ✓ Forecasting future trends.

Time series data are nearly always presented in the form of a chart or graph .

The horizontal (x axis) is divided into time intervals,and the Vertical (y) axis shows the value of dependant variable as they fluctuate over time.

Advantages of Time series Research:

- ✓ Easy to collect data.
- ✓ Easy to present in graphs.
- ✓ Easy to Interpret.
- ✓ Can forecast short term trends.
- ✓ Disadvantages of Time series Research:

Disadvantages of Time series Research:

- ❖ Data collection method may change over time.
- ❖ Difficult to show more than one variable at a time.
- ❖ Needs Qualitative research to explain fluctuations.
- ❖ Assumes present trends will continue unchanged.

1.4.9 Cross sectional Research:

- ❖ A cross sectional design provides a snapshot of the variables included in the study,at one particular point in time.
- ❖ In cross sectional research design generally use survey techniques to gather data,Eg.population census.
- ❖ Cross sectional research can be exploratory,descriptive or explanatory but it is most consistent with a descriptive approach to research.

Advantages of Cross sectional Research:

- Data on many variables.
- Data from large number of Subjects.
- Data on Attitude & Behaviours.
- Answers questions on Who, What, When, Where.
- Good for exploratory research.
- Generate hypothesis for future research.

Disadvantages of Cross sectional Research:

- o Increased chances of Error.
- o Increased cost.
- o Cannot Measure change.
- o Cannot establish cause and effect.
- o No control of Independent variable.
- o Static, time bound.

1.5 Objectives Of Research

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

1.6 Research Hypothesis:

- ➢ Hypothesis is an assumption, that can be tested and can be proved to be right or wrong.
- ➢ Hypothesis is a specific statement of prediction.

- Research Hypothesis is a predictive statement that relates an independent variable to a dependant variable.
- Hypothesis must contain atleast one independent variable and one dependant variable.

1.6.1 Characteristics Of Hypothesis:

a) ***Simplicity*** - It should be stated as far as possible in simple terms.
b) ***Objectivity*** - It should not include value judgments, relative terms or any moral preaching.
c) ***Theoretical Relevance*** - It should be consistent with a substantial body of established or known facts or existing theory.
d) ***Availability of Techniques*** – Statistical methods should be available for testing the proposed hypothesis.
e) ***Conceptual Clarity*** - It should be clear and precise. Specificity - It should be specific and limited in scope.
f) ***Consistency*** - It should be consistent with the objectives of research.
g) ***Testability*** - It should be capable of being tested. Expectancy - It should state the expected relationships between variables.

1.7 Research Questions:

The research question must be accurately and clearly defined.

Choosing a research question is the central element of both quantitative and qualitative research and in some cases it may precede construction of the conceptual framework of study.

The research question serves two purposes:

1. It determines where and what kind of research the writer will be looking for and
2. It identifies the specific objectives the study or paper will address.

A **research question** is a way of expressing your interest in a problem or phenomenon. Research questions are not necessarily an attempt to answer the many philosophical questions that often arise in schools, and they are certainly not intended to be an avenue for grinding personal axes regarding classroom or school issues. You may have more than one research question for a study, depending on the complexity and breadth of your proposed work. Each question should be clear and specific. Identifying a research question will provide greater focus to your research or clarify the direction of your investigation, whether the research is descriptive or experimental.

1.7.1 Characteristics of Good Research Questions

- Are specific.
- Are clear.
- Refer to the problem or phenomenon.
- Reflect the intervention in experimental research.
- Note the target group of participants.

How Do You Formulate A Good Research Question?

Choose a general topic of interest, and conduct preliminary research on this topic in current periodicals and journals to see what research has already been done. This will help determine what kinds of questions the topic generates.

Once you have conducted preliminary research, consider: Who is the audience? Is it an academic essay, or will it be read by a more general public? Once you have conducted preliminary research, start asking openended “How?” “What?” and Why?” questions. Then evaluate possible responses to those questions.

Possible Question: *Why are social networking sites harmful?*
An evaluation of this question reveals that the question is unclear: it does not specify which social networking sites or state what harm is being caused. Moreover, this question takes as a given that this "harm" exists. A clearer question would be the following:
Revised Question: *How are online users experiencing or addressing privacy issues on such social networking sites as Facebook and Twitter?*
While a good research question allows the writer to take an *arguable* position, it DOES NOT leave room for ambiguity.

Research Question in the Sciences and Social Sciences
While all research questions need to take a stand, there are additional requirements for research questions in the sciences and social sciences. That is, they need to have **repeatable** data. Unreliable data in the original research does not allow for a strong or arguable research question.
In addition, you need to consider what kind of problem you want to address. Is your research trying to accomplish one of these four goals?
1) Define or measure a specific fact or gather facts about a specific phenomenon.
2) Match facts and theory.
3) Evaluate and compare two theories, models, or hypotheses.
4) Prove that a certain method is more effective than other methods.
Moreover, the research question should address what the variables of the experiment are, their relationship, and state something about the testing of those relationships.
In essence, the research question that guides the sciences and social sciences should do the following three things:
1) Post a problem.
2) Shape the problem into a testable hypothesis.
3) Report the results of the tested hypothesis.
There are two types of data that can help shape research questions in the sciences and social sciences: quantitative and qualitative data.
While quantitative data focuses on the numerical measurement and analysis between variables, qualitative data examines the social processes that give rise to the relationships, interactions, and constraints of the inquiry.

1.8 Research Objectives:

The research objectives define the type and extend of information needed to achieve the research objectives.Research objectives set the purpose and focus of your research with the fundamental questions that will be addressed.
A research objective can be specified broadly, narrowly.

1.8.1 Framing Research Objectives:

Precise : Precise means that the terminology is understandable to the marketing manager and that is accurately capture the essence of each item to be researched.
Detailed: Detail is provided by elaborating,perhaps with examples,each item.
Clear: The objective is clear if there is no doubt as to what will be researched and how the information will be presented to the manager.
Operational: Research objective must be operational. The research objectives should define how the construct being evaluated is actually measured.

1.8.2 Types Of Research Objectives:

i. Market attractiveness Evaluation - Market sizing-estimating the size of total market.
ii. Customer Insight - Specify customer needs,Aspirations,Buying behaviours,decision models,preferences etc.
iii. Competitive forces – current and potential basis of competition in the market.

iv. Communication planning – What information sources do prospective customers pay attention to, how to reach them,opportunities for influencingtarget customers and which are most effective.
v. Product Testing – Evaluation of product improvements,Alternatives,Packaging etc.
vi. Concept testing – Evaluation of Potential products and solutions,clarification of needs,wants and preferences.
vii. Customer satisfaction – Measurement of quality of customer experience,perceptions,reaction,loyalty,intent,etc.
viii. Pricing – Testing of Price/feature/quality/packaging/positioning combinations,promotions,terms & conditions etc.

1.9 Role of theory in Research

Theory:

A theory is a set of systematically interrelated concepts,definitions,and propositions that are advanced to explain and predict phenomena.

Theories are simply generalizations that help us better understand Reality. If theory doesnot hold true in practice, then that theory holds no value.

A theory is a formal, testable explanation of some events that includes explanations of how things relate to one another.

1.9.1 Role of Theory in Research:

i)Prediction of Behaviour:

A theory enables to predict the behavior or characteristics of one phenomenon from the knowledge of another phenomenon.

Accomplishing the first goal allows the theorist to predict the behavior or characteristics of one phenomenon from the knowledge of another phenomenon's characteristics.

A researcher also wants to gain understanding. Predictions and understanding go in hand. To predict phenomena, one must have an explanation of why variables behave as they do. Theories provide explanations.

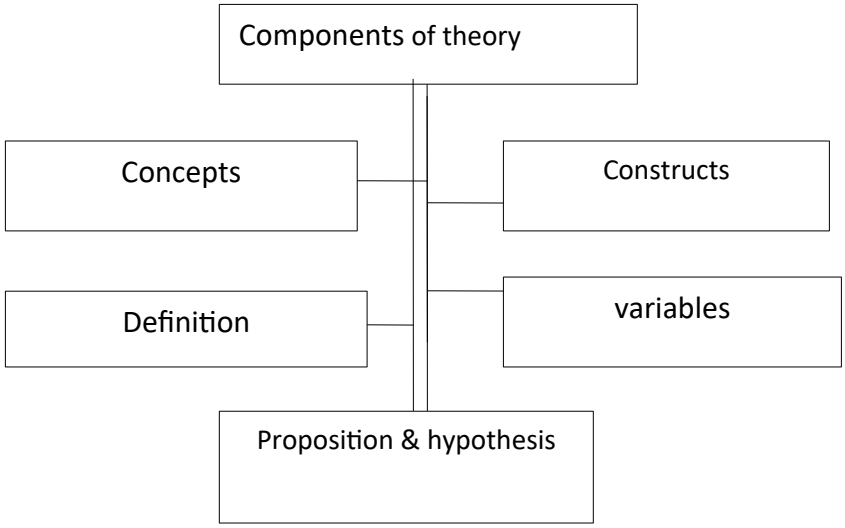
ii) Understanding Relationships:

To gain an understanding of the relationship among various phenomena.Before one can see what a proposition is,however he must discuss the nature of theoretical concepts.

iii) Other Roles:

- Suggest a problem for the study
- Gives a hypothesis to be tested.
- Provide a simple explanation about the observed relations regarding their relation to a phenomenon.
- Makes research findings Intelligible.
- Helps in the selection of variables or identification of classes of data to be collected.

Components of Theory:



Concepts:

To understand and communicate information about objects and events,there must be a common ground on which to do it.Concepts serve this purpose.

A concept is generally accepted collection of meanings or characteristics associated with certain events,objects,conditions,situations,and behaviours. Classifying and categorizing objects or events that have common characteristics beyond any single observation creates concepts.

Constructs:

A construct is a theoretical dimension that has been or potentially could be operationalised by one or more variables.

The terms concept and construct are often used in place of one another.

Definitions:

Definitions are one way to reduce this danger. For eg: Customer is defined as a patron; a patron, in turn,is defined as a customer or client of an establishment; a client is defined as one who employs the service of any professional as a patron of any shop.

Variables:

A Variable is a symbol of an event, act, characteristic,trait,attribute that can be measured, and to which we assign categorical values. Variables also take on values representing added categories,such as demographic variables of race or religion.

Proposition and Hypothesis:

A proposition is a statement about observable phenomena that may be judged as true or false. When a proposition is formulated for empirical testing, people call it as Hypothesis. It is a formal statement explaining some outcome. A hypothesis is a guess. For eg: A human resource manager may hypothesize that job candidates with certain majors will be more successful employee

CHAPTER II

RESEARCH DESIGN AND MEASUREMENT

2.1 Research Design:

2.1.1 Definition: The framework developed to control the collection of data is called research design. Research design is an absolute essentiality in research irrespective of the type of research (e.g., exploratory or descriptive), as it ensures that the data collected is appropriate, economical and accurate.

A sufficiently formulated research design would ensure that the information gathered is consistent with the study objectives and that the data are collected by accurate procedures. Since, research designs germinate from the objectives, the accuracy and adequacy of a research design depends on the unambiguous framing of the objectives.

2.1.2 Concepts Relating to Research Design:

1) Dependent and Independent variables :

Variables : A magnitude that varies is known as "variable"

Continuous variable : Values that can be expressed even in decimal poins are known as continuous variables

Eg: age (4 years 3 months)

Height (5.2 cm)

Weight (45.3 kg)

Non continuous Variables: Value that can be expressed only in integer values are called Non continuous variables

Eg: No. of students in a class (45)

No. of children in a family (3)

Statistically known as "discrete variables"

Dependent or Endogenous variables :

When the change in one variable depends on the change in other variable, it is known as dependent or Endogenous variable.

Demand ----- Price (independent)

Independent or Exogenous variable

The variable that causes the change in the dependent variable is known as independent or exogenous variable.

Demand (Dependent) ------- Price ,Income

Here demand is a dependent variable while price / income is an independent variable.

Extraneous variable :

The independent variable which is not directly related to the purpose of the study but affects the dependent variable is know as Extraneous variables.

- The influence caused by the extraneous variable on the dependent value is technically known as "Experimental Error"
- A research study or a Research design should always be framed in such a manner that the influence of 'Extraneous variables' on the dependent variable is completely controlled and the influence of the independent variable is clearly evident.

Control:

Good Research design should minimize the effect for Extraneous variables.

Confounded Relationship

The relationship between dependent and independent variable is said to be confounded by an extraneous variables.

Research Hypothesis:

When the formulated hypothesis is tested by adopting scientific methods, it is known as Research Hypothesis.

Experimental & Non Experimental Hypothesis testing:

- When the objective of the Research is to test the hypothesis, it is Research hypothesis.
- Research in which the independent variable are (handled with skill) manipulated, it is experimental hypothesis testing.
- When the variables are not manipulated, it is non experimental hypothesis testing.

Experimental & Control Groups:

- When a group is exposed to usual conditions in an experimental hypothesis, research it is control Groups.
- When the group is exposed to special or certain new conditions, it is experimental groups.

8. Treatments:

The different conditions to which the experimental & control groups are subject to is known as treatments.

9. Experiment: Fertilizers and crops)

Process of verifying the truth.

Absolute Experiment:

- Determine the fact

Comparative Experiment:

- Determine the impact in comparison with another fact.

10. Experimental units

- Pre-determined block to which different treatments are applied.

Eg : animal testing

2.1.3 Importance of Research Design:

- Facilitates the smooth flow of the various stages of Research.
- Helps yield maximum information with minimum effort, time and money.
- Helps to plan in advance data collecting and analysis techniques.
- Prepare with utmost care to avoid errors.
- Attain reliability

2.1.3 Importance of Research Design:

- Facilitates the smooth flow of the various stages of Research.
- Helps yield maximum information with minimum effort, time and money.
- Helps to plan in advance data collecting and analysis techniques.
- Prepare with utmost care to avoid errors.
- Attain reliability

2.1.4 Characteristics of a Good Research Design

- Posses the qualities of being flexible, suitable efficient & economical.
- Should minimize 'bias' and maximize reliability of data collection & Analysis.
- No experimental error should be allowed
- Should yield maximum information
- Research problem should be viewed from different angles or dimensions.

The choice of Research design depends on,

- Nature of the Research problem
- Objectives of the Research problem
- Skills / ability of the Researcher
- Methods of gathering information
- Availability of monetary support
- Time schedule

2.2 Types Of Research Design

Two types of research design are established according to the nature of the research objectives or types of research. They are:

- Exploratory design; and
- Conclusive design. (Descriptive researh and casual research)

2.2.1 Exploratory Research Design

It is appropriate when the research objective is to provide insights into (i) identifying the problems or opportunities (ii) defining the problem more precisely, (iii) gaining deeper insights into the variables operating in a situation (iv) identifying relevant courses of action (v) establishing priorities regarding

the potential significance of a problems or opportunities (vi) gaining additional insights before an approach can be developed and (vii) gathering information on the problems associated with doing conclusive research. Much research has been of an exploratory nature; emphasising on finding practices or policies that needed changing and on developing possible alternatives.
On examination of the objectives of exploratory research, it is well understood that it could be used at the initial stages of the decision making process. It allows the marketer to gain a greater understanding of something that the researcher doesn‟t know enough about. This helps the decision maker and the researcher in situations when they have inadequate knowledge of the problem situation and/or alternative courses of action. In short, exploratory research is used in the absence of tried models and definite concepts.

Exploratory research could also be used in conjunction with other research. As mentioned below, since it is used as a first step in the research process, defining the problem, other designs will be used later as steps to solve the problem. For instance, it could be used in situations when a firm finds the going gets tough in terms of sales volume, the researcher may develop use exploratory research to develop probable explanations. Analysis of data generated using exploratory research is essentially abstraction and generalization. Abstraction refers to translation of the empirical observations, measurements etc. into concepts; generalization means arranging the material so that it focuses on those structures that are common to all or most of the cases.

The exploratory research design is best characterised by its flexibility and versatility. This is so, because of the absence of the non-imperativeness of a structure in its design. It predominantly involves imagination, creativity, and ingenuity of the researcher. Examples of exploratory research are:
- survey of experts to validate an instrument
- pilot studies conducted to perform reliability check on a questionnaire
- use of secondary data in order to analyse it in a qualitative way
- qualitative research.

2.2.2 Conclusive Research Design

It involves providing information on evaluation of alternative courses of action and selecting one from among a number available to the researcher.

Conclusive research is again classified as:
(i) Descriptive research, and
(ii) Causal research.

(i) ***Descriptive Research:*** It is simple to understand as the name itself suggests that it involves describing something, for example:
(a) market conditions;
(b) characteristics or functions;
(c) estimate the percentage of customers in a particular group exhibiting the same purchase behaviour;
(d) perceptions of product characteristics; and
(e) to predict the pattern of behaviour of characteristic versus the other

Majority of research studies are descriptive studies. As research studies involve investigating the customers/consumers, collection of data includes interrogating the respondents in the market and data available from secondary data sources. However, it cannot be concluded that descriptive studies should be simply fact-gathering process. Descriptive study deals with the respondents in the market and hence, extreme caution has to be exercised in developing this study. Much planning should be done, objectives should be clear than exploratory studies.

In descriptive research, the data is collected for a specific and definite purpose and involves analysis and interpretation by the researcher. The major difference between exploratory and descriptive research is that descriptive research is characterised by the formulation of specific

objectives. The success of descriptive studies depends on the degree to which a specific hypothesis acts as a guide.

Descriptive studies restrict flexibility and versatility as compared to exploratory research. It involves a higher degree of formal design specifying the methods for selecting the sources of information and for collecting data from those sources. Formal design is required in order to ensure that the description covers all phases desired. It is also required to restrain collection of unnecessary data. Descriptive studies require a clear specification of the who, when, where, what, why and how.
While designing a descriptive research, the researcher should also have sufficient knowledge on the nature and type of statistical techniques he/she is going to use. This will greatly help to have the right design in place. Mostly descriptive studies are conducted using questionnaire, structured interviews and observations. The results of description studies are directly used for marketing decisions.

Descriptive studies are again classified into two types:
(a) Longitudinal
(b) Cross sectional
(a) Longitudinal research relies on panel data and panel methods.

It involves fixing a panel consisting of fixed sample of subjects that are measured repeatedly. The panel members are those who have agreed to provide information at a specific intervals over an extended period. For example, data obtained from panels formed to provide information on market shares are based on an extended period of time, but also allow the researcher to examine changes in market share over time. New members may be included in the panel as an when there is a dropout of the existing members or to maintain representativeness.

Panel data is analytical and possess advantages with respect to the information collected in the study. They are also considered to be more accurate than cross sectional data because panel data better handle the problem associated with the errors that arise in reporting past behaviour and the errors that arise because of the necessary interaction between interviewer and respondent.

(b) Cross-sectional research is the most predominantly and frequently used descriptive research design in marketing. It involves a sample of elements from the population ofinterest. The sample elements are measured on a number of characteristics.

There are two types of cross-sectional studies:
- Field studies and
- Surveys

It may appear that field studies and surveys are no different but the same. However, for practical reasons, they are classified into two categories cross sectional research. The fundamental difference lies in the depth of what these research cover. While survey has a larger scope, field study has greater depth. Survey attempts to be representative of some known universe and filed study is less concerned with the generation of large representative samples and is more concerned with the in-depth study of a few typical
situations.

Cross sectional design may be either single or multiple cross sectional design depending on the number of samples drawn from a population. In single cross sectional design, only one sample respondents is drawn whereas in multiple cross sectional designs, there are two or more samples of respondents. A type of multiple cross sectional design of special interest is Cohort analysis.
Cohort analysis consists of a series of surveys conducted at appropriate time intervals, where the cohort serves as the basic unit of analysis. A cohort is a group of respondents who experience the same event within the same time interval.

(a) ***Case Study****:* This study involves intensive study of a relatively small number of cases. In this method, much emphasis is on obtaining a complete description and understanding of factors in each case, regardless of the number involved. It could be used significantly, particularly when one is

seeking help on a problem in which interrelationships of number of factors are involved, and in which it is difficult to understand the individual factors without considering them in their relationships with each other. As in the case of exploratory research, case method is also used in conjunction with exploratory research as first step in a research process. It is of prime value when the researcher is seeking help on a market problem in which the interrelationships of a number of factors are involved, and in which it is difficult to understand the individual factors without considering themin their relationships with each other.

(ii) ***Causal research:*** It is used to obtain evidence of cause-and-effect relationships with is otherwise known as the independent-dependent relationship or the predictive relationships. This is an important type of research useful for marketers as this allows marketers to base their decision on assumed causal relationships. Causal research is done in the following situations:
(a) To identify which variables are the cause and which are the effect. In statistical terms, causal variables are called independent variables and effectual variables are called dependent variables.
(b) To determine the nature of the relationship between the causal variables and the effect to be predicted.
Causal research requires a strong degree of planning on the design as its success depends on the structure of the design.

2.2.2 Experimental Design:

Experiment is the process of examining the truth of a statistical Hypothesis related to some research problem.

Experiments are of two types,

1. Absolute Experiment.

2. Comparative Experiment.

Absolute Experiment:

When a researcher wants to determine the impact of a fertilizer on the yield of a crop, it is a case of Absolute Experiment.

Comparative Experiment:

When a researcher wants to determine the impact of one fertilizer as compared to the impact of some other fertilizer, it will be called as Comparative Experiment.

Research Design are of three types,

1. Research design in case of descriptive & diagnostic studies.

2. Research design in case of exploratory Research studies.

3. Research design in case of Hypothesis Testing Research Studies.

Research Design In case of Hypothesis Testing Research Studies:

- Hypothesis testing research studies are generally known as Experimental studies.
- The researcher test the casual relationship between the variables.
- Professor Fisher is considered as the pioneer of this type of studies. (Experimental Studies)
- He performed this study when he was working at a Agricultural Research Station in London.
- "His found out that, by dividing plots into different blocks and then by conducting experiments in each of these blocks, whatever in formations is collected and inference

drawn from them can be more reliable Professor Fisher laid three principles of Experimental Designs,

1. The Principal of Replication

2. The Principal of Randomization.

3. The Principal of Local Control.

The Principal of Replication :

"The Experiment should be repeated more than once".

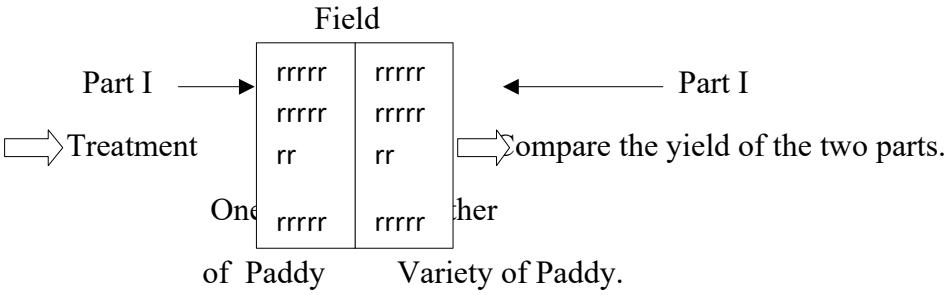
- The treatment is applied to many experimental units.
- The information collected and the inference drawn from these experimental units will be more reliable and statistically accruable.

Aim: To examine the effect of two varieties of paddy.

Example: A paddy field is divided into 2 parts. Grow one variety in one part and the other variety in the other. Then we compare the yield of the two parts.

Draw conclusion on that basis.

No Principle of Replication is



When Principle of Replication is used:

- First divide the field into several parts.
- Grow one variety in half of the parts and the other variety in the remaining parts.
- Collect the information of the two varieties and draw the conclusion by comparing both.

The Results so obtained will be more reliable and accurate compared to the results drawn without using the principle of Replication.

The Experiment can be repeated several times.

Field

Tre l

Rr	Rr	R	r	r
Rr	Rr	R	r	r
Rr	Rr	R	r	r
Rr	Rr	R	r	r
rr	Rr	R	r	r

⇨ Conclusion drawn.

↓ Parts

2. The Principle of Randomization:

- "Principle of Randomization "Provides us a protection against the effects of "Extraneous Variables."
- The variations or effects caused by these extraneous variables can be combined under the heading "Chance".

Example: When the researcher grows one variety of paddy in the first half of the field and the other variety in the next half of the field, there may be a possibility or chance that the soil fertility of the first half of the field may be different in comparison to the next half.

- In this case, he may go on to cultivate the two varieties of paddy in different parts of the field on the basis of some random sampling technique.

(i.e.) He may apply Randomization principle and protect himself from the effects of the Extraneous Factors.

- By using Randomization Principle a better estimate can be drawn.

⇨ When treated

Rr	rr	r	r	r
Rr	rr	r	r	r
Rr	rr	r	r	r
Rr	rr	r	r	r
Rr	rr	r	r	r

⇨

Conclusion drawn is more accurate.

(Can protect Effects of Extraneous Variables)

3. The Principle of Local Control:

- The extraneous variable which is a known source of variability can be made to vary extensively or deliberately over a wide range.
- Now the variability it causes can be measured and eliminated.
- In short, through the principle of Local Control, we can eliminate the variability due to extraneous factors from the experimental error.
- The extraneous variable is brought to a control.

Kinds of Experimental Design:

- Experimental Design refers to the framework of the structure of an experiment.

Classified into 2 Broad Categories,

1. Informal Experimental designs.
2. Formal Experimental designs,

I.Informal Experimental Designs:

Designed based only on the difference between the magnitudes or performance.

- Three Types,
    1. Before and after without control design.

2. After Only with Control design.
3. Before and after with control design.

Before and after without control design:

Consider a test group,

Step.1: The dependent variable is measured before introduction of the treatment.

Step.2: The treatment is introduced.

Step.3: The dependent variable is measured after the treatment has been introduced.

Step.4: Inference:

The effect of the Treatment : The level o the phenomenon after the treatment.

- The level of the phenomenon before the treatment.

Test Area Level of Phenomenon Treatment Level of Phenomenon

Introduced

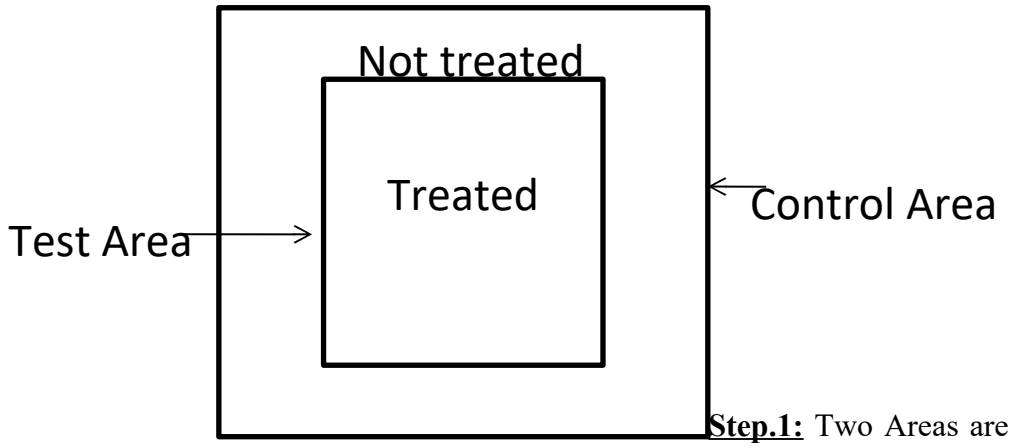
Before Treatment (X) ⟶ After Treatment (Y)

þ Effect of the Treatment = (Y) – (X)

Demerit:

With the passage of time, several extraneous variable may be there in the treatment effect.

(2) After only with control Design:



Step.1: Two Areas are selected, the control Area & the test area.

Step.2: The treatment is introduced in the test area alone.

Step.3: The dependent variable in both the areas are measured, at the same time.

Step.4: Treatment Effect is calculated by subtracting the value of the dependent variable in the control area from its value in the test area.

Treatment: 1

Effect = Value of dependent variable in the

control Area – Value of Dependent Variable in the test area.

Test Area	Treatment Introduced	Level of Phenomenon (Y)
Control Area	No → Treatment	Level of Phenomenon (Z)

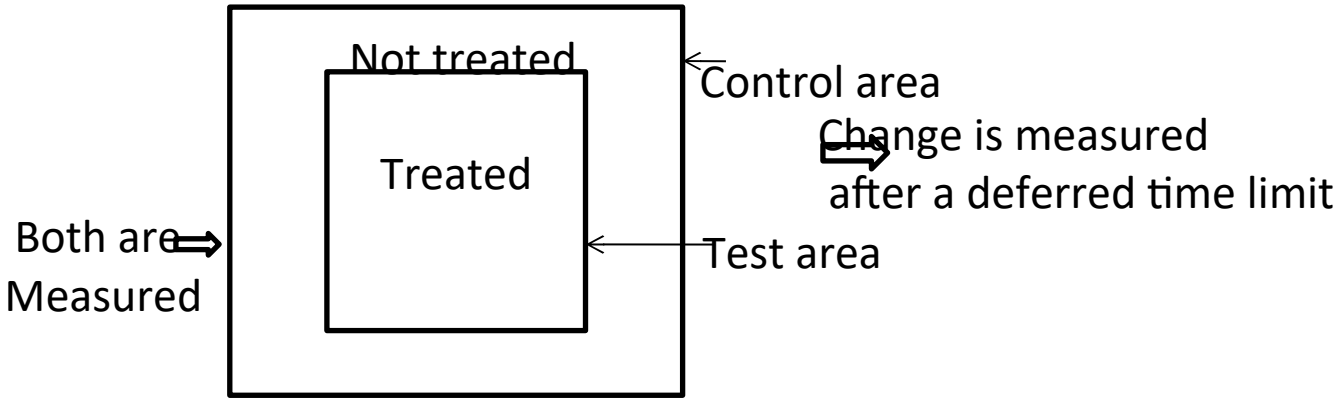
(3) Before And After with Control Design:

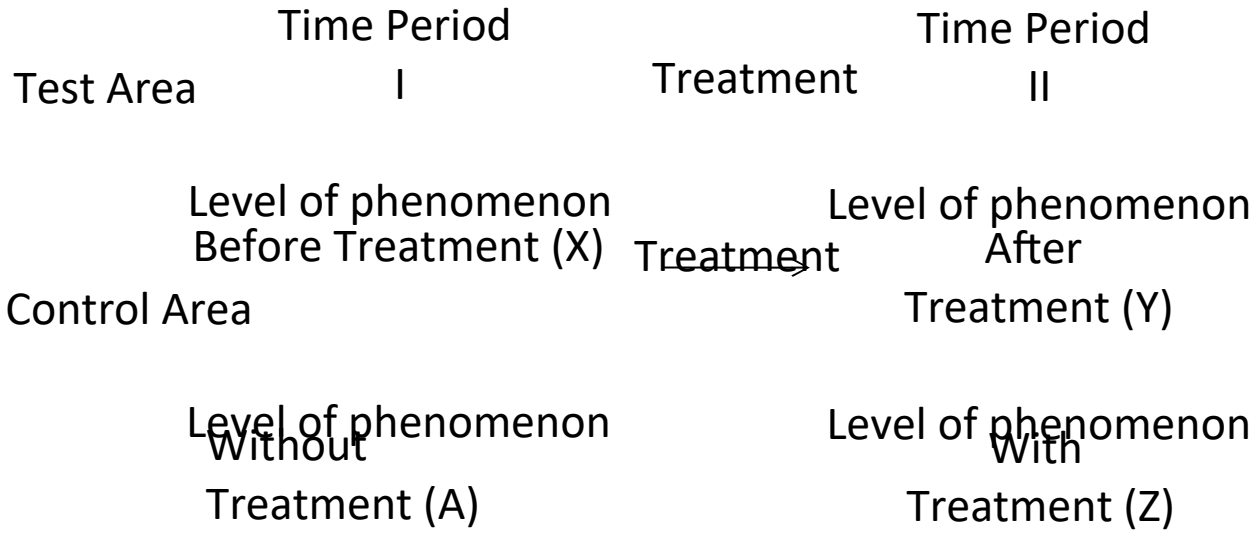
Step.1: In this design, two areas are selected and the dependent variables in both the areas are measured for an identical time period before treatment.

Step.2: Treatment is introduced only in the test area.

Step 3 : The dependent valuable is measured on both the areas (control area & test area) for an identical time period.

Setp4 : The effect of the treatment is determined by subtracting the change in the dependent valuable in the control area with the charge in the dependent valuable of the test area.





Treatment :-

Effect = ((Y) – (X)) – ((Z)-(A))

Merits: Avoids Extraneous variables resulting from passage of time and non

comparability of control and test areas.

II. Formal Experimental Design:

Offer relatively more control and use specific statistical procedures for analysis types,

1) Complete Randomized design (Generally called C.R. Design)
2) Randomized Block Design (R.B. Design)
3) Latin Square Design (L.S. Design)
4) Factorial Designs .

(1) Completely Randomized Design :-

- Involves two principals, the principle of replication and the principle of Randomized of the experimental designs.
- The items are randomly assigned do experimental treatments.
- This design is simpler and easier.
  Example: It the research has 2 items of 20 parts and if he wishes to test to under treatment B, this completely randomized design gives every possible group of 10 items selected from a set of 20, an equal chance of being assigned to treatment A & treatment B.
- One way analysis of variance (one way ANOVA) is used to analyze such a design.

Treatment A on X

&

Treatment B on Y

x	y	x	y
y	x	y	x
x	y	y	x
y	x	x	y
x	y		x

20 1 teams

(10+10)

2. Randomizes Block Design:-

- The subjects or items are first divided into groups, known as **"Blocks"**
- See that, the items in each group or black is homogenous.
- Randomly select items from each given block and assign treatment.
- Extraneous variables can be fixed and can be measured.
- The main feature of this study is, each treatment appears the same no of times in each block.
- This design is analyzed two way analyses of variance (two way ANOVA) technique.

3. Latin Square Design:-

- Used in Agricultural Research.
- L.S Design is used when two or more extraneous variables is found.

Example: Effect of fertilizer on the field of wheat is do be determined.

- Here along with the effect of fertilizer, the fertility of the soil must be considered.

- If the facility of the soil is not considered along with the fertilizer the result obtained may be dependable.
- Similarly the impact of the various seeds used many also vary the yield.
- To overcome this difficulty L.S design is used.
- Each fertilizer (X1, X2, X3, X4, X5) will appear 5 items but will be used only once in each row and in each coloumn.

Example: No treatment occurs more than once.

I II III IV V

X1	A	B	C	D	E
X2	B	C	D	E	A
X3	C	D	E	A	B
X4	D	E	A	B	C
X5	E	A	B	C	D

Conclusion :

- The field is divided into several blocks (I, II, III, IV & V) and there are variety of fertilizer (X1, X2, X3, X4, X5).
- But each fertilizer is used in each block only once.
- a two way **ANOVA** technique.

4. Factorial Design :

- Are used in experiments where the effect of the depended variable, when affected by more than one variable is to be determined.
- Used in social & economic studies where usually large no of factors affect a particular problem.
- Factorial design are of two types:
  I . Simple Factorial design

  II. Complex Factorial designs.

Simple Factorial Design :

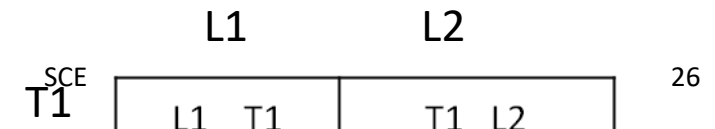
- When the effect of the dependent variable is affect by only two factors, it is simple factorial designs.
- Otherwise known as " TWO factors Factorial Design".

Complex Factorial Design :

- This design is used when more than two factors at a time affects the dependent variable.
- Or the design considers three or more independent variable.
- The greater the no of independent variable, the higher the order of interaction, analysis possible.

Advantage :

- Accuracy
- Can determine the effects of more variable in a single experiment.



2.3 Variables

- A variable is a characteristic, trait, or attribute of a person or thing that can be classified or measured.
- "Variable" = "more than one value"
    - Age, gender, height, weight, ethnicity, etc.
- 2 basic types of variables:
    - Quantitative
    - Qualitative
- Qualitative Variables
- A.k.a. = Categorical Variables
- "Unmeasurable" variables
    - Gender, eye color, political affiliation, etc.

- 2 types of categorical variables:
  Ordinal
    - Data values represent categories with some intrinsic order (for example, low, medium, high; strongly agree, agree, disagree, strongly disagree).

  Nominal
    - Data values represent categories with no intrinsic order (1 = Male, 2 = Female).
- Quantitative Variables
- Characteristics, attributes, or traits that can be measured.
- "Measurable" variables
- Age, height, weight, etc.
- 2 types of quantitative variables:
    - Discrete or *Scale* Variables
        - Data values are numeric values on an interval or ratio scale.
        - Values that cannot be fractionated or divided (whole numbers).
        - Test scores, game scores, age, etc.
    - Continuous Variables
        - Precision-based measurements
        - Strength, endurance, track & field times, height, weight, girth, etc.
- Variables
- Independent variables
- Dependent variables
- Extraneous variables
- Independent Variables (IV)
- Experimental treatment; treatment variable
- Will not change during the course of the study
- Antecedent to other variables
- 2 types of IVs:
    - Active
        - A variable that is actually manipulated by the researcher
        - Methods of training, form of reinforcement, type of nutritional supplement
    - Attribute
        - Preexisting attribute that cannot be manipulated
        - Gender, race, age, or grade level…

        - Comparison groups can be formed
- Dependent Variables (DV)
- Expected to change as a result of the treatment
- Not under the control of the researcher(s)
- IV → affects → DV
- Common IV and DV examples
- IVs
    - Exercise
    - Diet
    - Medicine(s)
    - Drugs
    - Motivation
    - Programs
    - Procedures
    - Methods
    - Techniques
- DVs
    - Performance
    - Fitness
    - Learning
    - Health
    - Knowledge
    - Behavior
- Extraneous Variables
- Error-producing variables (other than the IVs) that may impact the DV response.
- A.k.a.:
    - Intervening variables
    - Modifying variables
    - Confounding variables
- Overview of Variable Classifications
- Qualitative
    - Ordinal
    - Nominal
- Quantitative
    - Discrete (scale)
    - Continuous
- Independent variable (IV)
- Dependent variable (DV)
- Extraneous variable (EV)

2.4 The Measurement Process

Measurement is defined as the assignment of numbers to characteristics of objects or events according to rules. The definition of measurement clearly states that the researcher should know that the measurement scale measures the characteristics of the objects or event and not the objects or events.

The numerical application on all measurements and the analysis of numbers using mathematical or statistics involve one or more of the four characteristics of number system. Measurement of any property could be fitted into any of these characteristics.

2.4 .Levels Of Measurement

Researchers normally use four levels of measurement scales. They are:

a) Nominal scale
b) Ordinal scale
c) Interval scale
d) Ratio scale

2.4.1 Nominal Scale

Nominal scale are categorical scales used to identify, label or categorise objects or persons or events. A familiar example is the use of alternative numbering system by our Physical Education Teacher in our school days to engage us in a game. The teacher as a result would form two groups one labelled 1 and the other 2. The numbers 1 and 2 are assigned to two groups and the members belonging to group 1 would exclusively be a part of group 1 and the members belonging to group 2 would exclusively be a part of group 2. However, assigning the numbers does not indicate any order or position to the group it represents. Interchanging the numbers otherwise would also result in the same effect in that, the order or position would not change.

Nominal scales are the lowest form of measurement. The simple rule to be followed while developing a nominal scale: Do not assign the same numerals to different objects or events or different numbers to the same object or event. In marketing nominal scales are used substantially in many occasions. For example, nominal scale is used to identify and classify brands, sales regions, awareness of brands, working status of women etc.,

On data generated using nominal scale, the type of statistical analysis appropriate are mode, percentages, and the chi-square test. Mode alone could be used as a measure of central tendency.

Mean and median could be employed on nominal data since they involve higher level properties of the number system. Researchers should be careful enough to identify the type of scales before they apply any statistical technique. The researcher may not be able to make any meaning inference from the mean or median value obtained from nominal data.

2.4.2 Ordinal Scale

Ordinal scale is a ranking scale that indicates ordered relationship among the objects or events. It involves assigning numbers to objects to indicate the relative extent to which the objects possess some characteristic. It measure whether an object or event has the same characteristic than some other object or event. It is an improvement over nominal scale in that it indicates an order.

However, this scale does not indicate on how much more or less of the characteristic various objects or events possess. The term how much refers to ranks that it do not indicate if the second rank is a close second or a poor second to the first rank.

Data generated using ordinal scale appears as ranks where the object which has ranked first has more of the characteristic as compared to those objects ranked second or third. Hence, the important feature of ordinal scale over nominal scale is that it indicates relative position, not the magnitude of the difference between the objects. In research, ordinal scales are used to measure relative attitudes, opinions, perceptions etc., Most data collected by the process of interrogating people have ordinal properties. To illustrate, a marketer may be interested in knowing the preference of the customers across various brands. The customers may be requested to rank the products in terms of their preference for the products.

The numbers assigned to a particular object or event can never be changed in ordinal scales. Any violation of this principle would result in confounding results by the researcher. Mean is not an appropriate statistic for ordinal scale.

2.4.3 Interval Scale

Interval scale is otherwise called as rating scale. It involves the use of numbers to rate objects or events. It interval scales, numerically equal distances on the scale represent equal values in the characteristic being measured. Interval scale is an advancement over the ordinal scale that it has all the properties of an ordinal scale plus it allows the researcher to compare the differences between objects. It also possesses the property of equality of difference between each levels of measurement. The feature of this scale is that the difference between any two scale values is identical to the difference between any other two adjacent values of an interval scale. Examples of interval scales are the Fahrenheit and Celsius scales.

Interval scales also place restriction on the assignment of values to the scale points. The zero that could be assignment is a arbitrary zero rather than a natural zero. Arbitration involves freedom to place the zero value on any point. There is a constant or equal interval between scale values.

In research, most of the research on attitudes, opinions and perceptions are done using scales treated as interval scales. All statistical techniques that are employed on nominal and ordinal scales could also be employed on data generated using interval scales.

2.4.4 Ratio Scales

Ratio scales differ from interval scales in that it has a natural/absolute zero. It possesses all the properties of the normal, ordinal and interval scales. Data generated using ratio scales may be identified, classified into categories, ranked and compared with others properties. It could also be expressed in terms of relativity in that one can be expressed in terms of a division of the other. Hence, it may be called as relative scales.

Ratio scales have great many number of application in research. They include sales, market share, costs, ages, and number of customers. In all these cases, natural zero exists. All statistical techniques can be applied on ratio data.

2.5 Validity & Reliability Of Measurement

Research should always be based on absolutely correct, defectless and errorless measuring instruments, tools or procedures of measurement. For this purpose the acceptability of a measuring instrument should be tested on the principles of adherence to the standards of perfect reliability, confirmed practicality and verified validity. The reliability of an instrument can be ensured when it conforms to certain prescribed norms. It is not the physical form or shape but it is the accuracy of the prescribed standard content of the instrument that leads to acceptability. An instrument should be conveniently usable with verifiable validity. Perfection in measurement can be achieved if a researcher, at the outset, carries out appropriately, the prescribed tests of reliability, practical acceptability and validity of his tools of measurement.

2.5.1 Errors in Measurement

Errors in the course of measurement can be traced to a number of factors such as carelessness, negligence, ignorance in the usage of the instruments. If appropriate and defectless instruments are used and care is taken in the process of measurement, only then can accuracy in research be ensured.

In regard to survey-work, where the researcher obtains information through interviews, it is necessary, to judge as to whether the respondent is providing accurate facts or is biased. As situational factors also influence measurement, it is imperative that the researcher adopts his measuring procedures accordingly.

Research findings and conclusions can be reliable and acceptable if they are based on sound analysis carried out through appropriate procedures of error-free and perfect measuring tools.

2.6 Scaling Techniques

A well-designed research problem constitutes a well designed measurement process. The process of measurement is a fundamental aspect of any research. This is the step where you actually try to find out the reality by measuring it. Decision makers are more interested as the steps prior to this step are purely descriptive, and, this is the step where actual quantification happens.

Developing effective measures of marketing is not an easy task. The measures should be devoid of measurement errors. There may be disastrous situations where the marketer may be confused with the findings of the data. If he is well aware of the confounding results, then he may discard the findings the emerge from the data analysis. This requires lot of wisdom and knowledge in identifying if the data that resulted from the measurement is consistent, unambiguous etc., But unfortunately, marketers may not be interested in knowing or rather would not know the type of scales used to measure the aspects involved in the marketing problem. Any decision made based on the findings would lot of negative implications on the organisation. Hence, it is very imperative that the researcher is wise enough to develop measurement scales that capture the right property with appropriately.

The scaling techniques employed in research could be broadly classified into comparative and non comparative scale. Comparative scales as its name indicate derive their name from the fact that all ratings are comparisons involving relative judgements. It involves direct comparison of stimulus objects. It contains only ordinal or rank order properties. It is also otherwise called non metric scales in that it does not allow any numerical operations on it against all that could be applied on interval and ratio scales. Comparative scales involve the direct comparison of stimulus objects.

2.6.1 Comparative Scaling Techniques

Comparative scaling techniques consist of:

a) Paired comparison scaling
b) Rank order scaling
c) Constant sum scaling and
d) Q-sort.

2.6.1.1 Paired Comparison Scaling

Paired comparison scaling as its name indicates involves presentation of two objects and asking the respondents to select one according to some criteria. The data are obtained using ordinal scale. For example, a respondent may be asked to indicate his/her preference for TVs in a paired manner.
Paired comparison data can be analysed in several ways. In the above example, the researcher can calculate the percentage of respondents who prefer one particular brand of TV over the other. Under the assumption of transitivity, data generated using paired comparison technique could be converted to a rank order. Transitivity of preference implies that if a respondent prefers brand X over brand Y, and brand Y is preferred to Z, then brand X is preferred to Z. This may be done by determining the number of times each brand is preferred by preference, from most to least preferred.

Paired comparison technique is useful when the number of brands is limited, as it requires direct comparison and overt choice. However, it is not so, that possible comparison could not be made, but comparisons would become so much unwieldy.
The most common method of taste testing is done by paired comparison where the consumer may be, for example, asked to taste two different brands of soft drinks and select the one with the most appealing taste.

2.6.1.2 Rank Order Scaling

In rank order scaling is done by presenting the respondents with several objects simultaneously and asked to order or rank them based on a particular criterion. For example, the customers may rank their preference for TVs among several brands. In this scaling technique, ordinal scale is used. The consumers may be asked to rank several brands of television in an order, 1 being the most preferred brand, followed by 2, 3 and so on. Like paired comparison, it is also comparative in nature.
Data generated using this technique are employed with conjoint analysis because of the discriminatory potential of the scaling, stimulating the consumers to discriminate one brand from the other.

2.6.1.3 Constant Sum Scaling

This technique allows the respondents to allocate a constant sum of units, such as points, rupees or among a set of stimulus objects with respect to some criterion. The technique involves asking the respondents to assign 10 points to attributes of a sports utility vehicle. If the attribute is unimportant, then the respondents would want to enter zero.

The attributes are scaled by counting the points assigned to each one by al the respondents and dividing the number of respondents. This predominantly uses ordinal because of its comparative nature and the resulting lack of generalisability. Constant sum scaling has advantage in that it allows for discrimination among stimulus objects without requiring too much time. Its advantage involves allocation of more or fewer units than those specified.

2.6.1.4 Q-Sort

Q-sort refers to discriminating among a relatively large number of objects quickly. This technique uses a rank order procedure in which objects are sorted into piles based on similarity with respect to some criterion. A typical example quoted in Malhotra (2004) is as follows:

Respondents are given 100 attitude statements on individual cards and asked to place them into 11 piles, ranging from „most highly agreed with“ to „least highly agreed with“. The number of objects to

be sorted should not be less than 60 nor more than 140: 60 to 90 objects is a reasonable range. The number of objects to be placed in each pile is pre-specified, often to result in a roughly normal distribution of objects over the whole set.

2.6.2 Non-Comparative Scaling Techniques

Non-comparative scales or otherwise called as nomadic scales because only one object is evaluated at a time. Researchers use this scale allowing respondents to employ whatever rating standard seems appropriate to them and not specified by the researcher.

The respondents do not compare the object being rated either to another object or to some specified standard set by the researcher. Non-comparative techniques use continuous and itemised rating scales.

In such scales, each object is scaled independently of the other objects in the stimulus set, the resulting data is generally assumed to be interval or ratio scale.

2.6.2.1 Continuous Rating Scale

This is also otherwise called as graphic rating scale. This is a type of scale that offers respondents a form of continuum (such as a line) on which to provide a rating of an object.

Researchers develop continuous rating scale allowing the respondents to indicate their rating by placing a mark at the appropriate point on a line that runs from one end of the criterion variable to the other or a set of predetermined response categories. Here the respondents need not select marks already set the researcher.

There are several variations that are possible. The line may be vertical or horizontal; it may be unmarked or marked; if marked, the divisions may be few or as many as in the thermometer scale; the scale points may be in the form of numbers or brief descriptions.

Examples of continuous rating scale

Please evaluate the service quality of a restaurant by placing an x at the position on the
horizontal line that most reflects your feelings

Empathy

The worst ----------

The best

Continuous rating scales are easy to construct, however, the scoring may be cumbersome and unreliable. With the advent of computers in research, they are increasingly used, though, they otherwise provide little new information.

2.6.2.2 Itemized Rating Scales

This scale is similar to the graphic scale in that the individuals make their judgement independently, without the benefit of direct comparison. The respondents are provided with a scale that has a number or brief description associated with each category. This scale allows the respondents to choose from a more limited number of categories, usually five to seven, although 10 or more are rarely used. The categories are ordered in terms of scale position; and the respondents are required to select the specified category that best describes the object being rated. The categories are given verbal description, although this is not absolutely necessary.

These scales are widely used in research and nowadays, more complex types such as multi-item rating scales are used. There are few variants among itemised rating scales. They are Likert, Semantic differential and stapel scales.

Likert Scale

This scale is named after Renis Likert. This is the most widely used scale in research, in particular, in testing models. Several research studies are done using Likert scale. The respondents req ire to indicate a degree of agreement of disagreement with each of a series of statements about the stimulus objects. Example of a portion of a popularly used Likert scale to measure tangibility of service is given below.

Listed below are the tangibility of service rendered by a bank is given below.

Please indicatehow strongly you agree or disagree with each by using the following scale
1 = Strongly disagree
2 = Disagree
3 = Neither agree nor disagree
3 = Agree
5 = Strongly agree

To analyse the data generated using this scale, each statement is assigned a numerical score, ranging either from -2 to +2 through a zero or 1 to 5. The analysis can be conducted item wise or a total score (summated) or a mean can be calculated for each respondent by summing or averaging across items.

It is important in Likert scale that a consistent scoring procedure so that a high score reflects favourable response and a low score reflects unfavourable response. Any deviation in the form of reverse coding where the lowest value is given to a favourable responseand highest value is given to an unfavourable response should be clearly specified by the researcher. Usually, reverse coding is used when the statements indicate a negative concept andwhen used with other statements, reverse coding would give a positive effect.

Semantic Differential Scale

Semantic differential scale is a popular scaling technique next to Likert scale. In this scale, the respondents associate their response with bipolar labels that have semantic meaning. The respondents rate objects on a number of itemised, seven point rating scales bounded at each end by one of two bipolar adjectives such as “Excellent” and “Very bad”. The respondents indicate their response choosing the one that best describes their choice.

The points are marked either from - 3 to +3 through a zero or from 1 to 7. The middle value may be treated as a neutral position. The value zero in the first type is the neutral point and 4 in the second type is the neutral point. The resulting data are commonly analysed through profile analysis. In such analysis, the means or median values on each rating scale are calculated and compared by plotting or statistical analysis. This would help the researcher to determine the overall differences and similarities among the objects.

To assess differences across segments of respondents, the researcher can compare mean responses of different segments. This data generated using this scale could be employed with summary statistics such mean, though, there is a controversy on the employment of mean on this scale. Mean is typical of Interval and ratio scales whereas this scale theoretically is an ordinal scale. However, looking beyond this objection by statisticians, researchers invariably apply all statistical techniques on this scale.

The following example illustrates semantic differential scales
1) Pleasant ---------- unpleasant
2) Aggressive ---------- submissive
3) Exciting ---------- unexciting

CHAPTER III
DATA COLLECTION

3.1 Introduction:

Methods of data collection:

Types of Data:

While deciding about the method of data collection to be used for the study, the researcher should keep in mind two types of data viz., primary and secondary.

The ***primary data*** are those which are collected afresh and for the first time, and thus happen to be original in character.

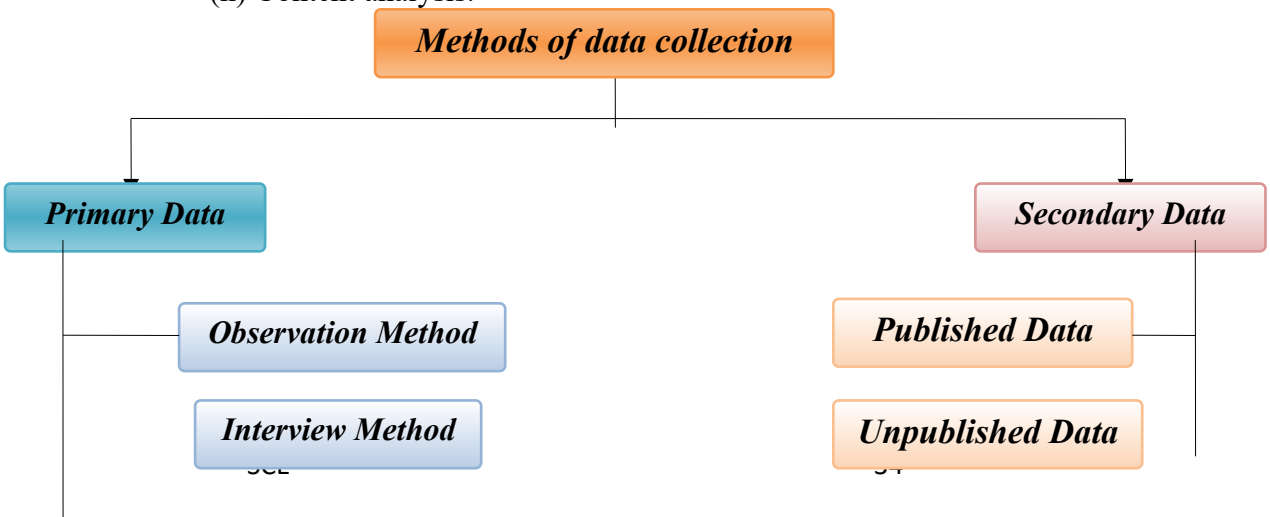
The ***secondary data,*** on the other hand, are those which have already been collected by someone else and which have already been passed through the statistical process. The researcher would have to decide which sort of data he would be using (thus collecting) for his study and accordingly he will have to select one or the other method of data collection.

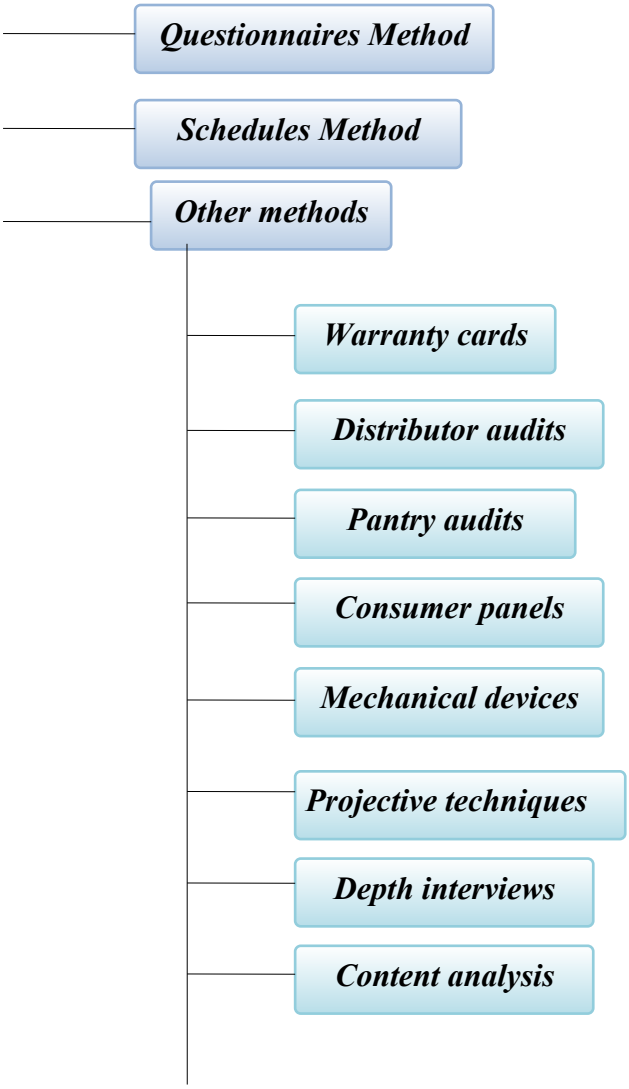
3.2 Collection of Primary Data

We collect primary data during the course of doing experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interviews.
This, in other words, means that there are several methods of collecting primary data, particularly in surveys and descriptive researches.

Important ones are:

1. Observation method,
2. Interview method,
3. Through questionnaires,
4. Through schedules, and
5. other methods which include
   - (a) Warranty cards;
   - (b) Distributor audits;
   - (c) Pantry audits;
   - (d) Consumer panels;
   - (e) Using mechanical devices;
   - (f) Through projective techniques;
   - (g) Depth interviews, and
   - (h) Content analysis.





Observation Method

The observation method is the most commonly used method specially in studies relating to behavioural sciences. In a way we all observe things around us, but this sort of observation is not scientific observation.

Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under the observation method, the information is sought by way of investigator's own direct observation without asking from the respondent.

Advantages:

a. Subjective bias is eliminated, if observation is done accurately.
b. The information obtained under this method relates to what is currently happening; it is not complicated by either the past behaviour or future intentions or attitudes.
c. It is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method.
d. This method is particularly suitable in studies which deal with subjects (i.e., respondents) who are not capable of giving verbal reports of their feelings for one reason or the other

Limitations:

a. It is an expensive method.
b. The information provided by this method is very limited.

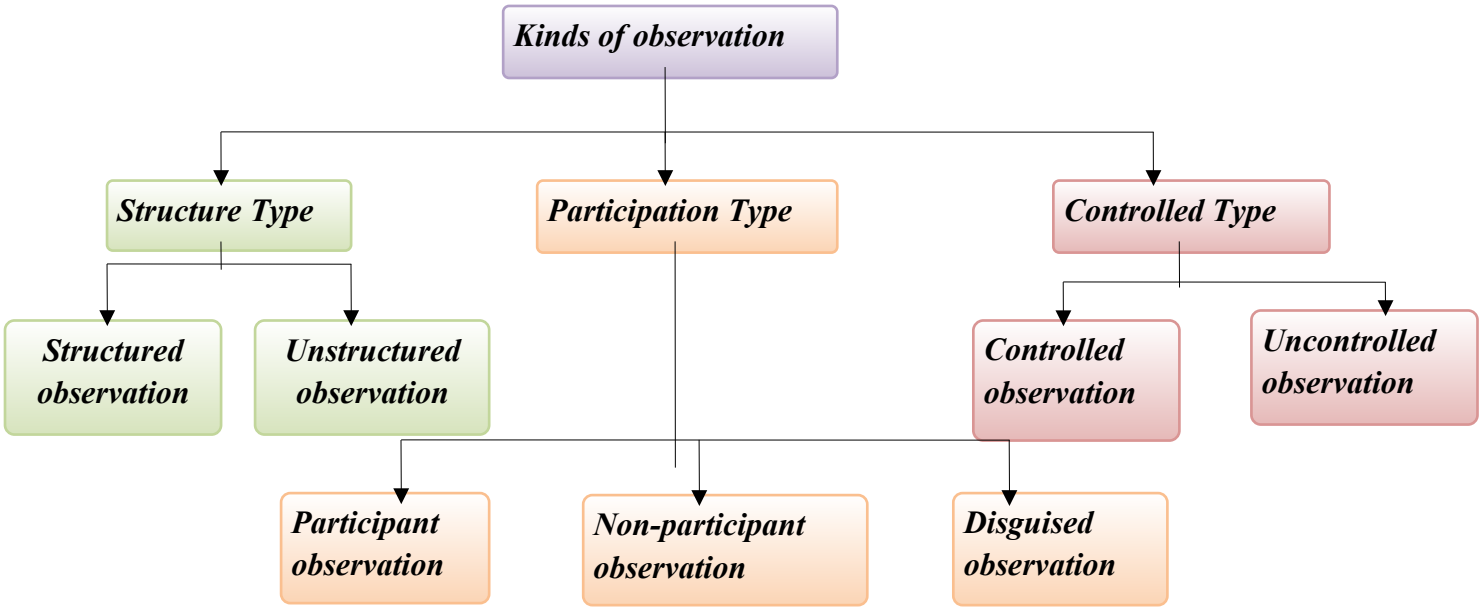
c. Sometimes unforeseen factors may interfere with the observational task.
d. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

Kinds of observation:

Structured observation
In case the observation is characterized by a careful definition of the units to be observed, the style of recording the observed information, standardized conditions of observation and the selection of pertinent data of observation, then the observation is called as *structured observation.*

Unstructured observation
When observation is to take place without these characteristics to be thought of in advance, the same is termed *as unstructured observation.*

Difference:
Structured observation is considered appropriate in descriptive studies, whereas in an exploratory study the observational procedure is most likely to be relatively unstructured.



Participant observation
If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the *participant observation.*

Merits
a. The researcher is enabled to record the natural behaviour of the group.
b. The researcher can even gather information which could not easily be obtained if he observes in a disinterested fashion.
c. The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule.

Demerits:
a. The observer may lose the objectivity to the extent he participates emotionally;
b. The problem of observation-control is not solved; and
c. It may narrow-down the researcher's range of experience.

Non-participant observation
When the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is often termed as *non-participant observation.*

Disguised observation
When the observer is observing in such a manner that his presence may be unknown to the people he is observing, such an observation is described as *disguised observation.*

Controlled observation*:*

a. When observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation.
b. Controlled observation, we use mechanical (or precision) instruments as aids to accuracy and standardization. Such observation has a tendency to supply formalized data upon which generalizations can be built with some degree of assurance.
c. Generally, controlled observation takes place in various experiments that are carried out in a laboratory or under controlled conditions.

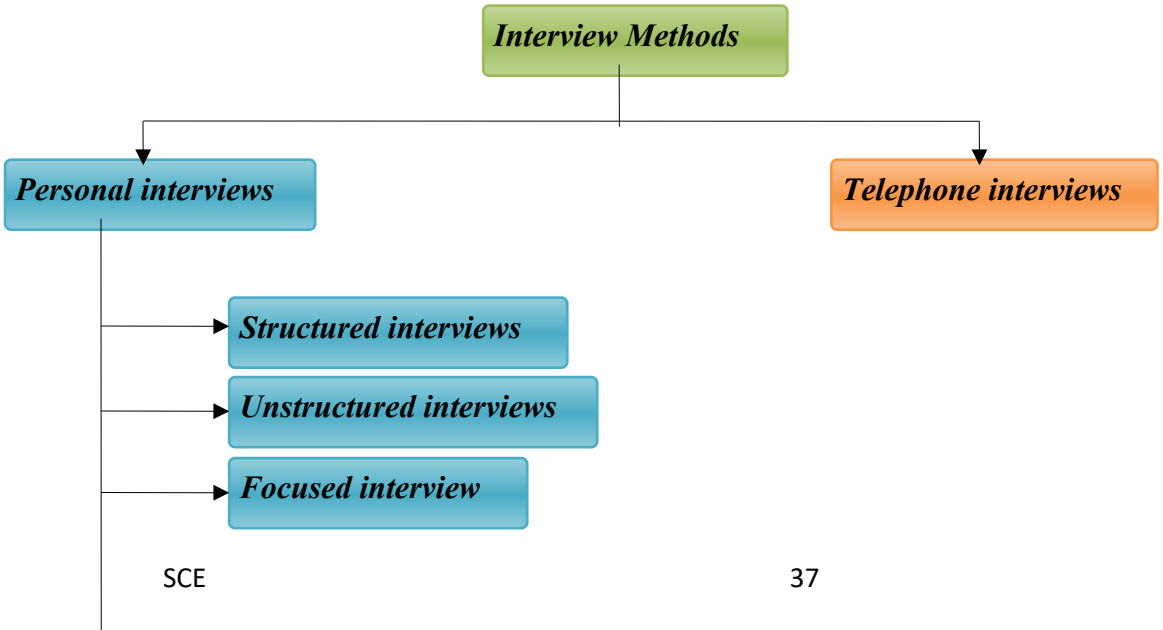
Uncontrolled observation:

a. If the observation takes place in the natural setting, it may be termed as uncontrolled observation, in non-controlled observation, no attempt is made to use precision instruments.
b. The major aim of this type of observation is to get a spontaneous picture of life and persons. It has a tendency to supply naturalness and completeness of behaviour, allowing sufficient time for observing it.
c. Uncontrolled observation is resorted to in case of exploratory researches.

Interview Method
The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses.

(a) ***Personal interviews:***

- Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons.
- This sort of interview may be in the form of direct personal investigation or it may be indirect oral investigation. In the case of direct personal investigation the interviewer has to collect the information personally from the sources concerned. He has to be on the spot and has to meet people from whom data have to be collected.
- This method is particularly suitable for intensive investigations.



Structured interviews:

- The method of collecting information through personal interviews is usually carried out in a structured way. As such we call the interviews as *structured interviews.*
- Such interviews involve the use of a set of predetermined questions and of highly standardized techniques of recording.
- In case of descriptive studies, we quite often use the technique of structured interview because of its being more economical, providing a safe basis for generalization and requiring relatively lesser skill on the part of the interviewer.

Unstructured interviews

- The *unstructured interviews* are characterized by a flexibility of approach to questioning. Unstructured interviews do not follow a system of pre-determined questions and standardized techniques of recording information.
- In a non-structured interview, the interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so requires.
- He may even change the sequence of questions.
- Unstructured interview, however, happens to be the central technique of collecting information in case of exploratory or formulative research studies.
- In case of *non-directive interview*, the interviewer's function is simply to encourage the respondent to talk about the given topic with a bare minimum of direct questioning. The interviewer often acts as a catalyst to a comprehensive expression of the respondents' feelings and beliefs and of the frame of reference within which such feelings and beliefs take on personal significance.

Focused interview

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

Clinical interview

The *clinical interview* is concerned with broad underlying feelings or motivations or with the course of individual's life experience.

The method of eliciting information under it is generally left to the interviewer's discretion.

Merits of the interview method:

- More information and that too in greater depth can be obtained.
- Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.
- There is greater flexibility under this method as the opportunity to restructure questions is always there, specially in case of unstructured interviews.
- Observation method can as well be applied to recording verbal answers to various questions.
- Personal information can as well be obtained easily under this method.
- Samples can be controlled more effectively as there arises no difficulty of the missing returns; non-response generally remains very low.

Weaknesses of the interview method:

- It is a very expensive method, specially when large and widely spread geographical sample is taken.
- There remains the possibility of the bias of interviewer as well as that of the respondent; there also remains the headache of supervision and control of interviewers.
- Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.
- This method is relatively more-time-consuming, specially when the sample is large and recalls upon the respondents are necessary.
- The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.

(b) ***Telephone interviews:***

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

Merits:

- It is more flexible in comparison to mailing method.
- It is faster than other methods i.e., a quick way of obtaining information.
- It is cheaper than personal interviewing method; here the cost per response is relatively low.
- Recall is easy; callbacks are simple and economical.
- There is a higher rate of response than what we have in mailing method; the non-response is generally very low.
- Replies can be recorded without causing embarrassment to respondents.
- Interviewer can explain requirements more easily.
- At times, access can be gained to respondents who otherwise cannot be contacted for one reason or the other.
- No field staff is required.
- Representative and wider distribution of sample is possible.

Demerits

- Little time is given to respondents for considered answers; interview period is not likely to exceed five minutes in most cases.
- Surveys are restricted to respondents who have telephone facilities.
- Extensive geographical coverage may get restricted by cost considerations.
- It is not suitable for intensive surveys where comprehensive answers are required to various questions.
- Possibility of the bias of the interviewer is relatively more.
- Questions have to be short and to the point; probes are difficult to handle.

Collection of Data through Questionnaires

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

Merits:

- There is low cost even when the universe is large and is widely spread geographically.
- It is free from the bias of the interviewer; answers are in respondents' own words.
- Respondents have adequate time to give well thought out answers.

- Respondents, who are not easily approachable, can also be reached conveniently.
- Large samples can be made use of and thus the results can be made more dependable and reliable.

Demerits:

- Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
- It can be used only when respondents are educated and cooperating.
- The control over questionnaire may be lost once it is sent.
- There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been despatched.
- There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
- It is difficult to know whether willing respondents are truly representative.
- This method is likely to be the slowest of all.

Before using this method, it is always advisable to conduct 'pilot study' (Pilot Survey) for testing the questionnaires.

Main aspects of a questionnaire:

1. ***General form:*** So far as the general form of a questionnaire is concerned, it can either be structured or unstructured questionnaire.

Structured questionnaires:

Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questions are presented with exactly the same wording and in the same order to all respondents.

Resort is taken to this sort of standardization to ensure that all respondents reply to the same set of questions. Structured questionnaires may also have fixed alternative questions in which responses of the informants are limited to the stated alternatives. Thus a highly structured questionnaire is one in which all questions and answers are specified and comments in the respondent's own words are held to the minimum.

Forms of Questions:

The form of the question may be either ***closed*** (i.e., of the type 'yes' or 'no') or ***open*** (i.e., inviting free response) but should be stated in advance and not constructed during questioning.

Merits:

- Structured questionnaires are simple to administer and relatively inexpensive to analyse.
- The provision of alternative replies, at times, helps to understand the meaning of the question clearly.

Limitations

For instance, wide range of data and that too in respondent's own words cannot be obtained with structured questionnaires. They are usually considered inappropriate in investigations where the aim happens to be to probe for attitudes and reasons for certain actions or feelings.

Unstructured or non-structured questionnaire

In an unstructured questionnaire, the interviewer is provided with a general guide on the type of information to be obtained, but the exact question formulation is largely his own responsibility and the replies are to be taken down in the respondent's own words to the extent possible; in some situations tape recorders may be used to achieve this goal.

Question sequence:

In order to make the questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chances of individual questions being misunderstood.

The question-sequence must be clear and smoothly-moving, meaning thereby that the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because they are likely to influence the attitude of the respondent and in seeking his desired cooperation. The opening questions should be such as to arouse human interest.

The following type of questions should generally be avoided as opening questions in a questionnaire:

- Questions that put too great a strain on the memory or intellect of the respondent;
- Questions of a personal character;
- Questions related to personal wealth, etc.

Question formulation and wording:

All questions should meet the following standards:

(a) Should be easily understood;

(b) Should be simple i.e., should convey only one thought at a time;

(c) Should be concrete and should conform as much as possible to the respondent's way of thinking.

Forms of questions

Concerning the form of questions, we can talk about two principal forms, viz.,

- Multiple choice questions and
- Open-end questions

Multiple choice or closed questions

The respondent selects one of the alternative possible answers put to him, whereas in the latter he has to supply the answer in his own words.

The question with only two possible answers (usually 'Yes' or 'No') can be taken as a special case of the multiple choice question, or can be named as a 'closed question.'

Advantages:

- It has the advantages of easy handling, simple to answer, quick and relatively inexpensive to analyse.
- They are most amenable to statistical analysis.
- Sometimes, the provision of alternative replies helps to make clear the meaning of the question.

Drawback

- But the main drawback of fixed alternative questions is that of "putting answers in people's mouths" i.e., they may force a statement of opinion on an issue about which the respondent does not infact having any opinion.
- They are not appropriate when the issue under consideration happens to be a complex one and also when the interest of the researcher is in the exploration of a process.

Open-ended questions:

- Open-ended questions which are designed to permit a free response from the respondent rather than one limited to certain stated alternatives are considered appropriate.
- Such questions give the respondent considerable latitude in phrasing a reply. Getting the replies in respondent's own words is, thus, the major advantage of open-ended questions.
- But one should not forget that, from an analytical point of view, open-ended questions are more difficult to handle, raising problems of interpretation, comparability and interviewer bias.

Essentials of a good questionnaire:

- To be successful, questionnaire should be comparatively short and simple i.e., the size of the questionnaire should be kept to the minimum.
- Questions should proceed in logical sequence moving from easy to more difficult questions. Personal and intimate questions should be left to the end.
- Technical terms and vague expressions capable of different interpretations should be avoided in a questionnaire.
- Questions may be dichotomous (yes or no answers), multiple choice (alternative answers listed) or open-ended.
- The latter types of questions are often difficult to analyse and hence should be avoided in a questionnaire to the extent possible.
- There should be some control questions in the questionnaire which indicate the reliability of the respondent.

Collection of Data through Schedules

- This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose.
- These enumerators along with schedules go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same in the proforma. In certain situations, schedules may be handed over to respondents and enumerators may help them in recording their answers to various questions in the said schedules. Enumerators explain the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question or the definition or concept of difficult terms.
- This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected.
- Enumerators should be intelligent and must possess the capacity of cross-examination in order to find out the truth. Above all, they should be honest, sincere, and hardworking and should have patience and perseverance.
- This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organisations. Population census all over the world is conducted through this method.

Difference between Questionnaires And Schedules

S.NO	***QUESTIONNAIRES***	***SCHEDULES***
1	It has sent through mail to informants to be answered as specified in a covering letter, but otherwise without further assistance from the sender	It is filled out by the research worker or the enumerator, who can interpret questions when necessary
2	It is relatively cheap and economical	It is relatively more expensive
3	Here no field staff required	Amount of money has to be spent in appointing enumerators and in importing training to them. Money is also spent in preparing schedules.
4	Non-response is usually high & Bias due to non-response often remains indeterminate.	Non-response is generally very low & the danger of interviewer bias and cheating
5	it is not always clear as to who replies	Identity of respondent is known
6	is likely to be very slow	The information is collected well in time as they are filled in by enumerators
7	Personal contact is generally not possible	Direct personal contact is established with respondents

8	Can be used only when respondents are literate and cooperative	The information can be gathered even when the respondents happen to be illiterate
9	Wider and more representative distribution of sample is possible	The difficulty in sending enumerators over a relatively wider area
10	Risk of collecting incomplete and wrong information is relatively more	The information collected is generally complete and accurate as enumerators can remove the difficulties, if any
11	Success of questionnaire method lies more on the quality of the questionnaire	Depends upon the honesty and competence of enumerators
12	The physical appearance of questionnaire must be quite attractive	To be filled in by enumerators and not by respondents
13	Observation method cannot used	Observation method can also be used

Some Other Methods of Data Collection

1. Warranty cards: Warranty cards are usually postal sized cards which are used by dealers of consumer durables to collect information regarding their products. The information sought is printed in the form of questions on the 'warranty cards' which is placed inside the package along with the product with a request to the consumer to fill in the card and post it back to the dealer.

2. Distributor or store audits: Distributor or store audits are performed by distributors as well as manufactures through their salesmen at regular intervals. Distributors get the retail stores audited through salesmen and use such information to estimate market size, market share, and seasonal purchasing pattern and so on. The data are obtained in such audits not by questioning but by observation.

Store audits are invariably panel operations, for the derivation of sales estimates and compilation of sales trends by stores are their principal '*raison detre*'. The principal advantage of this method is that it offers the most efficient way of evaluating the effect on sales of variations of different techniques of in-store promotion.

3. Pantry audits: Pantry audit technique is used to estimate consumption of the basket of goods at the consumer level. In this type of audit, the investigator collects an inventory of types, quantities and prices of commodities consumed. Thus in pantry audit data are recorded from the examination of consumer's pantry. The usual objective in a pantry audit is to find out what types of consumers buy certain products and certain brands, the assumption being that the contents of the pantry accurately portray consumer's preferences.

4. Consumer panels: An extension of the pantry audit approach on a regular basis is known as 'consumer panel', where a set of consumers are arranged to come to an understanding to maintain detailed daily records of their consumption and the same is made available to investigator on demands.

Mostly consume panels are of two types viz., the transitory consumer panel and the continuing consumer panel.

Transitory consumer panel: is set up to measure the effect of a particular phenomenon. Usually such a panel is conducted on a before-and-after-basis. Initial interviews are conducted before the phenomenon takes place to record the attitude of the consumer. A second set of interviews is carried out after the phenomenon has taken place to find out the consequent changes that might have occurred in the consumer's attitude. It is a favourite tool of advertising and of social research.

***Continuing consumer panel*:** is often set up for an indefinite period with a view to collect data on a particular aspect of consumer behaviour over time, generally at periodic intervals or may be meant to serve as a general purpose panel for researchers on a variety of subjects. Such panels have been used in the area of consumer expenditure, public opinion and radio and TV listenership among others. Most of these panels operate by mail. The representativeness of the panel relative to the population and the effect of panel membership on the information obtained after the two major problems associated with the use of this method of data collection.

5. Use of mechanical devices: The use of mechanical devices has been widely made to collect information by way of indirect means. Eye camera, Pupilometric camera, Psychogalvano-meter, Motion picture camera and Audiometer are the principal devices so far developed and commonly used by modern big business houses, mostly in the developed world for the purpose of collecting the required information.

Eye cameras are designed to record the focus of eyes of a respondent on a specific portion of a sketch or diagram or written material. Such an information is useful in designing advertising material.

Pupilometric cameras record dilation of the pupil as a result of a visual stimulus. The extent of dilation shows the degree of interest aroused by the stimulus.

Psychogalvano meter is used for measuring the extent of body excitement as a result of the visual stimulus.

Motion picture cameras can be used to record movement of body of a buyer while deciding to buy a consumer good from a shop or big store. Influence of packaging or the information given on the label would stimulate a buyer to perform certain physical movements which can easily be recorded by a hidden motion picture camera in the shop's four walls.

Audiometers are used by some TV concerns to find out the type of programmes as well as stations preferred by people. A device is fitted in the television instrument itself to record these changes. Such data may be used to find out the market share of competing television stations

6. Projective techniques (indirect interviewing techniques): Projective techniques for the collection of data have been developed by psychologists to use projections of respondents for inferring about underlying motives, urges, or intentions which are such that the respondent either resists to reveal them or is unable to figure out himself.

In projective techniques the respondent in supplying information tends unconsciously to project his own attitudes or feelings on the subject under study. Projective techniques play an important role in motivational researches or in attitude surveys.

The use of these techniques requires intensive specialised training. In such techniques, the individual's responses to the stimulus-situation are not taken at their face value.

Important projective techniques

- *Word association tests*
- *Sentence completion tests*
- *Story completion tests*
- *Verbal projection tests*
- *Pictorial techniques*
- *Play techniques*
- *Quizzes*, *tests and examinations*
- *Sociometry*

Word association tests: These tests are used to extract information regarding such words which have maximum association. In this sort of test the respondent is asked to mention the first word that comes to mind, ostensibly without thinking, as the interviewer reads out each word from a list. If the interviewer says *cold*, the respondent may say *hot* and the like ones. The general technique is to use a list of as many as 50 to 100 words.

Sentence completion tests: These tests happen to be an extension of the technique of word association tests. This technique permits the testing not only of words (as in case of word association tests), but of ideas as well and thus, helps in developing hypotheses and in the construction of questionnaires. This technique is also quick and easy to use, but it often leads to analytical problems, particularly when the response happens to be multidimensional.

Story completion tests: Such tests are a step further wherein the researcher may contrive stories instead of sentences and ask the informants to complete them. The respondent is given just enough of story to focus his attention on a given subject and he is asked to supply a conclusion to the story.

Verbal projection tests: These are the tests wherein the respondent is asked to comment on or to explain what other people do.

Pictorial techniques: There are several pictorial techniques. The important ones are as follows:
(a) ***Thematic apperception test (T.A.T.):*** The TAT consists of a set of pictures (some of the pictures deal with the ordinary day-to-day events while others may be ambiguous pictures of unusual situations) that are shown to respondents who are asked to describe what they think the pictures represent. The replies of respondents constitute the basis for the investigator to draw inferences about their personality structure, attitudes, etc.

(b) ***Rosenzweig test:*** This test uses a cartoon format wherein we have a series of cartoons with words inserted in 'balloons' above. The respondent is asked to put his own words in an empty balloon space provided for the purpose in the picture. From what the respondents write in this fashion, the study of their attitudes can be made.
(c) ***Rorschach test:*** This test consists of ten cards having prints of inkblots. The design happens to be symmetrical but meaningless. The respondents are asked to describe what they perceive in such symmetrical inkblots and the responses are interpreted on the basis of some pre-determined psychological framework. This test is frequently used but the problem of validity still remains a major problem of this test.

(d) ***Holtzman Inkblot Test (HIT):*** This test from W.H. Holtzman is a modification of the Rorschach Test explained above. This test consists of 45 inkblot cards (and not 10 inkblots as we find in case of Rorschach Test) which are based on colour, movement, shading and other factors involved in inkblot perception. Only one response per card is obtained from the subject (or the respondent) and the responses of a subject are interpreted at three levels of form appropriateness.

(e) ***Tomkins-Horn picture arrangement test****:* This test is designed for group administration. It consists of twenty-five plates, each containing three sketches that may be arranged in different ways to portray sequence of events. The respondent is asked to arrange them in a sequence which he considers as reasonable. The responses are interpreted as providing evidence confirming certain norms, respondent's attitudes, etc.

Play techniques: Under play techniques subjects are asked to improvise or act out a situation in which they have been assigned various roles. The researcher may observe such traits as hostility, dominance, sympathy, prejudice or the absence of such traits. These techniques have been used for knowing the attitudes of younger ones through manipulation of dolls. Dolls representing different racial groups are usually given to children who are allowed to play with them freely. The manner in which children organise dolls would indicate their attitude towards the class of persons represented by dolls. This is also known as *doll-play test*, and is used frequently in studies pertaining to sociology.

The choice of colour, form, words, the sense of orderliness and other reactions may provide opportunities to infer deep-seated feelings.

Quizzes, tests and examinations: This is also a technique of extracting information regarding specific ability of candidates indirectly. In this procedure both long and short questions are framed to test through them the memorising and analytical ability of candidates.

Sociometry: is a technique for describing the social relationships among individuals in a group. In an indirect way, sociometry attempts to describe attractions or repulsions between individuals by asking them to indicate whom they would choose or reject in various situations.

7. *Depth interviews*:

- Depth interviews are those interviews that are designed to discover underlying motives and desires and are often used in motivational research. Such interviews are held to explore needs, desires and feelings of respondents.
- Depth interviews require great skill on the part of the interviewer and at the same time involve considerable time. Unless the researcher has specialised training, depth interviewing should not be attempted.
- Depth interview may be projective in nature or it may be a non-projective interview. The difference lies in the nature of the questions asked. Indirect questions on seemingly irrelevant subjects provide information that can be related to the informant's behaviour or attitude towards the subject under study.
- Depth interviews do not necessarily have to be projective in nature; even non-projective depth interviews can reveal important aspects of psycho-social situation for understanding the attitudes of people.

8. Content-analysis:

Content-analysis consists of analysing the contents of documentary materials such as books, magazines, newspapers and the contents of all other verbal materials which can be either spoken or printed.

Bernard Berelson's name is often associated with the latter type of content analysis. "Content-analysis is measurement through proportion…. Content analysis measures pervasiveness and that is sometimes an index of the intensity of the force."

3.3 Secondary Data

Secondary data are the data that are in actual existence in accessible records, having been already collected and treated statistically by the persons maintaining the records. In other words, secondary data are the data that have been already collected, presented tabulated, treated with necessary statistical techniques and conclusions have been drawn. Therefore, collecting secondary data doesn't mean doing some original enumeration but it merely means obtaining data that have already been collected by some agencies, reliable persons, government departments, research

workers, dependable organisations etc. Secondary data are easily obtainable from reliable records, books, government publications and journals.

When once primary data have been originally collected, moulded by statisticians or statistical machinery, then it becomes secondary in the hands of all other persons who may be desirous of handling it for their own purpose or studies. It follows, therefore, that primary and secondary data are demarcated separately and that the distinction between them is of degree only. It a person 'X' collects some data originally, then the data is primary data to 'X' whereas the same data when used by another person 'Y' becomes secondary data to 'Y'.

3.3 Sources Of Secondary Data

The following are some of the sources of secondary data:

1. Central and State government publications.
2. Publications brought out by international organisation like the UNO, UNESCO, etc.
3. Foreign government publications.
4. Official publications as well as reports of municipalities, district parishads, etc.
5. Reports and publications of commissions - like U.G.C. education commission, tariff commission, chambers of commerce, co-operative societies, trade associations, banks, stock exchanges, business houses etc.
6. Well-know newspapers and journals like the *Economic Times, The Financial Express, Indian Journal of Economics, Commerce, Capital, Economical Eastern Economist,* etc.

7. Publications brought out by research institutions, universities as well as those published by research workers give considerable secondary data.
8. Through the Internet/website sources.

Though the given list of secondary data cannot be said to be thorough or complete, yet it can be pointed out that it fairly indicates the chief sources of secondary data. Also, besides the above mentioned data there are a number of other important sources, such as records of governments in various departments, unpublished manuscripts of eminent scholars, research workers, statisticians, economists, private organisations, labour bureaus and records of business firms.

Types Of Secondary Data

Secondary data are of two types. Data that are originated from within the company are called as internal data. If they are collected for some other purpose, they are internal secondary data. This poses significant advantage as they are readily available in the company at low cost. The most convenient example internal secondary data is the figures relating sales of the company.

Important internal source of secondary data is database marketing,
Database marketing involves the use of computers to capture and track customer profiles and purchase details. The information about customer profile would serve as the foundation for marketing programmes or as an internal source of information related to preference of customer‟s preference of a particular product.

Published external secondary data refers to the data available without the company. There is such a pool of published data available in the market that it is sometimes easy to underestimate what is available and thereby bypass relevant information. Several sources of external data are available.
They are:

General Business Data

Guides or small booklets containing information about a particular trade or business. Directories are helpful for identifying individuals or organisations that collect specific data. Indexes used to locate information on a particular topic in several different publications by using an index.

Non-governmental statistical data refers to published statistical data of great interest to researchers. Graphic and statistical analyses can be performed on these data to draw meaning inference.

Government Sources

Census data is a report published by the Government containing information about the population of the country. Other Government publications may be pertaining to availability of train tickets just before it leaves.

Computerised Databases

Online databases are databases consisting of data pertaining to a particular sector (e.g., banks) that is accessed with a computer through a telecommunication network

- Internet databases are available in internet portals that can be accessed, searched, and analysed on the internet.
- Offline databases are databases available in the form of diskettes and CD-ROM disks.

Bibliographic databases comprises of citations in articles published in journals, magazines, newspapers etc.

Numeric databases contain numerical and statistical information. For example, time series data about stock markets.

Directory databases provide information on individuals, organisations and service. E.g. Getit Yellow pages.

Special-purpose databases are databases developed online for a special purpose.

External Data-syndicated

In response to the growing need for data pertaining to markets, consumer etc., companies have started collecting and selling standardised data designed to serve the information needs of the shared by a number of organisations. Syndicated data sources can be urther classified as (a) consumer data (b) retail data (c) wholesale data (d) industrial data (e) advertising evaluation data and (f) media and audience data.

Consumer data relates to data about consumers purchases and the circumstances surrounding the purchase.

Retail data rely on retailing establishments for their data. The data collected focus on the products or services sold through the outlets and / or the characteristics of the outlets themselves.

- Wholesale data refers to data on warehouse shipment data to estimate sales at retail.
- Industrial data refers to substantially more syndicated data services available to consumer goods manufacturers than to industrial goods suppliers.
- Advertising evaluation data refers to money spent each year on media such as magazines and television with the expectation that these expenditures will result in sales.

Collection of Secondary Data

Secondary data means data that are already available i.e., they refer to the data which have already been collected and analysed by someone else.

Secondary data may either be published data or unpublished data.

Published data

Usually published data are available in:

(a) Various publications of the central, state are local governments;

(b) Various publications of foreign governments or of international bodies and their subsidiary organizations;
(c) Technical and trade journals;
(d) Books, magazines and newspapers;
(e) Reports and publications of various associations connected with business and industry, banks, stock exchanges, etc.;
(f) Reports prepared by research scholars, universities, economists, etc. in different fields; and
(g)Public records and statistics, historical documents, and other sources of published information.

Unpublished data
The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labour bureaus and other public/ private individuals and organisations.

Researcher must be very careful in using secondary data. He must make a minute scrutiny because it is just possible that the secondary data may be unsuitable or may be inadequate in the context of the problem which the researcher wants to study.
In this connection Dr. A.L. Bowley very aptly observes that it is never safe to take published statistics at their face value without knowing their meaning and limitations and it is always necessary to criticise arguments that can be based on them.
The researcher, before using secondary data, must see that they possess following characteristics:

Reliability of data:
Suitability of data: The data that are suitable for one enquiry may not necessarily be found suitable in another enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher.
Adequacy of data: If the level of accuracy achieved in data is found inadequate for the purpose of the present enquiry, they will be considered as inadequate and should not be used by the researcher.
The data will also be considered inadequate, if they are related to an area which may be either narrower or wider than the area of the present enquiry.

Selection of Appropriate Method For Data Collection
Thus, there are various methods of data collection. As such the researcher must judiciously select the method/methods for his own study, keeping in view the following factors:

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

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

Time factor:

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

Precision required:

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

Case Study Method

Meaning*:*

- The case study method is a very popular form of qualitative analysis and involves a careful and complete observation of a social unit, be that unit a person, a family, an institution, a cultural group or even the entire community.
- It is a method of study in depth rather than breadth. The case study places more emphasis on the full analysis of a limited number of events or conditions and their interrelations.
- The case study deals with the processes that take place and their interrelationship.
- Thus, case study is essentially an intensive investigation of the particular unit under consideration.
- The object of the case study method is to locate the factors that account for the behavior patterns of the given unit as an integrated totality.
- Pauline V. Young describes case study as "a comprehensive study of a social unit be that unit a person, a group, a social institution, a district or a community."

Characteristics:

- Under this method the researcher can take one single social unit or more of such units for his study purpose; he may even take a situation to study the same comprehensively.
- Here the selected unit is studied intensively i.e., it is studied in minute details.
- In the context of this method we make complete study of the social unit covering all facets.
- Under this method the approach happens to be qualitative and not quantitative. Mere quantitative information is not collected. Every possible effort is made to collect information concerning all aspects of life. As such, case study deepens our perception and gives us a clear insight into life.
- In respect of the case study method an effort is made to know the mutual inter-relationship of causal factors.
- Under case study method the behaviour pattern of the concerning unit is studied directly and not by an indirect and abstract approach.
- Case study method results in fruitful hypotheses along with the data which may be helpful in testing them, and thus it enables the generalised knowledge to get richer and richer. In its absence, generalised social science may get handicapped.

Assumptions:

- The assumption of uniformity in the basic human nature in spite of the fact that human behaviour may vary according to situations.
- The assumption of studying the natural history of the unit concerned.
- The assumption of comprehensive study of the unit concerned.

Major phases involved:

- Recognition and determination of the status of the phenomenon to be investigated or the unit of attention.
- Collection of data, examination and history of the given phenomenon.
- Diagnosis and identification of causal factors as a basis for remedial or developmental treatment.
- Application of remedial measures i.e., treatment and therapy (this phase is often characterized as case work).
- Follow-up programme to determine effectiveness of the treatment applied.

Advantages:

- Charles Horton Cooley, "case study deepens our perception and gives us a clearer insight into life…. It gets at behaviour directly and not by an indirect and abstract approach."
- Through case study a researcher can obtain a real and enlightened record of personal experiences which would reveal man's inner strivings, tensions and motivations that drive him to action along with the forces that direct him to adopt a certain pattern of behaviour.
- This method enables the researcher to trace out the natural history of the social unit and its relationship with the social factors and the forces involved in its surrounding environment.
- It helps in formulating relevant hypotheses along with the data which may be helpful in testing them. Case studies, thus, enable the generalised knowledge to get richer and richer.
- The method facilitates intensive study of social units which is generally not possible if we use either the observation method or the method of collecting information through schedules. This is the reason why case study method is being frequently used, particularly in social researches.
- Information collected under the case study method helps a lot to the researcher in the task of constructing the appropriate questionnaire or schedule for the said task requires thorough knowledge of the concerning universe.
- The researcher can use one or more of the several research methods under the case study method depending upon the prevalent circumstances. In other words, the use of different methods such as depth interviews, questionnaires, documents, study reports of individuals, letters, and the like is possible under case study method.
- Case study method has proved beneficial in determining the nature of units to be studied along with the nature of the universe. This is the reason why at times the case study method is alternatively known as "mode of organising data".
- This method is a means to well understand the past of a social unit because of its emphasis of historical analysis. Besides, it is also a technique to suggest measures for improvement in the context of the present environment of the concerned social units.

Limitations:

- Case situations are seldom comparable and as such the information gathered in case studies is often not comparable. Since the subject under case study tells history in his own words, logical concepts and units of scientific classification have to be read into it or out of it by the investigator.
- Read Bain does not consider the case data as significant scientific data since they do not provide knowledge of the "impersonal, universal, non-ethical, non-practical, repetitive aspects of phenomena."
- Real information is often not collected because the subjectivity of the researcher does enter in the collection of information in a case study.

- The danger of false generalisation is always there in view of the fact that no set rules are followed in collection of the information and only few units are studied.
- It consumes more time and requires lot of expenditure. More time is needed under case study method since one studies the natural history cycles of social units and that too minutely.
- The case data are often vitiated because the subject, according to Read Bain, may write what he thinks the investigator wants; and the greater the rapport, the more subjective the whole process is.
- Case study method is based on several assumptions which may not be very realistic at times, and as such the usefulness of case data is always subject to doubt.
- Case study method can be used only in a limited sphere. It is not possible to use it in case of a big society. Sampling is also not possible under a case study method.
- Response of the investigator is an important limitation of the case study method. He often thinks that he has full knowledge of the unit and can himself answer about it.

3.4 Questionnaires are frequently used in quantitative marketing research and social research. They are a valuable method of collecting a wide range of information from a large number of individuals, often referred to as respondents.

Adequate **questionnaire construction** is critical to the success of a survey. Inappropriate questions, incorrect ordering of questions, incorrect scaling, or bad questionnaire format can make the survey valueless, as it may not accurately reflect the views and opinions of the participants. A useful method for checking a questionnaire and making sure it is accurately capturing the intended information is to pretest among a smaller subset of target respondents.

Questionnaire construction issues

- Know how (and whether) you will use the results of your research before you start. If, for example, the results won't influence your decision or you can't afford to implement the findings or the cost of the research outweighs its usefulness, then save your time and money; don't bother doing the research.
- The research objectives and frame of reference should be defined beforehand, including the questionnaire's context of time, budget, manpower, intrusion and privacy.
- How (randomly or not) and from where (your sampling frame) you select the respondents will determine whether you will be able to generalize your findings to the larger population.
- The nature of the expected responses should be defined and retained for interpretation of the responses, be it preferences (of products or services), facts, beliefs, feelings, descriptions of past behavior, or standards of action.
- Unneeded questions are an expense to the researcher and an unwelcome imposition on the respondents. All questions should contribute to the objective(s) of the research.
- If you "research backwards" and determine what you want to say in the report (i.e., Package A is more/less preferred by X% of the sample vs. Package B, and y% compared to Package C) then even though you don't know the exact answers yet, you will be certain to ask all the questions you need - and only the ones you need - in such a way (metrics) to write your report.

- The topics should fit the respondents' frame of reference. Their background may affect their interpretation of the questions. Respondents should have enough information or expertise to answer the questions truthfully.
- The type of scale, index, or typology to be used shall be determined.
- The level of measurement you use will determine what you can do with and conclude from the data. If the response option is yes/no then you will only know how many or what percent of your sample answered yes/no. You cannot, however, conclude what the average respondent answered.

- The types of questions (closed, multiple-choice, open) should fit the statistical data analysis techniques available and your goals.

- Questions and prepared responses to choose from should be neutral as to intended outcome. A biased question or questionnaire encourages respondents to answer one way rather than another.Even questions without bias may leave respondents with expectations.
- The order or "natural" grouping of questions is often relevant. Prior previous questions may bias later questions.
- The wording should be kept simple: no technical or specialized words.
- The meaning should be clear. Ambiguous words, equivocal sentence structures and negatives may cause misunderstanding, possibly invalidating questionnaire results. Double negatives should be reworded as positives.

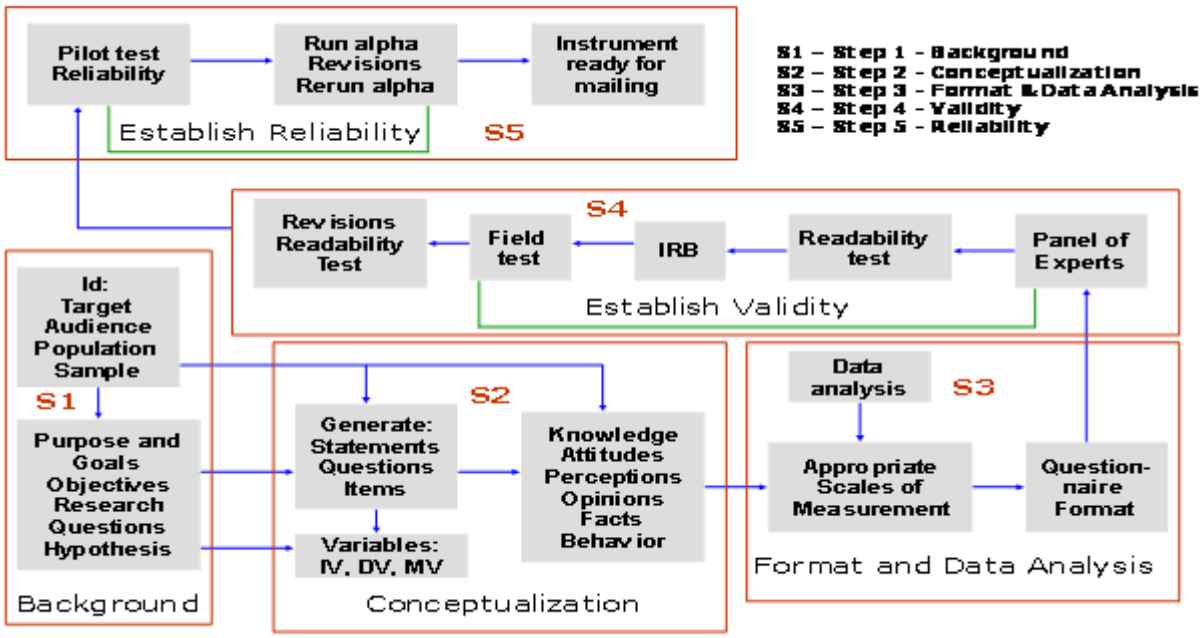
- If a survey question actually contains more than one issue, the researcher will not know which one the respondent is answering. Care should be taken to ask one question at a time.
- The list of possible responses should be collectively exhaustive. Respondents should not find themselves with no category that fits their situation. One solution is to use a final category for "other _____".
- The possible responses should also be mutually exclusive. Categories should not overlap. Respondents should not find themselves in more than one category, for example in both the "married" category and the "single" category - there may be need for separate questions on marital status and living situation.

- Writing style should be conversational, yet concise and accurate and appropriate to the target audience.
- Many people will not answer personal or intimate questions. For this reason, questions about age, income, marital status, etc. are generally placed at the end of the survey. This way, even if the respondent refuses to answer these "personal" questions, he/she will have already answered the research questions.
- "Loaded" questions evoke emotional responses and may skew results.

- Presentation of the questions on the page (or computer screen) and use of white space, colors, pictures, charts, or other graphics may affect respondent's interest or distract from the questions.
- Numbering of questions may be helpful.

Questionnaires can be administered by research staff, by volunteers or self-administered by the respondents. Clear, detailed instructions are needed in either case, matching the needs of each audience. Development of a valid and reliable questionnaire involves several steps taking considerable time. This article describes the sequential steps involved in the development and testing of questionnaires used for data collection. Figure 1 illustrates the five sequential steps involved in questionnaire development and testing. Each step depends on fine tuning and testing of previous steps that must be completed before the next step.

1.Sequence for Questionnaire/Instrument Development



Step 1--Background

In this initial step, the purpose, objectives, research questions, and hypothesis of the proposed research are examined. Determining who is the audience, their background, especially their educational/readability levels, access, and the process used to select the respondents (sample vs. population) are also part of this step. A thorough understanding of the problem through literature search and readings is a must. Good preparation and understanding of Step1 provides the foundation for initiating Step 2.

Step 2--Questionnaire Conceptualization

After developing a thorough understanding of the research, the next step is to generate statements/questions for the questionnaire. In this step, content (from literature/theoretical framework) is transformed into statements/questions. In addition, a link among the objectives of the study and their translation into content is established. For example, the researcher must indicate what the questionnaire is measuring, that is, knowledge, attitudes, perceptions, opinions, recalling facts, behavior change, etc. Major variables (independent, dependent, and moderator variables) are identified and defined in this step.

Step 3--Format and Data Analysis

In Step 3, the focus is on writing statements/questions, selection of appropriate scales of measurement, questionnaire layout, format, question ordering, font size, front and back cover, and proposed data analysis. Scales are devices used to quantify a subject's response on a particular variable. Understanding the relationship between the level of measurement and the appropriateness of data analysis is important. For example, if ANOVA (analysis of variance) is one mode of data analysis, the independent variable must be measured on a nominal scale with two or more levels (yes, no, not sure), and the dependent variable must be measured on a interval/ratio scale (strongly agree to strongly disagree).

Step 4--Establishing Validity

As a result of Steps 1-3, a draft questionnaire is ready for establishing validity. Validity is the amount of systematic or built-in error in measurement (Norland, 1990). Validity is established using a panel

of experts and a field test. Which type of validity (content, construct, criterion, and face) to use depends on the objectives of the study. The following questions are addressed in Step 4:

1. Is the questionnaire valid? In other words, is the questionnaire measuring what it intended to measure?
2. Does it represent the content?
3. Is it appropriate for the sample/population?
4. Is the questionnaire comprehensive enough to collect all the information needed to address the purpose and goals of the study?
5. Does the instrument look like a questionnaire?

Addressing these questions coupled with carrying out a readability test enhances questionnaire validity. The Fog Index, Flesch Reading Ease, Flesch-Kinkaid Readability Formula, and Gunning-Fog Index are formulas used to determine readability. Approval from the Institutional Review Board (IRB) must also be obtained. Following IRB approval, the next step is to conduct a field test using subjects not included in the sample. Make changes, as appropriate, based on both a field test and expert opinion. Now the questionnaire is ready to pilot test.

Step 5--Establishing Reliability

In this final step, reliability of the questionnaire using a pilot test is carried out. Reliability refers to random error in measurement. Reliability indicates the accuracy or precision of the measuring instrument (Norland, 1990). The pilot test seeks to answer the question, does the questionnaire consistently measure whatever it measures?

The use of reliability types (test-retest, split half, alternate form, internal consistency) depends on the nature of data (nominal, ordinal, interval/ratio). For example, to assess reliability of questions measured on an interval/ratio scale, internal consistency is appropriate to use. To assess reliability of knowledge questions, test-retest or split-half is appropriate.

Reliability is established using a pilot test by collecting data from 20-30 subjects not included in the sample. Data collected from pilot test is analyzed using SPSS (Statistical Package for Social Sciences) or another software. SPSS provides two key pieces of information. These are "correlation matrix" and "view alpha if item deleted" column. Make sure that items/statements that have 0s, 1s, and negatives are eliminated. Then view "alpha if item deleted" column to determine if alpha can be raised by deletion of items. Delete items that substantially improve reliability. To preserve content, delete no more than 20% of the items. The reliability coefficient (alpha) can range from 0 to 1, with 0 representing an instrument with full of error and 1 representing total absence of error. A reliability coefficient (alpha) of .70 or higher is considered acceptable reliability.

3.5 Sampling Fundamentals

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

3.5.1 Need For Sampling

Sampling is used in practice for a variety of reasons such as:

- Sampling can save time and money. A sample study is usually less expensive than a census study and produces results at a relatively faster speed.
- Sampling may enable more accurate measurements for a sample study is generally conducted by trained and experienced investigators.
- Sampling remains the only way when population contains infinitely many members.

- Sampling remains the only choice when a test involves the destruction of the item under study.
- Sampling usually enables to estimate the sampling errors and, thus, assists in obtaining information concerning some characteristic of the population.

3.5.2 Fundamental Definitions

Universe: From a statistical point of view, the term 'Universe'refers to the total of the items or units in any field of inquiry.

Population: the term 'population' refers to the total of items about which information is desired.

Elementary units: The attributes that are the object of study are referred to as characteristics and the units possessing them are called as elementary units. Quit often, we do not find any difference between population and universe, and as such the two terms are taken as interchangeable. However, a researcher must necessarily define these terms precisely.

The population or universe can be *finite* or *infinite*.

Finite population: The population is said to be finite if it consists of a fixed number of elements so that it is possible to enumerate it in its totality.

Example: For instance, the population of a city, the number of workers in a factory are examples of finite populations.

Infinite population: is that population in which it is theoretically impossible to observe all the elements. Thus, in an infinite population the number of items is infinite i.e., we cannot have any idea about the total number of items.

Example: The number of stars in a sky, possible rolls of a pair of dice are examples of infinite population.

Sampling frame: The elementary units or the group or cluster of such units may form the basis of sampling process in which case they are called as sampling units. A list containing all such sampling units is known as sampling frame.

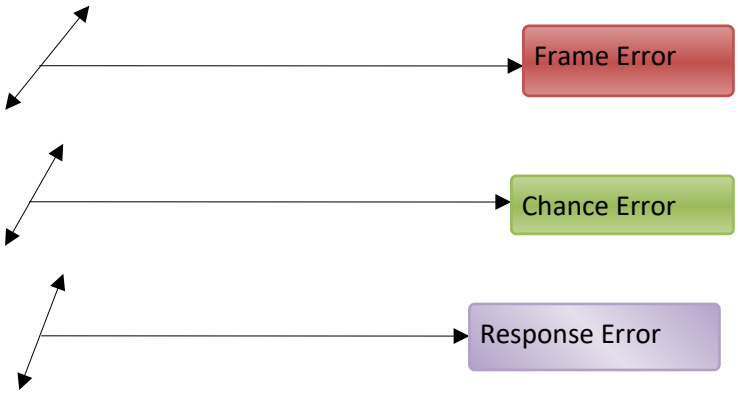
In most cases they are not identical because it is often impossible to draw a sample directly from population. As such this frame is either constructed by a researcher for the purpose of his study or may consist of some existing list of the population.

Sampling design: is a definite plan for obtaining a sample from the sampling frame. It refers to the technique or the procedure the researcher would adopt in selecting some sampling units from which inferences about the population is drawn. Sampling design is determined before any data are collected.

Statisitc(s) and parameter(s): A statistic is a characteristic of a sample, whereas a parameter is a characteristic of a population. Thus, when we work out certain measures such as mean, median, mode or the like ones from samples, then they are called statistic(s) for they describe the characteristics of a sample. But when such measures describe the characteristics of a population, they are known as parameter(s).

For instance, the population mean (μ) is a parameter, whereas the sample mean ($\bar{X}$) is a statistic. To obtain the estimate of a parameter from a statistic constitutes the prime objective of sampling analysis.

Sampling error: Sample surveys do imply the study of a small portion of the population and as such there would naturally be a certain amount of inaccuracy in the information collected. This inaccuracy may be termed as sampling error or error variance.



Sampling error = Frame error + Chance error + Response error

Sampling error is inversely related to the size of the sample i.e., sampling error decreases as the sample size increases and vice-versa.

A measure of the random sampling error can be calculated for a given sample design and size and this measure is often called the precision of the sampling plan. Sampling error is usually worked out as the product of the critical value at a certain level of significance and the standard error.

Non -sampling errors which may creep in during the process of collecting actual information and such errors occur in all surveys whether census or sample. We have no way to measure non-sampling errors.

Precision: Precision is the range within which the population average (or other parameter) will lie in accordance with the reliability specified in the confidence level as a percentage of the estimate ± or as a numerical quantity.

Confidence level and significance level: The confidence level or reliability is the expected percentage of times that the actual value will fall within the stated precision limits. Thus, if we take a confidence level of 95%, then we mean that there are 95 chances in 100 (or .95 in 1) that the sample results represent the true condition of the population within a specified precision range against 5 chances in 100 (or .05 in 1) that it does not. Precision is the range within which the answer may vary and still be acceptable; confidence level indicates the likelihood that the answer will fall within that range, and the significance level indicates the likelihood that the answer will fall outside that range.

We can always remember that if the confidence level is 95%, then the significance level will be (100 – 95) i.e., 5%; if the confidence level is 99%, the significance level is (100 – 99) i.e., 1%, and so on.

Sampling distribution: We are often concerned with sampling distribution in sampling analysis.

If we take certain number of samples and for each sample compute various statistical measures such as mean, standard deviation, etc., then we can find that each sample may give its own value for the statistic under consideration.

We can have sampling distribution of mean, or the sampling distribution of standard deviation or the sampling distribution of any other statistical measure. It may be noted that each item in a sampling

distribution is a particular statistic of a sample. The sampling distribution tends quite closer to the normal distribution if the number of samples is large.

Important Sampling Distributions

Some important sampling distributions, which are commonly used, are: (1) sampling distribution of mean; (2) sampling distribution of proportion; (3) student's't' distribution; (4) *F* distribution; and (5) Chi-square distribution. A brief mention of each one of this sampling distribution will be helpful.

3.6 Sample Size and Its Determination

In sampling analysis the most ticklish question is: What should be the size of the sample or how large or small should be '*n*'? If the sample size ('*n*') is too small, it may not serve to achieve the objectives and if it is too large, we may incur huge cost and waste resources. As a general rule, one can say that the sample must be of an optimum size i.e., it should neither be excessively large nor too small.

Technically, the sample size should be large enough to give a confidence interval of desired width and as such the size of the sample must be chosen by some logical process before sample is taken from the universe. Size of the sample should be determined by a researcher keeping in view the following points:

(i) ***Nature of universe:*** Universe may be either homogenous or heterogeneous in nature. If the items of the universe are homogenous, a small sample can serve the purpose. But if the items are heteogenous, a large sample would be required. Technically, this can be termed as the dispersion factor.

(ii) ***Number of classes proposed:*** If many class-groups (groups and sub-groups) are to be formed, a large sample would be required because a small sample might not be able to give a reasonable number of items in each class-group.

(iii) ***Nature of study:*** If items are to be intensively and continuously studied, the sample should be small. For a general survey the size of the sample should be large, but a small sample is considered appropriate in technical surveys.

(iv) ***Type of sampling:*** Sampling technique plays an important part in determining the size of the sample. A small random sample is apt to be much superior to a larger but badly selected sample.

(v) ***Standard of accuracy and acceptable confidence level:*** If the standard of accuracy or the level of precision is to be kept high, we shall require relatively larger sample. For doubling the accuracy for a fixed significance level, the sample size has to be increased fourfold.

(vi) ***Availability of finance:*** In practice, size of the sample depends upon the amount of money available for the study purposes. This factor should be kept in view while determining the size of sample for large samples result in increasing the cost of sampling estimates.

(vii) ***Other considerations:*** Nature of units, size of the population, size of questionnaire, availability of trained investigators, the conditions under which the sample is being conducted, the time available for completion of the study are a few other considerations to which a researcher must pay attention while selecting the size of the sample.

There are two alternative approaches for determining the size of the sample. The first approach is "to specify the precision of estimation desired and then to determine the sample size necessary to insure it" and the second approach "uses Bayesian statistics to weigh the cost of additional information against the expected value of the additional information." The first approach is capable of giving a mathematical solution, and as such is a frequently used technique of determining '*n*'.

The limitation of this technique is that it does not analyze the cost of gathering information *vis-a vis* the expected value of information. The second approach is theoretically optimal, but it is seldom used

because of the difficulty involved in measuring the value of information. Hence, we shall mainly concentrate here on the first approach.

3.7 Sampling Design

Census and Sample Survey

All items in any field of inquiry constitute a '***Universe' or 'Population***.' A complete enumeration of all items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry, when all items are covered, no element of chance is left and highest accuracy is obtained. But in practice this may not be true.

The respondents selected should be as representative of the total population as possible in order to produce a miniature cross-section. The selected respondents constitute what is technically called a ***'sample'*** and the selection process is called ***'sampling technique.***' The survey so conducted is known as ***'sample survey'***.

Implications of a Sample Design

A ***sample design*** is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample.

Sample design is determined ***before data are collected***. There are many sample designs from which a researcher can choose. Some designs are relatively more precise and easier to apply than others. Researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

3.7.2 Steps in Sample Design

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

(i) **Type of universe:** The first step in developing any sample design is to clearly define the set of objects, technically called the Universe, to be studied. The universe can be finite or infinite.

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

(iii) **Source list:** It is also known as 'sampling frame' from which sample is to be drawn. It contains the names of all items of a universe (in case of finite universe only).

(iv) **Size of sample:** This refers to the number of items to be selected from the universe to constitute a sample. This is a major problem before a researcher. The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility.

(v) **Parameters of interest:** In determining the sample design, one must consider the question of the specific population parameters which are of interest.

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

(vii) **Sampling procedure:** Finally, the researcher must decide the type of sample he will use i.e., he must decide about the technique to be used in selecting the items for the sample. In fact, this technique or procedure stands for the sample design itself.

3.7.3 Factors Influencing size of sample :

Parameters of Interest:

- The items or parameters are selected based on the researchers own interest.

Budgetary constraint :

- Cost consideration exercises a major influence.

a) **Sampling Procedure:**

- The type or technique used by the researcher to select the items.
- The technique should be selected so that for a given sample size & budget, the sampling error must be very small or negligible.

Sampling Error: may be caused (In case of Non probability sampling) due to

(1) Interviewer Bias

(2) Mistakes

(3) Non response problems

(4) Questionnaire design flaws

(5) Data processing & analysis errors

In case of probability sampling, (homogenous items) the sampling error is negligible since the sample is more accurate.

Characteristics of a Good sample :

- Should bind a truly representative sample.
- Small sampling error
- Should fit into the budgetary constraints.
- Result should be applicable in general.

Characteristics of sample techniques :

1) Much cheaper
2) Saves time
3) Much reliable
4) Suitable for carrying out different surveys
5) Scientific in Nature

Advantages of sampling:

1) Very accurate
2) Economical in Nature
3) Very reliable
4) Suitable for different surveys
5) Less time consumption
6) In case of large universe, sampling method is the only practical method for collecting the data.

3.7.4 Criteria of Selecting a Sampling Procedure

In this context one must remember that two costs are involved in a sampling analysis viz., the cost of collecting the data and the cost of an incorrect inference resulting from the data.

Researcher must keep in view the two causes of incorrect inferences viz.,

a. Systematic bias and
b. Sampling error

Systematic bias

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

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

2. Defective measuring device: If the measuring device is constantly in error, it will result in systematic bias. In survey work, systematic bias can result if the questionnaire or the interviewer is biased.

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

4. Indeterminancy principle: Sometimes we find that individuals act differently when kept under observation than what they do when kept in non-observed situations.

5. Natural bias in the reporting of data: Natural bias of respondents in the reporting of data is often the cause of a systematic bias in many inquiries. There is usually a downward bias in the income data collected by government taxation department, whereas we find an upward bias in the income data collected by some social organisation.

Sampling errors

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

Sampling error can be measured for a given sample design and size. The measurement of sampling error is usually called the '***precision of the sampling plan'***. If we increase the sample size, the precision can be improved.

But increasing the size of the sample has its own limitations viz.,

a. A large sized sample increases the cost of collecting data and
b. Also enhances the systematic bias.

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

3.8 Different Types of Sample Designs

There are different types of sample designs based on two factors viz.,

a. The representation basis and
b. The element selection technique.

On the representation basis, the sample may ***be probability sampling or it may be non-probability sampling***.

Probability sampling is based on the *concept of random selection*, whereas *non-probability sampling is 'non-random' sampling.*

On element selection basis, the sample may be either *unrestricted or restricted.*

When each sample element is drawn individually from the population at large, then the sample so drawn is known as *'unrestricted sample'*, whereas all other forms of sampling are covered under the term *'restricted sampling'*.

The following chart exhibits the sample designs as explained above.

Element selection technique	Representation basis	
	Probability sampling	Non-probability sampling
Unrestricted sampling	Simple random sampling	Haphazard sampling or convenience sampling
Restricted sampling	Complex random sampling (such as cluster sampling, systematic sampling, stratified sampling, etc.,)	Purposive sampling (such as quota sampling, judgement sampling)

Non-probability sampling: Non-probability sampling is *that sampling procedure which does not afford any basis for estimating the probability* that each item in the population has of being included in the sample. Non-probability sampling is also known by different names such as ***deliberate sampling, purposive sampling and judgement sampling***. In this type of sampling, items for the sample are selected deliberately by the researcher; his choice concerning the items remains supreme. In such a design, personal element has a great chance of entering into the selection of the sample. The investigator may select a sample which shall yield results favourable to his point of view and if that happens, the entire inquiry may get vitiated. Thus, there is always the danger of bias entering into this type of sampling technique. But in the investigators are impartial, work without bias and have the necessary experience so as to take sound judgement, the results obtained from an analysis of deliberately selected sample may be tolerably reliable.

Probability sampling: Probability sampling is also known as ***'random sampling' or 'chance sampling'***. Under this sampling design, ***every item of the universe has an equal chance*** of inclusion in the sample. It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process. Here it is blind chance alone that determines whether one item or the other is selected.

3.8.1 Probability Sampling Techniques:

Probability sampling techniques are broadly classified as simple random sampling, systematic sampling, and stratified sampling.

i)Simple Random Sampling

This is the most important and widely used probability sampling technique. They gain much significance because of their characteristic of being used to frame the concepts and arguments in statistics. Another important feature is that it allows each element in the population to have a known and equal probability of selection. This means that every element is selected independently of every other element. This method resembles lottery method where a in a system names are placed in a box, the box is shuffled, and the names of the winners are then drawn out in an unbiased manner. Simple random sampling has a definite process, though not, so rigid. It involves compilation of a sampling frame in which each element is assigned a unique identification number.

Random numbers are generated either using random number table or a computer to determine which elements to include in the sample. For example, a researcher is interested in investigating the behavioral pattern of customers while making a decision on purchasing a computer.

Accordingly, the researcher is interested in taking 5 samples from a sampling frame containing 100 elements. The required sample may be chosen using simple random sampling technique by arranging the 100 elements in an order and starting with row 1 and column 1 of random table, and going down the column until 5 numbers between 1 and 100 are selected. Numbers outside this range are ignored. Random number tables are found in every statistics book.

It consists of a randomly generated series of digits from 0 – 9. To enhance the readability of the numbers, a space between every 4^{th} digit and between every 10^{th} row is given. The researcher may begin reading from anywhere in the random number table, however, once started the researcher should continue to read across the row or down a column. The most important feature of simple random sampling is that it facilitates representation of the population by the sample ensuring that the statistical conclusions are valid.

Systematic Sampling

This is also another widely used type of sampling technique. This is used because of its ease and convenience. As in the case of simple random sampling, it is conducted choosing a random starting point and then picking every element in succession from the sampling frame. The sample interval, i, is determined by dividing the population size N by the sample size n and rounding to the nearest integer.

Consider a situation where the researcher intends to choose 10 elements from a population of 100. In order to choose these 10 elements, number the elements from one to 100. Within 20 population elements and a sample of size 10, the number is 10/100 = 1/10, meaning that one element in 10 will be selected. The sample interval will, therefore, be 10. This means that after a random start from any point in the random table, the researcher has to choose every 10^{th} element.

Systematic sampling is almost similar to simple random sampling in that each population element has a known and equal probability of selection. However, the difference lies in that simple random sampling allows only the permissible samples of size n drawn have a known and equal probability of selection. The remaining samples of size n have a zero probability of being selected.

Stratified sampling

Stratified sampling is a two-way process. It is distinguished from the simple random sampling and systematic sampling, in that:

a) It requires division of the parent population into mutually exclusively and exhaustive subsets;

b) A simple random sample of elements is chosen independently from each group or subset. Therefore, it characterises that, every population element should be assigned to one and only stratum and no population elements should be omitted. Next, elements are selected from each stratum by simple random sampling technique. Stratified sampling differs from quota sampling in that the sample elements are selected probabilistically rather than based on convenience or on judgemental basis.

Strata are created by a divider called the stratification variable. This variable divides the population into strata based on homogeneity, heterogeneity, relatedness or cost. Sometimes, more than one

variable is used for stratification purpose. This type of sampling is done in order to get homogenous elements within each strata and, the elements between each strata should have a higher degree of heterogeneity. The number of strata to be formed for the research is left to the discretion of the researcher, though, researchers agree that the optimum number of strata may be 6.

The reasons for using stratified sampling are as follows:
a) it ensures representation of all important sub-populations in the sample;
b) the cost per observation in the survey may be reduced;
c) it combines the use of simple random sampling with potential gains in precision;
d) estimates of the population parameters may be wanted for each sub-population and;
e) increased accuracy at given cost.

3.8.2 Non-probability Sampling Methods

Non-probability sampling does not involve random selection. It involves personal judgement of the researcher rather than chance to select sample elements. Sometimes this judgement is imposed by the researcher, while in other cases the selection of population elements to be includes is left to the individual field workers. The decision maker may also contribute to including a particular individual in the sampling frame. Evidently, non probability sampling does not include elements selected probabilistically and hence, leaves an degree of „sampling error" associated with the sample.

Sampling error is the degree to which a sample might differ from the population. Therefore, while inferring to the population, results could not be reported plus or minus the sampling error.

In non-probability sampling, the degree to which the sample differs from the population remains unknown However, we cannot come to a conclusion that sampling error is an inherent of non probability sample.
Non-probability samples also yield good estimates of the population characteristics. Since, inclusion of the elements in the sample are not determined in a probabilistic way, the estimates obtained are not statistically projectable to the population.
The most commonly used non-probability sampling methods are convenience sampling, judgment sampling, quota sampling, and snowball sampling.

Convenience Sampling

Convenience samples are sometimes called accidental samples because the elements included in the sample enter by „accident". It is a sampling technique where samples are obtained from convenient elements. This refers to happening of the element at the right place at the right time, that is, where and when the information for the study is being collected. The selection of the respondents is left to the discretion of the interviewer. The popular examples of convenience sampling include (a) respondents who gather in a church (b) students in a class room (c) mall intercept interviews without qualifying the respondents for the study (d) tear-out questionnaire included in magazines and (e) people on the street. In the above examples, the people may not be qualified respondents.

However, form part of the sample by virtue of assembling in the place where the researcher is conveniently placed.
Convenience sampling is the least expensive and least time consuming of all sampling techniques. The disadvantage with convenience sampling is that the researcher would have no way of knowing if the sample chosen is representative of the target population.

Judgement Sampling

This is a form of convenience sampling otherwise called as purposive sampling because the sample elements are chosen since it is expected that they can serve the research purpose. The sample elements are chosen based on the judgement that prevails in the researcher‟s mind about the prospective individual. The researcher may use his wisdom to conclude that a particular individual may be a representative of the population in which one is interested.

The distinguishing feature of judgment sampling is that the population elements are purposively selected. Again, the selection is not based on that they are representative, but rather because they can offer the contributions sought. In judgement sampling, the researcher may be well aware of the characteristics of the prospective respondents, in order that, he includes the individual in the sample. It may be possible that the researcher has ideas and insights about the respondent‟s requisite experience and knowledge to offer some perspective on the research question.

Quota Sampling

Quota sampling is another non-probability sampling. It attempts to ensure that the sample chosen by the researcher is a representative by selecting elements in such a way that the proportion of the sample elements possessing a certain characteristic is approximately the same as the proportion of the elements with the characteristic in the population.

Quota sampling is viewed as two-staged restricted judgemental sampling technique. The first stage consists of developing control categories, or quotas, of population elements. Control characteristics involve age, sex, and race identified on the basis of judgement. Then the distribution of these characteristics in the target population is determined. For example, the researcher may use control categories in that, he/she intends to study 40% of men and 60% of women in a population. Sex is the control group and the percentages fixed are the quotas.

In the second stage, sample elements are selected based on convenience or judgement. Once the quotas have been determined, there is considerable freedom to select the elements to be included in the sample. For example, the researcher may not choose more than 40% of men and 60% of women in the study. Even if the researcher comes across qualified men after reaching the 40% mark, he/she would still restrict entry of men into the sample and keep searching for women till the quota is fulfilled.

Snowball Sampling

This is another popular non-probability technique widely used, especially in academic research. In this technique, an initial group of respondents is selected, usually at random. After being interviewed, these respondents are asked to identify others who belong to the target population of interest. Subsequent respondents are selected based on the information provided by the selected group members. The group members may provide information based on their understanding about the qualification of the other prospective respondents. This method involves probability and non-probability methods. The initial respondents are chosen by a random method and the subsequent respondents are chosen by non-probability methods.

DATA PREPARATION & ANALYSIS

3.1 Data Analysis And Interpretation

A researcher's important function is the appropriate interpretation of different types of statistical data with the help of his tools. The preliminary statistical work consists of collection, classification, tabulation, presentation and analysis of data. The most important part of the statistical work consists in the proper use of the statistical tools in the interpretation of data.
The most commonly used tools are 'Mean, Median, Mode; Geometric Mean, Measures of Dispersion such as Range; Mean Deviation, Standard Deviation and also other measures such as Coefficient of Correlation, Index Numbers etc. It is necessary to note that technical interpretation of data has to be combined with a high degree of sound judgement, statistical experience, skill and accuracy.

3.2 Data Editing And Coding

Processing comprises the task of editing, coding classification and tabulation. In spite of a careful collection by a researcher, there may be a possibility for errors of omission and commission arising and it is for this purpose that the process of editing becomes necessary.
The editor, while examining certain responses of the respondents, may find some mistakes in the form of incomplete, vague or irrelevant answers. Such inconsistent answers have to be eliminated or suitably and reasonably modified.

In case a researcher is confronted with a very large volume of data then it is imperative to use 'computer processing'. For this purpose necessary statistical packages such as SPSS etc. may be used. Computer technology can prove to be a boon because a huge volume of complex data can be processed speedily with greater accuracy.

3.2.1 Interpretation Of Data In General Management And Social Sciences

Data pertaining to economic, psychological sociological or managerial phenomena necessarily requires appropriate interpretation through the use of analytical procedures based on inductive or deductive logical reasoning. Further, proper statistical methods will have to applied for scientific analysis.

Depending upon the nature of the data which may be nominal, ordinal, interval or ratio level, a researcher has to carefully evaluate the appropriateness and precision in the use of 'Parametric' or 'Non-parametric' tests of hypothesis. It may be noted that generally the nominal level data is weak whereas the ratio level data is comparatively strong.

Statistical analysis can be classified as (i) descriptive and (ii) inferential.

Depending upon the nature of researcher's problem, relevant sampling methods are used for obtaining data. However, for the purpose of hypothesis testing, parametric or non-parametric tests may be used depending upon the fact whether the assumptions in regard to population are based on 'distribution' or 'distribution-free characteristics'.

4.2.2 Interpretation Of Financial Ratios

Financial ratio analysis is a study of ratios between various items or groups of items in financial statements. Financial ratios can be broadly classified into the following categories:
1. Liquidity ratios
2. Leverage ratios
3. Turnover ratios
4. Profitability ratios

Liquidity Ratios

Liquidity refers to the ability of a firm to meet its obligations in the short run, usually one year. Liquidity ratios are generally based on the relationship between current assets and current liabilities.

The important liquidity ratios are:
(a) ***Current Ratio:*** Current assets include cash, current investments, debtors, inventories (stocks), loans and advances, and prepaid expenses. Current liabilities represent liabilities that are expected to mature in the next twelve months.
These comprise

1. loans, secured or unsecured, that are due in the next twelve months and
2. (ii) current liabilities and provisions.The current ratio thus measures the ability of the firm to meet its current liabilities.

(b) ***Acid-Test Ratio (also called the quick ratio):*** Quick assets are defined as current assets excluding inventories.
It is a fairly stringent measure of liquidity. It is based on those current assets, which are highly liquid. Inventories are excluded because they are deemed to be the least liquid component of the current assets.
(c) ***Cash Ratio:*** Because cash and bank balance and short term marketable securities are the most liquid assets of a firm.

Leverage Ratios
Financial leverage refers to the use of debt finance. While debt capital is a cheaper source of finance, it is also a riskier source of finance. Leverage ratios help in accessing the risk arising from the use of debt capital. Two types of ratios are commonly used to analyze financial leverage:
(i) Structural ratios
(ii) Coverage ratios
Structural ratios are based on the proportions of debt and equity in the financial structure of the firm. Coverage ratios show the relationship between debt serving commitments and sources for meeting these burdens.

The important structural ratios are:
(a) ***Debt-Equity Ratio:*** It shows the relative contributions of creditors and owners.
The numerator of this ratio consists of all debt, short-term as well as long-term, and the denominator consists of net worth plus preferential capital.
(b) ***Debt-Assets Ratio:*** It measures the extent to which borrowed funds support the firm's assets.
The numerator of this ratio includes all debts, short-term as well long-term, and the denominator of this ratio is total of all assets.

(c) ***Interest Coverage Ratio (also called "times interest earned"):*** A high interest coverage ratio means that the firm can easily meet the interest burden even if profit before interest and taxes suffer a considerable decline. A low interest coverage ratio may result in financial embarrassment when profit before interest and taxes decline.

Though widely used, this ratio is not a very appropriate measure because the source of interest payment is cash flow before interest and taxes.

(d) ***Fixed Charges Coverage Ratio:*** This ratio shows how many times the cash flow before interest and taxes covers all fixed financing charges. In the denominator of this ratio, only the repayment of loan is adjusted upwards for the tax factor because the loan repayment amount, unlike interest, is not tax deductible.

(e) ***Debt Service Coverage Ratio***

Turnover Ratios

Turnover ratios also referred to as activity ratios or assets management ratios, measure how efficiently the assets are employed by a firm. The important turnover ratios are:
(a) ***Inventory Turnover:*** It measures how fast the inventory is moving through the firm and generating sales. It reflects the efficiency of inventory management.
(b) ***Debtors' Turnover:*** It shows how many times accounts receivable (debtors) turnover during the year.
(c) ***Average Collection Period:*** It represents the number of days' worth of credit sales that is locked in debtors.
(d) ***Fixed Assets Turnover:*** This ratio measures sales per rupee of investment in fixed assets.
This ratio is supposed to measure the efficiency with which fixed assets are employed.
(e) ***Total Assets Turnover:*** This ratio measures how efficiently assets are employed overall.

Profitability Ratios
They reflect the final result of business operations. There are two types of profitability ratios:
(i) Profit margin ratios
(ii) Rate of return ratios

The important profit margin ratios are:
(a) ***Gross Profit Margin Ratio:*** The ratio shows the margin left after meeting manufacturing costs. It measures the efficiency of the production as well as pricing.
(b) ***Net Profit Margin Ratio:*** This ratio shows the earnings left for shareholders as a percentage of net sales.
(c) ***Return on Total Assets:*** It is measure of how efficiently the capital is employed. To ensure internal consistency, the following variant of return on total assets may be employed:
(a) ***Earning Power:*** It is a measure of operating profitability.
(b) ***Return on Equity:*** it is a measure of great interest to equity shareholder. The numerator of this ratio is equal to profit after tax less preference dividends. The denominator includes all contributions made by equity shareholders. It is also called the return on net worth.

4.3 Classification And Tabulation

Classification is the process of sorting 'similar' things from among a group of objects with different characteristics. In other words, heterogeneous data is divided into separate homogeneous classes according to characteristics that exist amongst different individuals or quantities constituting the data. Thus, fundamentally classification is dependent upon similarities and resemblances among the items in the data.

The main object of classification is to present vividly, in a simplified and quickly intelligible form, a mass of complex data. Without condensing details in a classified form it is difficult to compare quickly, interpret thoroughly and analyse properly different sets of quantitative and qualitative phenomena. The basic requirements of good classification are stability, non ambiguity, flexibility and comparability.

4.3.1 Descriptive and Quantitative Classification

Depending on the characteristics of the data, they can be broadly categorized into two separate and distinct groups - descriptive and numerical. Descriptive characteristics are those that can be described in words and are expressible in qualitative terms. Numerical characteristics are quantitative in nature. For instance, literacy, sex, caste and religion are descriptive characteristics. Height, weight, age, income and expenditure are numerically expressible characteristics. Descriptive or qualitative classification is termed classification according to attributes. Numerical or quantitative classification of data in certain class intervals is termed as classification in terms of classes with certain intervals, or classification according to class intervals.

4.3.2 Simple and Manifold Classification

Classification based on attributes may be either simple or manifold. In the case of simple classification, only one attribute is studied. That is, the data is classified into two separate classes

under a single attribute. For instance, data collected on literacy in the country can be classified into two distinct classes: literate and illiterate. Since this process is quite simple, it is known as simple classification.

On the other hand, analysing and classifying collected data under several attributes in different classes is called manifold classification. For example, if each of the two classes, literate and illiterate, is divided into males and females, then there would be four classes. If classified further on a regional basis, there would be a number of other classes. Such a process of classification of data into a number of classes and classes within classes is known as manifold classification.

4.3.3 Classification According to Class Intervals

Phenomena like income, heights and weights are all quantitatively measurable and data on them can be classified into separate class intervals of uniform length. For instance, the marks obtained by a group of 50 candidates in a subject at an examination can be classified into the following classes: 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70 etc. Each class has a lower and an upper limit and the number of candidates getting marks between these two limits of the same class interval is called the frequency of the respective class. To give an example, if 12 candidates get between 40 and 50 marks, 12 is the frequency of the class 40-50.

Number of Classes

The number of classes into which particular data should be classified depends upon the mass of data. The larger the mass, the more should be the number of classes. Usually data is classified into not less than six classes and not more than 20 classes, depending upon the mass and size of the data and the length of the class intervals. The fundamental object of classifying data is to get the maximum possible amount of information most precisely. According to Sturges' Rule, the number of class intervals (n) = 1 + 3.322 log N, where N = total number of observations.

Length of Class Intervals

The uniform length of class intervals depends upon the difference between the extreme items in the data-the largest item and the smallest item-and the number of classes required. For example, if in the data on marks secured by 250 candidates in a subject at an examination, 0 and 93 are the lowest and highest marks respectively and 10 classes are required, each class would then have a class interval length of 10. Ordinarily class intervals are fixed in such a way as to enable easy calculation and precision.

Class Limits

The choice of class limits is determined by the mid-value of a class interval, which should as far as possible be identical with the arithmetic average of the items occurring in that class interval.

4.4 Tabulation

Tabulation is the process of arranging given quantitative data based on similarities and common characteristics in certain rows and columns so as to present the data vividly for quick intelligibility, easy comparability and visual appeal.

Components of a Statistical Table

A statistical table comprises a title, a head-note, a stub head and stub details, captions and columns under the captions, field of the table under different column heads, footnotes and, source notes.

Here's a sample:

Title: Students studying in different classes in X, Y, Z Commerce College.

Head-Note: Data relates to the academic year for ex.1998-99.

Purpose of Statistical Tables

Statistical tables are of two types: general purpose table and special purpose table.

1. ***General Purpose Table:*** This is primarily meant to present the entire original data on a subject. Such presentation of numerical data in a tabular form is especially useful as a source of information and reference for constructing different special purpose tables.

2. ***Special Purpose Table:*** As its name implies, this is a statistical table that specially presents and emphasizes certain phases or significant aspects of the information given in a general purpose table. Presenting data in a special table not only makes it easy to understand specific data, it also facilitates easy comparison and clear-cut interpretation.

4.4.1 Types of Tabulation

1. ***One-way Table (single tabulation):*** A one-way table gives answers to questions about one characteristic of the data.
2. ***Two-way Table (double tabulation):*** A two-way table gives information about two interrelated characteristics of a particular type of data.
3. ***Three-way Table (Triple Tabulation):*** A three-way table answers questions relating to three interrelated characteristics of a given data.
4. ***Higher Order Table (Manifold Tabulation):*** This table gives information under several main heads and subheads on questions relating to a number of interrelated characteristics.

4.4.2 Rules and Precautions on Tabulation

1. Every statistical table should be given an appropriate title to indicate the nature of the data.
The title should be simple, intelligible and unambiguous and should not be too lengthy or too short.
2. If necessary, the title may be further explained with a suitable head-note.
3. Different types of data require different types of tabulation.

It has to be decided at the outset whether one or more tables would be necessary to fit in the data precisely and suitably. A single simple table is appealing to the eye provided it is prepared properly. Several tables or a large table make comparisons difficult.

4. The stub heads and the main heads should be consistent with the nature of the data and be very clear.
5. The main headings under the caption should be as few as possible in keeping with the requirements of space and type of data. If the main headings are few, comparison between different sets of data becomes easy.
6. The entire arrangement of data should be appropriate, compact and self-explanatory so that it is not necessary to rearrange the data in any manner.
7. Comparisons between different figures such as totals and averages-are easier if they are arranged vertically and not horizontally.
8. In order to show important parts of the data (under main heads) distinctly, it is necessary to draw thick double or multiple ruled lines.

9. Depending upon the nature of the data, items in the stub column may be arranged according to:
(i) Alphabetical order.
(ii) Geographical importance.
(iii) Customary classification.
(iv) Chronological order.
(v) Size or magnitude.

10. Figures in the data that are estimates, approximate or revised should be indicated by an alphabet, asterisk, number or any other symbol. An explanation should be given in the footnote.
11. The different units used in the data should be indicated in the column heads. For example: 'figures in rupees', 'figures in metres', and so on.
12. The source of the data should be indicated under the footnote. It is necessary to mention the source for further references and other details and also for indicating the reliability of the data.

4.5 Statistical Data Analysis

The data generated using the questionnaire is analysed and inference made out of the data could be used by the decision maker. The fundamental question that arises in the minds of the researcher is: "What technique should be used to analyse the collected data?"
The collected data may be coded as per the description given in the scaling lesson. The researcher should ensure that he/she does not deviate from the scaling principles enumerated in the scaling lesson. The researcher can create a master file containing the coded information of all the items included in the questionnaire.
Data analysis technique depends on the level of measurement and the type of sample the researcher uses.
Ordinal data may be subjected to median and interquartile range. Under inferential statistics, non parametric techniques such as Kolmogorov Smirnov test, Mann Whitney test, Kruskal Wallis, and Friedman two- way ANOVA are used. Interval and ratio scale may be subjected to mean and standard deviation. Under inferential statistics, z test, t –test, one-way ANOVA, correlation and regression.

4.6 Concepts Relating to Testing of Hypothesis

Null Hypothesis & Alternative Hypothesis (Statistical Analysis)

Null Hypothesis: Denoted by H0. If both the variables (say male or female) or (Head or Tail) are equally good, it is Null Hypothesis.

Alternative Hypothesis: Denoted by Ha or H1. If one variable is considered superior to other or vice versa or if there is a difference, it is alternative hypothesis.

Mean Population (u) or (p)

Total / No. of variables

Null Hypothesis

Ho : u = 100

Alternative Hypothesis

Ha : u = 100

Ha : u > 100

Ha : u < 100

4.6.1 Aspects to be considered while formulating Null Hypothesis

1) The researcher always tries to reject Null hypothesis since Alternative Hypothesis should be proved.

2) Null hypothesis when it is actually true, when rejected involves great risk, the level of significance should be considered.

3) Null hypothesis should be very specific (No approximation)

The level of significance:

- Important concept of hypothesis testing.
- It is a certain percentage chosen with great 'care, reason and thought'

  (eg) let us consider the level of significance to be 5%. It means the Researcher takes a risk of rejecting Null hypothesis (Ho) by 5% when Ho is actually true.

3. Decision Rule

- The researcher should make a decision, if to accept or Reject Ho.
- The decision rule should be decided on the number of items to be tested and the basic of which to accept or reject.

4. **Type I and Type II Errors**

(i) Researcher may reject Ho, when it is true – Type I Error (which must have been accepted).

(ii) Researcher may accept Ho, when it is false – Type II Error (which must have been rejected)

5. **One tailed and Two tailed Tests**:

(i) One tailed test rejects the Null hypothesis when the sample mean is either greater or lower than the hypothesized value of the population mean.

Two tailed Test: When the sample mean is both greater and lower than the hypothesized value of the population mean.

4.6.2 Procedure for Hypothesis Testing:

1. Testing hypothesis refers whether the formulated hypothesis is valid or not
2. Whether to Accept or Reject Null Hypothesis.

(i) **Making a formal statement:**

- Making a formal statement of the null hypothesis and alternative hypothesis.

(ii) Selecting a significant level of testing

- A pre-determined level of significance should be specified.
- Either 5% or 1% level can be considered for the purpose.

(iii) Deciding the Distribution to use:

- Choice should be made generally relates to Normal distribution or t-distribution.

(iv) Selection of random sample & computing an Appropriate value

- Selection of Random sample
- Computing suitable value
- Drawing a sample for furnishing Empirical data.

(v) Calculation of Probability:

- The diverged results from the expected results, when Ho is true.

(vi) Comparing the probability:

- By making a comparison with the assumed significance level.
- If the value is less than or equal to Ho, in case of one-tailed test, Ho is rejected. Here type I error is committed.
- If the value is greater than the mean, Ho is accepted. Were type-I error is committed.

- compile, compare & compute the data and come out with the inference.

Null Hypothesis: The null hypothesis is the proposition or proposal that implies no effect on the phenomena.

Alternative Hypothesis: is the one predictive statement that implies some effect on the phenomena.

4.6.3 Rules For Hypothesis Testing

Basic analysis of the data involves testing of hypothesis. Lot of confusion prevails in developing a hypothesis.

In simple terms,

- Hypothesis refers to assumption of a relationship between two variables or difference between two or more groups.
- Hypothesis also contains the direction of relationship between the variables concerned.

Examples for hypothesis is given below:
(a) The purchasing power of the consumers is positively related to the availability of surplus income.
(b) Customers belonging to the Northern states in India have a different taste preference than those from Northern States.

Hypotheses are of two types: (a) Null hypothesis and (b) Alternative hypothesis. A simple rule may be followed to develop a hypothesis:
1. What we hope or expect to be able to conclude as a result of the test usually should be placed in alternative hypothesis.
2. The null hypothesis should contain a statement of equality (=) and an alternative hypothesis contains a > or < than sign.
3. The null is the hypothesis that is tested.
4. The null and alternate hypothesis are complementary.

4.7 How To Select A Particular Test

An appropriate statistical test for analysing a given set of data is selected on the basis of:
Scaling of the data: Is the measurement scale nominal, ordinal, interval or ratio; Dependence, Independence of the measurements; Types of samples: Independent or dependent samples;
Number of samples (groups) studied and; Specific requirements such as sample size, shape of population distribution, are also used for considering the choice of a statistical test.

There are two types of samples: Independent and dependent samples. Two samples are independent sample if the sample selected from one of the populations has no effect or bearing on the sample selected from the other population. E.g., responses collected from Tamilians, Keralites, Kannadigas etc. They are exclusive groups of respondents where a Tamilian is exclusive in nature in that he does not take part in the other groups. Similarly, a Kannadiga is exclusive in nature in his membership in his group in that he does not take part in any other groups.

Dependent samples, also called related or correlated or matched samples, are ones in which the response of the nth subject in one sample is partly a function of the response of the nth subject in an earlier sample. Examples of dependent samples include before-during-after samples of the same people or matched response of similar people.

The nature of the samples is also considered while deciding on the appropriateness of the statistical test. The following are the conditions to be followed while choosing the tests:
Does the test involve one sample, two samples or k samples
If 2 samples or k samples are involved, are the individual cases independent or related.

The selection of an appropriate statistical test rests with two criteria:
(a) Type of scale used (Nominal, ordinal, interval or ratio)
(b) Type and the size of the samples. Type relates to whether the samples are independent or dependent.
The hypothesis of type two mentioned in the example above could be tested using two types of statistical tests. They are:
(a) Parametric tests
(b) Non-parametric tests

A simple understanding of the characteristics of the tests reveal that the term „parametric" is derived from the term parameter which is a descriptive measure computed from or used to describe a population of data. Parametric tests are used to test hypothesis with interval and ratio measurements and non parametric tests are used to test hypothesis involving nominal and ordinal data. Parametric tests are more powerful than non–parametric tests. Explanation of parametric and non parametric tests in detail is beyond the scope of this study material.

There are few simple, easy to understand assumptions made while applying a parametric test.
They are:

- The observations must be independent – that is, the selection of any one case should not affect the chances for any other case to be included in the sample.
- The observations should be drawn from normally distributed populations.
- These populations should have equal variances.
- The measurement scales should be at least interval so that arithmetic operations can be used with them.
- Non-parametric tests do not have any assumptions of such kind.
- This is the advantage of nonparametric tests over parametric tests.
- Hypothesis of the type 1 may be tested using Correlation and regression

Correlation is a test of association only between two variables. It uses only interval and ratio scale. Such correlations are called as Karl Pearson bi–variate correlation. Correlation of a special type employed on ordinal data is called Rank Correlation. This is otherwise called as Spearman Rank correlation.

However, correlation will never tell the researcher about the independent – dependent relationship. Correlation analysis will give a result r called the correlation coefficient. R value ranges from -1 to +1 through a O. As r value approaches 1, the strength of the association increases and as it approaches 0, it decreases. R value will be associated with a positive or negative sign.

Positive sign refers to positive correlation where the change in one variable causes change in the other variable in the same direction whereas a negative sign indicates inverse relationship.

Regression is a powerful technique dealing with two or more than two number of variables. Regression analysis will tell the researcher about the independent and dependent relationship. It deals with one dependent variable and any number of independent variables. Regression analysis involving only one independent variable, is called simple regression and that involves more than one independent variables is called multiple regression.

4.8 Validity of data.

4.8.1 Qualitative Vs Quantitative data analyses

There are two general approaches to gathering and reporting information: qualitative and quantitative approaches. The qualitative approach to research is focused on understanding a phenomenon from a closer perspective. The quantitative approach tends to approximate phenomena from a larger number of individuals using survey methods. In this research corner, I describe methods that are generally

used in each strand of research. Each approach has its benefits and detriments, and is more suitable to answering certain kinds of questions.

Qualitative Approach

The qualitative approach to gathering information focuses on describing a phenomenon in a deep comprehensive manner. This is generally done in interviews, open-ended questions, or focus groups. In most cases, a small number of participants participate in this type of research, because to carry out such a research endeavor requires many resources and much time. Interviews can vary from being highly structured and guided by open-ended questions, or be less structured and take the form of a conversational interview. Because of the investment in this type of research and the relatively few number of participants, findings from qualitative research cannot be generalized to the whole population. However, such research serves as a spring board for larger studies and deeper understanding that can inform theory, practice, and specific situations.

Example from youth mentoring research:

Ahrens, DuBois, Garrison, Spencer, Richardson, & Lozano (2011) used semi-structured interviews to outline themes of mentor characteristics and factors that youth perceive to influence mentor relationships. They spoke with participants on the phone and asked them open-ended questions. In identifying barriers and facilitators for relationship initiation and maintenance, Ahrens et al. provide important points of inquiry to be used in a larger scale survey-based research. One of the cautions in using qualitative approaches is that the findings apply only to this small group of 23 individuals. This research was crucial in providing evidence that these factors should be examined and further elaborated through quantitative methods prior to making any wide-range recommendation.

Benefits of the qualitative approach:

Using open-ended questions and interviews allows researchers and practitioners to understand how individuals are doing, what their experiences are, and recognize important antecedents and outcomes of interest that might not surface when surveyed with pre-determined questions. Although qualitative research can be thought of as anecdotal, when pooled across a number of participants it provides a conceptual understanding and evidence that certain phenomena are occurring with particular groups or individuals.

- Allows identification of new and untouched phenomena
- Can provide a deeper understanding of mechanisms
- Gives a one-on-one and anecdotal information
- Provides verbal information that may sometimes be converted to numerical form
- May reveal information that would not be identified through pre-determined survey questions

Limitations:

- Cannot generalize to the general population
- Challenges in applying statistical methods
- Difficulty in assessing relations between characteristics

Quantitative Approach

The quantitative approach to gathering information focuses on describing a phenomenon across a larger number of participants thereby providing the possibility of summarizing characteristics across groups or relationships. This approach surveys a large number of individuals and applies statistical techniques to recognize overall patterns in the relations of processes. Importantly, the use of surveys

can be done across groups. For example, the same survey can be used with a group of mentors that is receiving training (often called the intervention or experimental groups) and a group of mentors who does not receive such a training (a control group). It is then possible to compare these two groups on outcomes of interest, and determine what influence the training had. It is also relatively easy to survey people a number of times, thereby allowing the conclusion that a certain features (like matching) influence specific outcomes (well-being or achievement later in life).

Example from youth mentoring research:

Grossman and Rhodes (2002) examined duration of matched relationships in over 1,100 Big Brothers Big Sisters mentor-mentee matches. Because the information they used was survey-based and numerical, they were able to employ statistical techniques examining how duration of match was related to different outcomes of interest.

In using a variety of statistical techniques, they concluded that "youth who were in [matched mentoring] relationships that lasted a year or longer reported improvements in academic, psychosocial, and behavioral outcomes" (p. 213). If Grossman and Rhodes had not used survey-based quantitative research, they would not have had such a large sample of matches and therefore could not generalize to matches in general. In addition, with a smaller number of participants, it is challenging to apply some statistical techniques to examine emerging patterns across such a large group of mentored matches. The current rule of thumb to using complex statistical modeling is that you need a sample of at least 130 participants. However, for more complex modeling that controls for characteristics, a larger pool of participants is needed.

Benefits of the quantitative approach:

Using survey methods across a large group of individuals enables generalization. For example, if policy makers wanted to instantiate a policy about mentor training, they would likely require some evidence that this training actually works. Interviewing a few individuals, or conducting a focus group with forty matches, might be reflective of specific cases in which the mentoring training worked, however, it would not provide strong evidence that such training is beneficial overall. Stronger support for successful training would be evident if using quantitative methods.

- Enables gathering information from a relatively large number of participant
- Can conduct in a number of groups, allowing for comparison
- Allows generalizing to broader population
- Provides numerical or rating information
- Informative for instantiating policy or guidelines
- Lends to statistical techniques that allow determining relations between variables (think of better word)

Limitations:

- Difficulty in recognizing new and untouched phenomena
- Caution in interpretation without a control group

In summary, the qualitative and quantitative approaches to research allow a different perspective of situations or phenomena. These two main approaches to research are highly informative, especially if used in combination. Each approach has its benefits and detriments, and being aware of the methods used to gather information can help practitioners and policy-makers understand the extent to which research findings can be applied.

4.9 Bivariate and Multivariate statistical techniques

4.9.1 Bivariate analysis is one of the simplest forms of quantitative (statistical) analysis.[1] It involves the analysis of two variables (often denoted as *X*, *Y*), for the purpose of determining the empirical relationship between them.In order to see if the variables are related to one another, it is common to measure how those two variables simultaneously change together (see also covariance)

Bivariate analysis can be helpful in testing simple hypotheses of association and causality – checking to what extent it becomes easier to know and predict a value for the dependent variable if we know a case's value of the independent variable (see also correlation)

Bivariate analysis can be contrasted with univariate analysis in which only one variable is analysed.Furthermore, the purpose of a univariate analysis is descriptive. Subgroup comparison – the descriptive analysis of two variables – can be sometimes seen as a very simple form of bivariate analysis (or as univariate analysis extended to two variables). The major differentiating point between univariate and bivariate analysis, in addition to the latter's looking at more than one variable, is that the purpose of a bivariate analysis goes beyond simply descriptive: it is the analysis of the relationship between the two variable.Bivariate analysis is a simple (two variable) special case of multivariate analysis (where multiple relations between multiple variables are examined simultaneously.

Types of Bivariate analysis

Common forms of bivariate analysis involve creating a percentage table or a scatterplot graph and computing a simple correlation coefficient.[1] The types of analysis that are suited to particular pairs of variables vary in accordance with the level of measurement of the variables of interest (e.g. nominal/categorical, ordinal, interval/ratio). If the dependent variable—the one whose value is determined to some extent by the other, independent variable— is a categorical variable, such as the preferred brand of cereal, then probit or logit regression (or multinomial probit or multinomial logit) can be used. If both variables are ordinal, meaning they are ranked in a sequence as first, second, etc., then a rank correlation coefficient can be computed. If just the dependent variable is ordinal, ordered probit or ordered logit can be used. If the dependent variable is continuous—either interval level or ratio level, such as a temperature scale or an income scale—then simple regression can be used.

If both variables are time series, a particular type of causality known as Granger causality can be tested for, and vector autoregression can be performed to examine the intertemporal linkages between the variables.

4.9.2 Multivariate analysis (**MVA**) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest.

Uses for multivariate analysis include:

- Design for capability (also known as capability-based design)
- Inverse design, where any variable can be treated as an independent variable
- Analysis of Alternatives (AoA), the selection of concepts to fulfill a customer need
- Analysis of concepts with respect to changing scenarios
- Identification of critical design drivers and correlations across hierarchical levels.

Multivariate analysis can be complicated by the desire to include physics-based analysis to calculate the effects of variables for a hierarchical "system-of-systems." Often, studies that wish to use multivariate analysis are stalled by the dimensionality of the problem. These concerns are often eased through the use of surrogate models, highly accurate approximations of the physics-based code. Since surrogate models take the form of an equation, they can be evaluated very quickly. This becomes an enabler for large-scale MVA studies: while a Monte Carlo simulation across the design space is

difficult with physics-based codes, it becomes trivial when evaluating surrogate models, which often take the form of response surface equations.

4.9.3 Factor analysis

Overview: Factor analysis is used to uncover the latent structure (dimensions) of a set of variables. It reduces attribute space from a larger number of variables to a smaller number of factors. Factor analysis originated a century ago with Charles Spearman's attempts to show that a wide variety of mental tests could be explained by a single underlying intelligence factor.

Applications:

• To reduce a large number of variables to a smaller number of factors for data modeling

• To validate a scale or index by demonstrating that its constituent items load on the same factor, and to drop proposed scale items which cross-load on more than one factor.

• To select a subset of variables from a larger set, based on which original variables have the highest correlations with some other factors.

• To create a set of factors to be treated as uncorrelated variables as one approach to handling multi-collinearity in such procedures as multiple regression

Factor analysis is part of the general linear model (GLM) family of procedures and makes many of the same assumptions as multiple regression, but it uses multiple outcomes.

4.9.4 Discriminant analysis

It is described by the Number of categories possessed by the criterion variable.

If 2 category – Two group discriminant analysis.

More than 2 means – Multi Discriminant analysis.

Steps:

- ☺ Formulate the problem
- ☺ Estimate the discriminant function co efficient
- ☺ Determine the significance of discriminant function
- ☺ Interpret the results
- ☺ Assess validity of Discriminant Analysis.

4.9.5 Cluster analysis

Clustering analysis is a newly developed computer-oriented data analysis technique. It is a product of many research fields: statistics, computer science, operations research, and pattern recognition. because of the diverse backgrounds of researchers, clustering analysis has many different names. In biology, clustering analysis is called “taxonomy”.

In pattern recognition it is called “unsupervised learning.” Perhaps the most confusing name of all, the term “classification” sometimes also denotes clustering analysis. Since classification may denote discriminant analysis, which is totally different from clustering analysis, it is perhaps important to distinguish these two terms.

4.9.6 Multiple regression and correlation

4.9.7 Multidimensional scaling

Multidimensional scaling (**MDS**) is a means of visualizing the level of similarity of individual cases of a dataset. It refers to a set of related ordination techniques used in information visualization, in particular to display the information contained in a distance matrix.

It represents the Preferences and perceptions of respondents by means of a visual display.

An MDS algorithm aims to place each object in *N*-dimensional space such that the between-object distances are preserved as well as possible. Each object is then assigned coordinates in each of the *N* dimensions.

The number of dimensions of an MDS plot *N* can exceed 2 and is specified a priori. Choosing *N*=2 optimizes the object locations for a two-dimensional scatter plot.

Objectives:

- Develop techniques of full text analysis
- Select an MDA Procedure (Matric & Non metric).
- Decide on the number of dimensions.
- Lable the dimension and interpret the configuration
- Assess Reliability and validity

Uses of MDS:

- Image Measurement
- Market Segmentation
- New product development
- Assessing the Advertisement Effectiveness
- Pricing Analysis
- Attitude scale construction
- Vendor Evaluation & advertising Media selection.

Applications

Applications include scientific visualisation and data mining in fields such as cognitive science, information science, psychophysics, psychometrics, marketing and ecology. New applications arise in the scope of autonomous wireless nodes that populate a space or an area. MDS may apply as a real time enhanced approach to monitoring and managing such populations.

Furthermore, MDS has been used extensively in geostatistics, for modeling the spatial variability of the patterns of an image (by representing them as points in a lower-dimensional space),and natural language processing.

Marketing

In marketing, MDS is a statistical technique for taking the preferences and perceptions of respondents and representing them on a visual grid, called perceptual maps. By mapping multiple attributes and multiple brands at the same time, a greater understanding of the marketplace and of consumers' perceptions can be achieved, as compared with a basic two attribute perceptual map.

Comparison and advantages

Potential customers are asked to compare pairs of products and make judgments about their similarity. Whereas other techniques (such as factor analysis, discriminant analysis, and conjoint analysis) obtain underlying dimensions from responses to product attributes identified by the researcher, MDS obtains the underlying dimensions from respondents' judgments about the similarity of products. This is an important advantage. It does not depend on researchers' judgments. It does not require a list of attributes to be shown to the respondents. The underlying dimensions come from respondents' judgments about pairs of products. Because of these advantages, MDS is the most common technique used in perceptual mapping.

4.10 Application Of Statistical Software For Data Analysis.

i) SPSS – Statistical package for Social science

SPSS Statistics is a software package used for statistical analysis. Long produced by SPSS Inc., it was acquired by IBM in 2009. The current versions (2014) are officially named **IBM SPSS Statistics**. Companion products in the same family are used for survey authoring and deployment (IBM SPSS Data Collection), data mining (IBM SPSS Modeler), text analytics, and collaboration and deployment (batch and automated scoring services).

The software name stands for **Statistical Package for the Social Sciences** (**SPSS**), reflecting the original market, although the software is now popular in other fields as well, including the health sciences and marketing.

ii) SAS – Statistical Analysis System

- SAS (**Statistical Analysis System**; not to be confused with **SAP) is a software** suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics. It is the largest market-share holder for advanced analytics.

iii) STATA

- **Stata** is a general-purpose statistical software package created in 1985 by StataCorp. Most of its users work in research, especially in the fields of economics, sociology, political science, biomedicine and epidemiology.

CHAPTER V

REPORT DESIGN, WRITING AND ETHICS IN BUSINESS RESEARCH

5.1 INTRODUCTION

Interpretation and Report Writing

After collecting and analyzing the data, the researcher has to accomplish the task of drawing inferences followed by report writing. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. All this analytical information and consequential inference(s) may well be communicated, preferably through research report, to the consumers of research results who may be either an individual or a group of individuals or some public/private organisation.

5.1.1 Meaning of Interpretation

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study.

The task of interpretation has two major aspects viz.,

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

(ii) The establishment of some explanatory concepts.

"In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses."

5.1.2 Importance of Interpretation:

Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic component of research process because of the following reasons:

- It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings.
- Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.
- Researcher can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.
- The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such a study have to be interpreted on a *post-factum* basis in which case the interpretation is technically described as '*post factum*' interpretation.

5.1.3 Technique of Interpretation

- Interpretation is an art that one learns through practice and experience. The researcher may, at times, seek the guidance from experts for accomplishing the task of interpretation. The technique of interpretation often involves the following steps:
- Researcher must give reasonable explanations of the relations which he has found and he must interpret the lines of relationship in terms of the underlying processes and must try to find out the thread of uniformity that lies under the surface layer of his diversified research findings. In fact, this is the technique of how generalization should be done and concepts be formulated.

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

5.1.4 Precautions in Interpretation

It is, therefore, absolutely essential that the task of interpretation be accomplished with patience in an impartial manner and also in correct perspective.

- Researcher must pay attention to the following points for correct interpretation:
- At the outset, researcher must invariably satisfy himself that
    - the data are appropriate, Trust worthy and adequate for drawing inferences;
    - the data reflect good homogeneity; and that
    - Proper analysis has been done through statistical methods.
- The researcher must remain cautious about the errors that can possibly arise in the process of interpreting results. Errors can arise due to false generalization and/or due to wrong interpretation of statistical measures, such as the application of findings beyond the range of observations, identification of correlation with causation and the like. Another major pitfall is the tendency to affirm that definite relationships exist on the basis of confirmation of particular hypotheses. In fact, the positive test results accepting the hypothesis must be interpreted as "being in accord" with the hypothesis, rather than as "confirming the validity of the hypothesis". The researcher must remain vigilant about all such things so that false generalization may not take place. He should be well equipped with and must know the correct use of statistical measures for drawing inferences concerning his study.
- He must always keep in view that the task of interpretation is very much intertwined with analysis and cannot be distinctly separated. As such he must take the task of interpretation as a special aspect of analysis and accordingly must take all those precautions that one usually observes while going through the process of analysis viz., precautions concerning the reliability of data, computational checks, validation and comparison of results.
- He must never lose sight of the fact that his task is not only to make sensitive observations of relevant occurrences, but also to identify and disengage the factors that are initially hidden to the eye. This will enable him to do his job of interpretation on proper lines. Broad generalization should be avoided as most research is not amenable to it because the coverage may be restricted to a particular time, a particular area and particular conditions. Such restrictions, if any, must invariably be specified and the results must be framed within their limits.
- The researcher must remember that "ideally in the course of a research study, there should be constant interaction between initial hypothesis, empirical observation and theoretical conceptions. It is exactly in this area of interaction between theoretical orientation and empirical observation that opportunities for originality and creativity lie."

5.2 Content of Individual Sections of the Research Report

The research report is presented in a particular format. The format varies from place to place and from discipline to discipline. It is advisable to find out what format is required in a given institution, place or discipline before making presentations.

Essentially, research report comprises three major sections as follows:
The Preliminaries
The Main Body (chapters 1 to 5)
The Appendix

The preliminary pages include:
Title page ---------- (i)
Certification ---------- (ii)
Dedication ---------- (iii)
Acknowledgement ---------- (iv)
Table of Content ---------- (v)
List of Tables ---------- (vi)
List of Figures ---------- (vii)
Abstract ---------- (viii)

The main body of the report comprises of five chapters arranged in the following format:

Chapter 1: Introduction
1.1 Background of The Study – In this subsection, the researcher traces the conditions and factors that made the study necessary. The essence is to provide the necessary background information that will show the reader the conditions, circumstances and factors that give rise to the problem under investigation.

1.2 Statement of Problem – Having traced the development of the problem in section 1 above, the problem is now formulated and defined in concise and precise terms.

1.3 Objectives of The Study – This refers to what the research will accomplish. It represents the statement of the purpose of the study. The statement of objectives are broken down to itemized specific statements.

1.4 Relevant Research Questions – As mentioned earlier, research questions represent the major questions for which the researcher seeks to answers in the course of his or her investigation.

1.5 Statement of Hypotheses – Hypotheses refers to intelligent guesses which the researcher formulates to guide his or her search for the solution to the problem.

1.6 Scope of The Study – This deals with the extent of the research problem the researcher will be concerned with in the present study, given his or her competence, time available, interest and resources.

1.7 Significance of The Study – this refers to the Rationale or Importance of the study. It includes the major contributions of the study both to theory and general knowledge.

1.8 Definition of Terms – Here ter ms or concepts peculiar to the study are defined by the researcher.

Chapter 2: Review of Related Literature.
This chapter presents the review of a literature relevant to the research topic usually organized under relevant sub-headings such as:

2.1 Historical Background – where the history of the problem under investigation is presented.

2.2 Theories Relevant To The Research Questions And Hypotheses – theories that gave rise to the research questions and hypotheses are discussed in this section.

2.3 Current Literature Based On Each of The Relevant Variables Postulated By The Theories.

Chapter 3: Research Methodology. This chapter is made up of the following sections:

3.1 Restatement of The Relevant Research Questions And Hypotheses – where the research questions, as well as the hypotheses are restated in ter ms of the method the question will be answered and the method the hypotheses will be tested.

3.2 Research Design And Sources of Data – this relates to the general approach in carrying out the study. Here the researcher specifies the type of research design for the study and presents the sources of data, both primary and secondary.

3.3 Characteristics of The Population of Study – the characteristics of the study population is represented here in relation to status, location and size. 108

3.4 Sampling Design, Procedures And Determination of The Sample Size – the type of sampling technique to be used, and the way in which the sample size for the study is determined will be presented in this section.

3.5 Questionnaire Design, Distribution And Collection of Responses – this section specifies the type of questionnaire design to be used, and the way or method in which the questionnaire will be distributed and responses collected.

3.6 Procedures For Data Processing – this section presents the techniques for data processing and analysis.

3.7 Limitations of The Research Method – the limitation of the applicable research method will be presented and specified in this section.

Chapter 4: Data Presentation and Analysis. This chapter is made up of three relevant sections:

4.1 Presentation And Analysis of Data According To Responses to The Research *Questions – in this section, the researcher isolates the responses of the questions in* the questionnaire which answers the stated research questions and analyzes them accordingly.

4.2 Presentation And Analysis of Data Based On Test of Hypotheses – in this section the researcher presents the data to be used in testing the stated hypotheses, followed by the tests.

4.3 Analysis of Other Relevant Data – data not directly related to either the research questions or hypotheses, but relevant in the study will be presented and analyzed in this section.

Chapter 5: Summary of Findings, Recommendations And Conclusions.
The relevant sections in this chapter include:
5.1 Summary of Findings
5.2 Recommendations Based On Research Findings
5.3 Conclusion.

5.3 Writing the Report

Students often give inadequate attention to reporting their findings and conclusions. This is unfortunate. A well-presented study will often impress the reader more than a study with greater scientific quality but a weaker presentation. Report-writing skills are especially valuable to the junior executive or management trainee who aspires to rise in an organization. A well-written study frequently enhances career prospects.

Prewriting Concerns

Before writing, one should ask again, "What is the purpose of this report?" Responding to this question is one way to crystallize the problem. The second prewriting question is, "Who will read the report?" Thought should be given to the needs, temperament, and biases of the audience. You should not distort facts to meet these needs and biases but should consider them while developing the presentation.

Knowing who will read the report may suggest its appropriate length. Generally, the higher the report goes in an organization, the shorter it should be. Another consideration is technical background—the gap in subject knowledge between the reader and the writer. The greater the gap, the more difficult it is to convey the full findings meaningfully and concisely.

The third prewriting question is, "What are the circumstances and limitations under which you are writing?" Is the nature of the subject highly technical? Do you need statistics? Charts? What is the importance of the topic? A crucial subject justifies more effort than a minor one. What should be the scope of the report? How much time is available? Deadlines often impose limitations on the report.

Finally, "How will the report be used?" Try to visualize the reader using the report. How can the information be made more convenient? How much effort must be given to getting the attention and interest of the reader? Will the report be read by more than one person? If so, how many copies should be made? What will be the distribution of the report?

The Outline Once the researcher has made the first analysis of the data, drawn tentative conclusions, and completed statistical significance tests, it is time to develop an outline. A useful system employs the following organization structure:

I. Major Topic Heading
  A. Major subtopic heading
    1. Subtopic
      a. Minor subtopic
        (1) Further detail
        (a) Even further detail

Software for developing outlines and visually connecting ideas simplifies this on ceonerous task. Two styles of outlining are widely used—the **topic outline** and the **sentence outline.** In the topic outline, a key word or two is used. The assumption is that the writer knows its significance and will later remember the nature of the argument represented by that word or phrase or, alternatively, the outliner knows that a point should be made but is not yet sure how to make it.

The sentence outline expresses the essential thoughts associated with the specific topic. This approach leaves less development work for later writing, other than elaboration and explanation to improve readability. It has the obvious advantages of pushing the writer to make decisions on what to include and how to say it. It is probably the best outlining style for the inexperienced researcher because it divides the writing job into its two major components—what to say and how to say it.

Writing the Draft

Once the outline is complete, decisions can be made on the placement of graphics, tables, and charts. Each should be matched to a particular section in the outline. It is helpful to make these decisions before your first draft. While graphics might be added later or tables changed into charts, it

is helpful to make a first approximation of the graphics before beginning to write. Choices for reporting statistics will be reviewed later in this chapter.

Each writer uses different mechanisms for getting thoughts into written form. Some will write in longhand, relying on someone else to transcribe their prose into word-processed format. Others are happiest in front of a word processor, able to add, delete, and move sections at will. Whichever works for you is the best approach to use. Computer software packages check for spelling errors and provide a thesaurus for looking up alternative ways of expressing a thought. A CD-ROM can call up the 20-volume *Oxford English Dictionary,* believed to be the greatest dictionary in any language. Common word confusion (*there* for *their, to* for *too,* or *effect* for *affect*) will not be found by standard spelling checkers. Advanced programs will scrutinize your report for grammar, punctuation, capitalization, doubled words, transposed letters, homonyms, style problems, and readability level. The style checker will reveal misused words and indicate awkward phrasing.

Readability Sensitive writers consider the reading ability of their audience to achieve high readership. You can obtain high readership more easily if the topic interests the readers and is in their field of expertise. In addition, you can show the usefulness of the report by pointing out how it will help the readers. Finally, you can write at a level that is appropriate to the readers' reading abilities. To test writing for difficulty level, there are standard **readability indexes.** The Flesch Reading Ease Score gives a score between 0 and 100. The lower the score, the harder the material is to read. The Flesch-Kincaid Grade Level and Gunning's Fog Index both provide a score that corresponds with the grade level needed to easily read and understand the document.

Although it is possible to calculate these indexes by hand, some software packages will do it automatically. The most sophisticated packages allow you to specify the preferred reading level. Words that are above that level are highlighted to allow you to choose an alternative. Advocates of readability measurement do not claim that all written material should be at the simplest level possible. They argue only that the level should be appropriate for the audience. They point out that comic books score about 6 on the Gunning scale (that is, a person with a sixth-grade education should be able to read that material). *Time* usually scores about 10, while *The Atlantic* is reported to have a score of 11 or 12. Material that scores much above 12 becomes difficult for the public to read comfortably. Such measures obviously give only a rough idea of the true readability of a report. Good writing calls for a variety of other skills to enhance reading comprehension.

Comprehensibility Good writing varies with the writing objective. Research writing is designed to convey information of a precise nature. Avoid ambiguity, multiple meanings, and allusions. Take care to choose the right words—words that convey thoughts accurately, clearly, and efficiently. When concepts and constructs are used, they must be defined, either operationally or descriptively. Words and sentences should be carefully organized and edited. Misplaced modifiers run rampant in carelessly written reports. Subordinate ideas mixed with major ideas make the report confusing to readers, forcing them to sort out what is important and what is secondary when this should have been done for them. Finally, there is the matter of pace.

Pace is defined as,

The rate at which the printed page presents information to the reader . . . The proper pace in technical writing is one that enables the reader to keep his mind working just a fraction of a second behind his eye as he reads along. It logically would be slow when the information is complex or difficult to understand; fast when the information is straightforward and familiar. If the reader's mind lags behind his eye, the pace is too rapid; if his mind wanders ahead of his eye (or wants to) the pace is too slow.

If the text is overcrowded with concepts, there is too much information per sentence. By contrast, sparse writing has too few significant ideas per sentence. Writers use a variety of methods to adjust the pace of their writing:

• Use ample white space and wide margins to create a positive psychological effect on the reader.
• Break large units of text into smaller units with headings to show organization of the topics.
• Relieve difficult text with visual aids when possible.
• Emphasize important material and deemphasize secondary material through sentence construction and judicious use of italicizing, underlining, capitalization, and parentheses.

• Choose words carefully, opting for the known and short rather than the unknown and long. Graduate students, in particular, seem to revel in using jargon, pompous constructions, and long or arcane words. Naturally, there are times when technical terms are appropriate. Scientists communicate efficiently with jargon, but the audiences for most applied research are not scientifically trained and need more help than many writers supply.

• Repeat and summarize critical and difficult ideas so readers have time to absorb them.
• Make strategic use of service words. These are words that "do not represent objects or ideas, but show relationship. Transitional words, such as the conjunctions, are service words. So are phrases such as 'on the other hand,' 'in summary,' and 'in contrast.'

Tone Review the writing to ensure the tone is appropriate. The reader can, and should, be referred to, but researchers should avoid referring to themselves. One author notes that the "application of the 'you' attitude . . . makes the message sound like it is written to the reader, not sent by the author. A message prepared for the reader conveys sincerity, personalization, warmth, and involvement on the part of the author." To accomplish this, remove negative phrasing and rewrite the thought positively. Do not change your recommendations or your findings to make them positive. Instead, review the phrasing. Which of the following sounds better?

End users do not want the Information Systems Department telling them what software to buy. End users want more autonomy over their computer software choices.
The messages convey the same information, but the positive tone of the second message does not put readers from the Information Systems Department on the defensive.

Final Proof It is helpful to put the draft away for a day before doing the final editing. Go to the beach, ride a bicycle in the park, or see a movie—do anything that is unrelated to the research project. Then return to the report and read it with a critical eye. Does the writing flow smoothly? Are there transitions where they are needed? Is the organization apparent to the reader? Do the findings and conclusions adequately meet the problem statement and the research objectives? Are the tables and graphics is playing the proper information in an easy-to-read format? After assuring yourself that the draft is complete, write the executive summary.

5.3.1 Significance of Report Writing

- Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written.
- As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others.
- The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge.
- All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process.
- But the general opinion is in favour of treating the presentation of research results or the writing of report as part and parcel of the research project.

- Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research.
- This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

5.3.2 Different Steps in Writing Report

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

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

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

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

4. Rewriting and polishing of the rough draft: This step happens to be most difficult part of all formal writing. Usually this step requires more time than the writing of the rough draft. The careful revision makes the difference between a mediocre and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation.

The researcher should also "see whether or not the material, as it is presented, has unity and cohesion; does the report stand upright and firm and exhibit a definite pattern, like a marble arch? Or does it resemble an old wall of moldering cement and loose brick."4 In addition the researcher should give due attention to the fact that in his rough draft he has been consistent or not. He should check the mechanics of writing—grammar, spelling and usage.

5. Preparation of the final bibliography: Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended to the research report, is a list of books in some way pertinent to the research which has been done. It should contain all those works which the researcher has consulted.

The bibliography should be arranged alphabetically and may be divided into two parts; the first part may contain the names of books and pamphlets, and the second part may contain the names of magazine and newspaper articles. Generally, this pattern of bibliography is considered convenient and satisfactory from the point of view of reader, though it is not the only way of presenting bibliography. The entries in bibliography should be made adopting the following order:

For books and pamphlets the order may be as under:
1. Name of author, last name first.
2. Title, underlined to indicate italics.
3. Place, publisher, and date of publication.

4. Number of volumes.

Example
Kothari, C.R., *Quantitative Techniques,* New Delhi, Vikas Publishing House Pvt. Ltd., 1978.
For magazines and newspapers the order may be as under:
1. Name of the author, last name first.
2. Title of article, in quotation marks.
3. Name of periodical, underlined to indicate italics.
4. The volume or volume and number.
5. The date of the issue.
6. The pagination.

Example
Robert V. Roosa, "Coping with Short-term International Money Flows", *The Banker,* London, September, 1971, p. 995.
The above examples are just the samples for bibliography entries and may be used, but one should also remember that they are not the only acceptable forms. The only thing important is that, whatever method one selects, it must remain consistent.

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

Layout of the Research Report
The layout of the report means as to what the research report should contain. A comprehensive layout of the research report should comprise

- Preliminary pages;
- The main text; and
- The end matter.

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

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

- Introduction;
- Statement of findings and recommendations;
- The results;
- The implications drawn from the results; and
- The summary

(i) Introduction: The purpose of introduction is to introduce the research project to the readers. It should contain a clear statement of the objectives of research i.e., enough background should be given

to make clear to the reader why the problem was considered worth investigating. A brief summary of other relevant research may also be stated so that the present study can be seen in that context.

The hypotheses of study, if any, and the definitions of the major concepts employed in the study should be explicitly stated in the introduction of the report.

The methodology adopted in conducting the study must be fully explained.

The statistical analysis adopted must also be clearly stated. In addition to all this, the scope of the study should be stated and the boundary lines be demarcated. The various limitations, under which the research project was completed, must also be narrated.

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

(iii) ***Results:*** A detailed presentation of the findings of the study, with supporting data in the form of tables and charts together with a validation of results, is the next step in writing the main text of the report. This generally comprises the main body of the report, extending over several chapters. The result section of the report should contain statistical summaries and reductions of the data rather than the raw data.

"Nevertheless, it is still necessary that he states clearly the problem with which he was concerned, the procedure by which he worked on the problem, the conclusions at which he arrived, and the bases for his conclusions."

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

- A statement of the inferences drawn from the present study which may be expected to apply in similar circumstances
- The conditions of the present study which may limit the extent of legitimate generalizations of the inferences drawn from the study
- The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them

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

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

(C) End Matter

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

5.3.3 Types of Reports

The results of a research investigation can be presented in a number of ways viz., a technical report, a popular report, an article, a monograph or at times even in the form of oral presentation. Which

method(s) of presentation to be used in a particular study depends on the circumstances under which the study arose and the nature of the results.
A *technical report* is used whenever a full written report of the study is required whether for recordkeeping or for public dissemination.
A *popular report* is used if the research results have policy implications.

(A) Technical Report
In the technical report the main emphasis is on

- The methods employed,
- Assumptions made in the course of the study,
- The detailed presentation of the findings including their limitations and supporting data.

A general outline of a technical report can be as follows:
1. Summary of results: A brief review of the main findings just in two or three pages.
2. ***Nature of the study****:* Description of the general objectives of study, formulation of the problem in operational terms, the working hypothesis, the type of analysis and data required, etc.
3. ***Methods employed:*** Specific methods used in the study and their limitations. For instance, in sampling studies we should give details of sample design viz., sample size, sample selection, etc.
4. ***Data:*** Discussion of data collected, their sources, characteristics and limitations. If secondary data are used, their suitability to the problem at hand be fully assessed. In case of a survey, the manner in which data were collected should be fully described.
5. ***Analysis of data and presentation of findings:*** The analysis of data and presentation of the findings of the study with supporting data in the form of tables and charts be fully narrated. This, in fact, happens to be the main body of the report usually extending over several chapters.
6. ***Conclusions****:* A detailed summary of the findings and the policy implications drawn from the results be explained.
7. ***Bibliography****:* Bibliography of various sources consulted be prepared and attached.
8. ***Technical appendices****:* Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones.
9. ***Index****:* Index must be prepared and be given invariably in the report at the end.
The order presented above only gives a general idea of the nature of a technical report; the order of presentation may not necessarily be the same in all the technical reports. This, in other words, means that the presentation may vary in different reports; even the different sections outlined above will not always be the same, nor will all these sections appear in any particular report.
It should, however, be remembered that even in a technical report, simple presentation and ready availability of the findings remain an important consideration and as such the liberal use of charts and diagrams is considered desirable.

(B) Popular Report
The popular report is one which gives emphasis on simplicity and attractiveness. The simplification should be sought through clear writing, minimization of technical, particularly mathematical, details and liberal use of charts and diagrams. Attractive layout along with large print, many subheadings, even an occasional cartoon now and then is another characteristic feature of the popular report.
Besides, in such a report emphasis is given on practical aspects and policy implications.
1. ***The findings and their implications****:* Emphasis in the report is given on the findings of most practical interest and on the implications of these findings.
2. ***Recommendations for action:*** Recommendations for action on the basis of the findings of the study is made in this section of the report.
3. ***Objective of the study****:* A general review of how the problem arises is presented along with the specific objectives of the project under study.
4. ***Methods employed****:* A brief and non-technical description of the methods and techniques used, including a short review of the data on which the study is based, is given in this part of the report.
5. ***Results****:* This section constitutes the main body of the report wherein the results of the study are presented in clear and non-technical terms with liberal use of all sorts of illustrations such as charts, diagrams and the like ones.

6. *Technical appendices*: More detailed information on methods used, forms, etc. is presented in the form of appendices. But the appendices are often not detailed if the report is entirely meant for general public.

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

Oral Presentation

- At times oral presentation of the results of the study is considered effective, particularly in cases where policy recommendations are indicated by project results. The merit of this approach lies in the fact that it provides an opportunity for give-and-take decisions which generally lead to a better understanding of the findings and their implications.
- But the main demerit of this sort of presentation is the lack of any permanent record concerning the research details and it may be just possible that the findings may fade away from people's memory even before an action is taken.
- In order to overcome this difficulty, a written report may be circulated before the oral presentation and referred to frequently during the discussion. Oral presentation is effective when supplemented by various visual devices.
- Use of slides, wall charts and blackboards is quite helpful in contributing to clarity and in reducing the boredom, if any. Distributing a board outline, with a few important tables and charts concerning the research results, makes the listeners attentive who have a ready outline on which to focus their thinking. This very often happens in academic institutions where the researcher discusses his research findings and policy implications with others either in a seminar or in a group discussion.
- Thus, research results can be reported in more than one ways, but the usual practice adopted, in academic institutions particularly, is that of writing the Technical Report and then preparing several research papers to be discussed at various forums in one form or the other. But in practical field and with problems having policy implications, the technique followed is that of writing a popular report.
- Researches done on governmental account or on behalf of some major public or private organizations are usually presented in the form of technical reports.

5.3.4 Mechanics of Writing A Research Report

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

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

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

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

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

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

(a) The footnotes serve two purposes viz., the identification of materials used in quotations in the report and the notice of materials not immediately necessary to the body of the research text but still of supplemental value. In other words, footnotes are meant for cross references, citation of authorities and sources, acknowledgement and elucidation or explanation of a point of view. It should always be kept in view that footnote is not an end nor a means of the display of scholarship. The modern tendency is to make the minimum use of footnotes for scholarship does not need to be displayed.

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

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

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

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

(i) ***Regarding the single-volume reference***

1. Author's name in normal order (and not beginning with the last name as in a bibliography) followed by a comma;
2. Title of work, underlined to indicate italics;
3. Place and date of publication;
4. Pagination references (The page number).

Example

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

(ii) ***Regarding multivolumed reference***

1. Author's name in the normal order;
2. Title of work, underlined to indicate italics;
3. Place and date of publication;
4. Number of volume;
5. Pagination references (The page number).

(iii) ***Regarding works arranged alphabetically***

For works arranged alphabetically such as encyclopedias and dictionaries, no pagination reference is usually needed. In such cases the order is illustrated as under:

Example 1

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

Example 2

"Mary Wollstonecraft Godwin," *Dictionary of national biography.*

But if there should be a detailed reference to a long encyclopedia article, volume and pagination reference may be found necessary.

(iv) ***Regarding periodicals reference***
1. Name of the author in normal order;
2. Title of article, in quotation marks;
3. Name of periodical, underlined to indicate italics;
4. Volume number;
5. Date of issuance;
6. Pagination.

(v) ***Regarding anthologies and collections reference***
Quotations from anthologies or collections of literary works must be acknowledged not only by author, but also by the name of the collector.

(vi) ***Regarding second-hand quotations reference***
In such cases the documentation should be handled as follows:
1. Original author and title;
2. "Quoted or cited in,";
3. Second author and work.

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

(vii) ***Case of multiple authorship***

If there are more than two authors or editors, then in the documentation the name of only the first is given and the multiple authorship is indicated by "et al." or "and others". Subsequent references to the same work need not be so detailed as stated above. If the work is cited again without any other work intervening, it may be indicated as *ibid,* followed by a comma and the page number. A single page should be referred to as p.,

but more than one page be referred to as pp. If there are several pages referred to at a stretch, the practice is to use often the page number, for example, pp. 190ff, which means page number 190 and the following pages; but only for page 190 and the following page '190f'. Roman numerical is generally used to indicate the number of the volume of a book. Op. cit. (opera citato, in the work cited) or Loc. cit. (loco citato, in the place cited) are two of the very convenient abbreviations used in the footnotes. Op. cit. or Loc. cit. after the writer's name would suggest that the reference is to work by the writer which has been cited in detail in an earlier footnote but intervened by some other references.

7. ***Punctuation and abbreviations in footnotes:*** The first item after the number in the footnote is the author's name, given in the normal signature order. This is followed by a comma. After the comma, the title of the book is given: the article (such as "A", "An", "The" etc.) is omitted and only the first word and proper nouns and adjectives are capitalized. The title is followed by a comma.
Information concerning the edition is given next. This entry is followed by a comma. The place of publication is then stated; it may be mentioned in an abbreviated form, if the place happens to be a famous one such as Lond. for London, N.Y. for New York, N.D. for New Delhi and so on. This entry is followed by a comma. Then the name of the publisher is mentioned and this entry is closed by a comma. It is followed by the date of publication if the date is given on the title page. If the date appears in the copyright notice on the reverse side of the title page or elsewhere in the volume, the comma should be omitted and the date enclosed in square brackets [c 1978], [1978]. The entry is followed by a comma.

Then follow the volume and page references and are separated by a comma if both are given. A period closes the complete documentary reference. But one should remember that the documentation regarding acknowledgements from magazine articles and periodical literature follow a different form as stated earlier while explaining the entries in the bibliography.

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

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

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

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

Precautions for Writing Research Reports

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

1. While determining the length of the report (since research reports vary greatly in length), one should keep in view the fact that it should be long enough to cover the subject but short enough to maintain interest. In fact, report-writing should not be a means to learning more and more about less and less.
2. A research report should not, if this can be avoided, be dull; it should be such as to sustain reader's interest.
3. Abstract terminology and technical jargon should be avoided in a research report. The report should be able to convey the matter as simply as possible. This, in other words, means that report should be written in an objective style in simple language, avoiding expressions such as "it seems," "there may be" and the like.
4. Readers are often interested in acquiring a quick knowledge of the main findings and as such the report must provide a ready availability of the findings. For this purpose, charts, graphs and the statistical tables may be used for the various results in the main report in addition to the summary of important findings.
5. The layout of the report should be well thought out and must be appropriate and in accordance with the objective of the research problem.
6. The reports should be free from grammatical mistakes and must be prepared strictly in accordance with the techniques of composition of report-writing such as the use of quotations, footnotes, documentation, proper punctuation and use of abbreviations in footnotes and the like.
7. The report must present the logical analysis of the subject matter. It must reflect a structure wherein the different pieces of analysis relating to the research problem fit well.

8. A research report should show originality and should necessarily be an attempt to solve some intellectual problem. It must contribute to the solution of a problem and must add to the store of knowledge.
9. Towards the end, the report must also state the policy implications relating to the problem under consideration. It is usually considered desirable if the report makes a forecast of the probable future of the subject concerned and indicates the kinds of research still needs to be done in that particular field.
10. Appendices should be enlisted in respect of all the technical data in the report.
11. Bibliography of sources consulted is a must for a good report and must necessarily be given.
12. Index is also considered an essential part of a good report and as such must be prepared and appended at the end.
13. Report must be attractive in appearance, neat and clean, whether typed or printed.
14. Calculated confidence limits must be mentioned and the various constraints experienced in conducting the research study may also be stated in the report.
15. Objective of the study, the nature of the problem, the methods employed and the analysis techniques adopted must all be clearly stated in the beginning of the report in the form of introduction

4.8 Need of Executive Summary

Executive summary provides the reader with an overview of the reports essential information.It is possible for us to understand the full content of the report,only by reading the executive summary.

It contains a brief summary about the subject matter, background problem,scope,method of analysis,important findings,conclusion & recommendations.

Need of Executive Summary:

- It crystallizes the thoughts.
- Sets Priorities
- Provides foundations of full plan.

Tips for Effective Executive Summary:

- Last task to be done.
- Use Bulletins.
- It has to provide a fast overview.
- It contain basic information about each section.
- It should contain Important facts.
- It should contain Main conclusions.
- It must be a Short summary.

4.9 Chapterization

The subject matter of the Project is divided into different parts.

Arrange them in a systematic way & mention which aspect of the work will be studied in which chapter.

Eg: Chapter I : Introduction

Chapter II: Company information

Chapter III : Research Methodology

Chapter IV: Findings & Recommendations

4.10 Role of Audience

- The information resulting from the study is ultimately of importance to marketing managers, who will use the results to make decisions.
- Thus, the report has to be understood by them;
- The report should not be too technical and not too much jargon should be used.
- This is a particular difficulty when reporting the results of statistical analysis where there is a high probability that few, if any, of the target audience have a grasp of statistical concepts.
- Hence, for example, there is a need to translate such terms as standard deviation, significance level, confidence interval etc. into everyday language.
- This is sometimes not an easy task but it may be the case that researchers who find it impossible do not themselves have a sufficiently good grasp of the statistical methods they have been using.
- Qualitative research also presents difficulties. The behavioural sciences have their own vocabulary, much of which is not encountered in everyday speech.
- Examples include: cognitive dissonance, evoked set, perception, needs versus wants, self-actualisation.
- It should be noted that these are extreme examples; many words, phrases and concepts used a very precise way by behavioural scientists are also present in everyday speech but often in a less precise or different way. This also presents opportunities for misunderstandings.

4.11 Readability

Readability is the ease with which text can be read and understood. Various factors to measure readability have been used, such as "speed of perception," "perceptibility at a distance," "perceptibility in peripheral vision," "visibility," "the reflex blink technique," "rate of work" (e.g., speed of reading), "eye movements," and "fatigue in reading."

Readability is distinguished from legibility which is a measure of how easily individual letters or characters can be distinguished from each other. Readability can determine the ease in which computer program code can be read by humans, such as through embedded documentation.

Definition

Readability has been defined in various ways, e.g. by: The Literacy Dictionary,Jeanne Chall and Edgar Dale,G. Harry McLaughlin,William DuBay.

Easy reading helps learning and enjoyment, so what we write should be easy to understand

While many writers and speakers since ancient times have used plain language, in the 20th century there was much more focus on reading ease. Much of the research has focused on matching texts to people's reading skills. This has used many successful formulas: in research, government, teaching, publishing, the army, doctors, and business. Many people, and in many languages, have been helped by this. By the year 2000, there were over 1,000 studies on readability formulas in professional journals about their validity and merit.The study of reading is not just in teaching. Research has shown that much money is wasted by companies in making texts hard for the average reader to read.

Comprehension

- ✓ Good Writing varies with the writing objectives.
- ✓ Convey Information Precisely.
- ✓ Avoid Ambiguity, Multiple meanings and allusion.
- ✓ Right words have to be used.
- ✓ Words and sentences are carefully analyzed and edited.

- ✓ Keep the pace.
- ✓ Use ample space and wide margins.
- ✓ Break large units into smaller units.
- ✓ Emphasize Important Materials.

4.12 Tone

Tone and Audience Awareness

We often consider the tone that we're using when we *speak* to others, but we sometimes forget that our **tone—our attitude towards the topic and/or reader**—can also be pretty obvious when we *write*.

To understand the effect that tone can have on your writing, consider what might happen if we attempted to convey the same piece of information using these types of tone:

Casual	Formal
Preachy	Informative
Sarcastic	Serious
Condescending	Understanding

For Example: In 2003, the Supreme Court ruled that the University of Michigan could indeed use racial quotas as part of the law school admissions process.

you can identify the tone being used in each of the similar sentences below:

1. Good luck trying to get into U of M's law school if you're not a minority in this country!
2. Though the quota system at U of M may deter some white male applicants, it's important to remember that race is only one factor in the lengthy admissions process.
3. The university admissions staff appears to be unaware that our forefathers fought and died for equality within this nation—such deserved equality is not possible within the university's prestigious law school.

How does tone relate to "audience awareness"?

One of the most important factors in determining the appropriate tone that you should use in your paper is an understanding of your audience.

To gain an understanding of your audience's expectations, try asking yourself the following questions:

- Is your audience familiar with the text/topic?
- Are they educated?
- What is their background? (Where are they from? What is their political affiliation? What do they do for a living?)
- How old are they?
- Do they agree or disagree with your stance on the issue?

All of these factors influence how your audience will interpret the words on the page; therefore, they should influence your tone as you write them.

Remember! Just as you might speak differently in front of the elderly than you might speak in front of your peers, you may have to adjust your tone and possibly the type of information you provide based on the type of audience you expect to read your essay

5.10 Final Proof

- ✓ Final editing of the draft should be taken up after a gap atleast a day.It helps in identifying mistakes if any.
- ✓ The report has to be assessed in different views.
- ✓ After doing unrelated activities, again the report should be read.
- ✓ After assuring that the draft is complete, write the Executive Summary.

5.11 What Are Research Ethics?

As in other aspects of business, all parties in research should exhibit ethical behavior. **Ethics** are norms or standards of behavior that guide moral choices about our behavior and our relationships with others. The goal of ethics in research is to ensure that no one is harmed or suffers adverse consequences from research activities. This objective is usually achieved. However, unethical activities are pervasive and include violating nondisclosure agreements, breaking respondent confidentiality, misrepresenting results, deceiving people, invoicing irregularities, avoiding legal liability, and more.

The recognition of ethics as a problem for economic organizations was revealed in a survey where 80 percent of the responding organizations reported the adoption of an ethical code. Surprisingly, the evidence that this effort has improved ethical practices is questionable. The same study reports limited success for codes of conduct that attempt to restrain improper behavior.

There is no single approach to ethics. Advocating strict adherence to a set of laws is difficult because of the unforeseen constraint put on researchers. Because of Germany's war history, for example, the government forbids many types of medical research. Consequently, the German people do not benefit from many advances in biotechnology and may have restricted access to genetically altered drugs in the future. Alternatively, relying on each individual's personal sense of morality is equally problematic. Consider the clash between those who believe death is deliverance from a life of suffering and those who value life to the point of preserving it indefinitely through mechanical means. Each value system claims superior knowledge of moral correctness.

Clearly, a middle ground between being completely code governed and ethical relativism is necessary. The foundation for that middle ground is an emerging consensus on ethical standards for researchers. Codes and regulations guide researchers and sponsors. Review boards and peer groups help researchers examine their research proposals for ethical dilemmas. Many design-based ethical problems can be eliminated by careful planning and constant vigilance. In the end, responsible research anticipates ethical dilemmas and attempts to adjust the design, procedures, and protocols during the planning process rather than treating them as an afterthought. Ethical research requires personal integrity from the researcher, the project manager, and the research sponsor.

Because integrity in research is vital, we are discussing its components early in this book and emphasizing ethical behavior throughout our coverage. Our objective is to stimulate an ongoing exchange about values and practical research constraints in the chapters that follow. This chapter is organized around the theme of ethical treatment of respondents, clients, research sponsors, and other researchers. We also highlight appropriate laws and codes, resources for ethical awareness, and cases for application.

5.11.1 Ethical Treatment of Participants

When ethics are discussed in research design, we often think first about protecting the rights of the participant, respondent, or subject. Whether data are gathered in an experiment, interview, observation, or survey, the respondent has many rights to be safeguarded. In general, research must be designed so a respondent does not suffer physical harm, discomfort, pain, embarrassment, or loss of privacy. To safeguard against these, the researcher should follow three guidelines:

1. Explain study benefits.
2. Explain respondent rights and protections.
3. Obtain informed consent.

Benefits

Whenever direct contact is made with a respondent, the researcher should discuss the study's benefits, being careful to neither overstate nor understate the benefits. An interviewer should begin an introduction with his or her name, the name of the research organization, and a brief description of the purpose and benefit of the research. This puts respondents at ease, lets them know to whom they are speaking, and motivates them to answer questions truthfully. In short, knowing why one is being asked questions improves cooperation through honest disclosure of purpose. Inducements to participate, financial or otherwise, should not be disproportionate to the task or presented in a fashion that results in coercion.

Sometimes the actual purpose and benefits of your study or experiment must be concealed from the respondents to avoid introducing bias. The need for concealing objectives leads directly to the problem of deception.

Deception

Deception occurs when the respondents are told only part of the truth or when the truth is fully compromised. Some believe this should never occur. Others suggest two reason for deception: (1) to prevent biasing the respondents before the survey or experiment and (2) to protect the confidentiality of a third party (e.g., the sponsor).

Deception should not be used in an attempt to improve response rates. The benefits to be gained by deception should be balanced against the risks to the respondents. When possible, an experiment or interview should be redesigned to reduce reliance on deception. In addition, the respondents' rights and well-being must be adequately protected. In instances where deception in an experiment could produce anxiety, a subject's medical condition should be checked to ensure that no adverse physical harm follows. The American Psychological Association's Ethics Code states that the use of deception is inappropriate unless deceptive techniques are justified by the study's expected scientific, educational, or applied value and equally effective alternatives that do not use deception are not feasible.3 And finally, the respondents must have given their informed consent before participating in the research.

Informed Consent

Securing **informed consent** from respondents is a matter of fully disclosing the procedures of the proposed survey or other research design before requesting permission to proceed with the study. There are exceptions that argue for a signed consent form. When dealing with children, it is wise to have a parent or other person with legal standing sign a consent form. When doing research with medical or psychological ramifications, it is also wise to have a consent form. If there is a chance the data could harm the respondent or if the researchers offer only limited protection of confidentiality, a signed form detailing the types of limits should be obtained. For most business research, oral consent is sufficient

Debriefing Participants

Debriefing involves several activities following the collection of data:

- Explanation of any deception.

• Description of the hypothesis, goal, or purpose of the study.
• Poststudy sharing of results.
• Poststudy follow-up medical or psychological attention.

First, the researcher shares the truth of any deception with the participants and the reasons for using deception in the context of the study's goals. In cases where severe reactions occur, follow-up medical or psychological attention should be provided to continue to ensure the participants remain unharmed by the research.

Even when research does not deceive the respondents, it is a good practice to offer them follow-up information. This retains the goodwill of the respondent, providing an incentive to participate in future research projects. For surveys and interviews, respondents can be offered a brief report of the findings. Usually, they will not request additional information. Occasionally, however, the research will be of particular interest to a respondent. A simple set of descriptive charts or data tables can be generated for such an individual. For experiments, all participants should be debriefed in order to put the experiment into context. Debriefing usually includes a description of the hypothesis being tested and the purpose of the study. Participants who were not deceived still benefit from the debriefing session. They will be able to understand why the experiment was created.

The researchers also gain important insight into what the participants thought about during and after the experiment. This may lead to modifications in future research designs. Like survey and interview respondents, participants in experiments and observational studies should be offered a report of the findings. To what extent do debriefing and informed consent reduce the effects of deception? Research suggests that the majority of participants does not resent temporary deception and may have more positive feelings about the value of the research after debriefing than those who didn't participate in the study. Nevertheless; deception is an ethically thorny issue and should be addressed with sensitivity and concern for research participants.

Rights to Privacy

Privacy laws in the United States are taken seriously. All individuals have a right to privacy, and researchers must respect that right. The importance of the right to privacy is illustrated with an example.

An employee of MonsterVideo, a large video company, is also a student at the local university. For a research project, this student and his team members decide to compare the video-viewing habits of a sample of customers. Using telephone interviews, the students begin their research. After inquiring about people's viewing habits and the frequency of rentals versus purchases, the students move on to the types of films people watch. They find that most respondents answer questions about their preferences for children's shows, classics, bestsellers, mysteries, and science fiction. But the cooperation ceases when the students question the viewing frequency of pornographic movies.

Without the guarantee of privacy, most people will not answer these kinds of questions truthfully, if at all. The study then loses key data. The privacy guarantee is important not only to retain validity of the research but also to protect respondents. In the previous example, imagine the harm that could be caused by releasing information on the viewing habits of certain citizens. Clearly, the confidentiality of survey answers is an important aspect of the respondents' right to privacy.

Once the guarantee of **confidentiality** is given, protecting that confidentiality is essential. The researcher protects respondent confidentiality in several ways:

• Obtaining signed nondisclosure documents.

• Restricting access to respondent identification.

• Revealing respondent information only with written consent.

• Restricting access to data instruments where the respondent is identified.

• Nondisclosure of data subsets.

Researchers should restrict access to information that reveals names, telephone numbers, addresses, or other identifying features. Only researchers who have signed nondisclosure, confidentiality forms should be allowed access to the data. Links between the data or database and the identifying information file should be weakened. Individual interview response sheets should be inaccessible to everyone except the editors and data entry personnel. Occasionally, data collection instruments should be destroyed once the data are in a data file. Data files that make it easy to reconstruct the profiles or identification of individual respondents should be carefully controlled. For very small groups, data should not be made available because it is often easy to pinpoint a person within the group. Employee-satisfaction survey feedback in small units can be easily used to identify an individual through descriptive statistics alone. These last two protections are particularly important in human resources research.

But privacy is more than confidentiality. A **right to privacy** means one has the right to refuse to be interviewed or to refuse to answer any question in an interview. Potential participants have a right to privacy in their own homes, including not admitting researchers and not answering telephones. And they have the right to engage in private behavior in private places without fear of observation. To address these rights, ethical researchers do the following:

• Inform respondents of their right to refuse to answer any questions or participate in the study.
• Obtain permission to interview respondents.
• Schedule field and phone interviews.
• Limit the time required for participation.
• Restrict observation to public behavior only.

Data Collection in Cyberspace

Some ethicists argue that the very conduct that results in resistance from respondents—interference, invasiveness in their lives, denial of privacy rights—has encouraged researchers to investigate topics online that have long been the principal commodity of offline investigation. The novelty and convenience of communicating by computer has led researchers to cyberspace in search of abundant sources of data. Whether we call it the "wired society," "digital life," "computer-mediated communication," or "cyberculture," the growth of cyberstudies causes us to question how we gather data online, deal with participants, and present results.

In a special ethics issue of *Information Society,* scholars involved in cyberspace research concluded:

All participants agree that research in cyberspace provides no special dispensation to ignore ethical precepts. Researchers are obligated to protect human subjects and "do right" in electronic venues as in more conventional ones. Second, each participant recognizes that cyberspace poses complex ethical issues that may lack exact analogs in other types of inquiry. The ease of covert observation, the occasional blurry distinction between public and private venues, and the difficulty of obtaining the informed consent of subjects make cyber-research particularly vulnerable to ethical breaches by even the most scrupulous scholars. Third, all recognize that because research procedures or activities may be permissible or not precluded by law or policy, it does not follow that they are necessarily ethical or allowable. Fourth, all agree that the individual researcher has the ultimate responsibility for assuring that inquiry is not only done honestly, but done with ethical integrity.7

Issues relating to cyberspace in research also relate to data mining. The information collection devices available today were once the tools of the spy, the science fiction protagonist, or the superhero. Smart cards, biometrics (finger printing, retinal scans, facial recognition), electronic monitoring (closed circuit television, digital camera monitoring), global surveillance, and genetic identification (DNA) are just some of the technological tools being used by today's organizations to track and understand employees, customers, and suppliers. The data mining of all this information, collected from advanced and not necessarily obvious sources, offers infinite possibilities for research abuse.

The primary ethical data-mining issues in cyberspace are privacy and consent. Smart cards, those ubiquitous credit-card-sized devices that imbed personal information on a computer chip that is then matched to purchase, employment, or other behavior data, offer the researcher implied consent to participant surveillance. But the benefits of card use may be enough to hide from an unsuspecting user the data-mining purpose of the card. For example, The Kroger Co., one of the largest grocers in the United States, offers significant discounts for enrollment in its Kroger Plus Shopper's Card program.8 Retailers, wholesalers, medical and legal service providers, schools, government agencies, and resorts, to name a few, use smart cards or their equivalent.

In most instances, participants provide, although sometimes grudgingly, the personal information requested by enrollment procedures. But in others, like when smart cards are used with those convicted of crimes and sentenced to municipal or state correction facilities or those attending specific schools, enrollment is mandatory. In some instances, mandatory sharing of information is initially for personal welfare and safety—like when you admit yourself for a medical procedure and provide detailed information about medication or prior surgery. But in others, enrollment is for less critical but potentially attractive monetary benefits—for example, free car care services when a smart card is included with the keys to a new vehicle. The bottom line is that the organization collecting the information gains a major benefit: the potential for better understanding and competitive advantage.

General privacy laws may not be sufficient to protect the unsuspecting in the cyberspace realm of data collection. The 15 European Union (EU) countries started the new century by passing the European Commission's Data Protection Directive. Under the directive, commissioners can prosecute companies and block websites that fail to live up to its strict privacy standards. Specifically, the directive prohibits the transmission of names, addresses, ethnicity, and other personal information to any country that fails to provide adequate data protection. This includes direct mail lists, hotel and travel reservations, medical and work records, and orders for products among a host of others.

U.S. industry and government agencies have resisted regulation of data flow. But the EU insists that it is the right of every citizen to find out what information about them is in a database and correct any mistakes. Few U.S. companies would willingly offer such access due to the high cost; 10 a perfect example of this reluctance is the difficulty individuals have correcting erroneous credit reports, even when such information is based on stolen personal identity or credit card transactions.

Yet questions remain regarding the definition of specific ethical behaviors for cyber-research, the sufficiency of existing professional guidelines, and the issue of ultimate responsibility for respondents. If researchers are responsible for the ethical conduct of their research, are they solely responsible for the burden of protecting participants from every conceivable harm?

Ethics and the Sponsor

There are also ethical considerations to keep in mind when dealing with the research client or sponsor. Whether undertaking product, market, personnel, financial, or other research, a sponsor has the right to receive ethically conducted research.

Confidentiality Some sponsors wish to undertake research without revealing themselves. They have a right to several types of confidentiality, including sponsor nondisclosure, purpose nondisclosure, and findings nondisclosure.

Companies have a right to dissociate themselves from the sponsorship of a research project. This type of confidentiality is called **sponsor nondisclosure.** Due to the sensitive nature of the management dilemma or the research question, sponsors may hire an outside consulting or research firm to complete research projects. This is often done when a company is testing a new product idea, to avoid potential consumers from being influenced by the company's current image or industry standing. Or if a company is contemplating entering a new market, it may not wish to reveal its plans

to competitors. In such cases, it is the responsibility of the researcher to respect this desire and devise a plan that safeguards the identity of the research sponsor.

Purpose nondisclosure involves protecting the purpose of the study or its details. A research sponsor may be testing a new idea that is not yet patented and may not want the competition to know of its plans. It may be investigating employee complaints and may not want to spark union activity. Or the sponsor might be contemplating a new public stock offering, where advance disclosure would spark the interest of authorities or cost the firm thousands or millions of dollars. Finally, even if a sponsor feels no need to hide its identity or the study's purpose, most sponsors want the research data and findings to be confidential, at least until the management decision is made. Thus sponsors usually demand and receive **findings nondisclosure** between themselves or their researchers and any interested but unapproved parties.

Right to Quality Research

An important ethical consideration for the researcher and the sponsor is the sponsor's **right to quality** research. This right entails:

- Providing a research design appropriate for the research question.
- Maximizing the sponsor's value for the resources expended.
- Providing data handling and reporting techniques appropriate for the data collected.

From the proposal through the design to data analysis and final reporting, the researcher guides the sponsor on the proper techniques and interpretations. Often sponsors will have heard about a sophisticated data-handling technique and will want it used even when it is inappropriate for the problem at hand. The researcher should guide the sponsor so this does not occur. The researcher should propose the design most suitable for the problem. The researcher should not propose activities designed to maximize researcher revenue or minimize researcher effort at the sponsor's expense.

Finally, we have all heard the remark, "You can lie with statistics." It is the researcher's responsibility to prevent that from occurring. The ethical researcher always follows the analytical rules and conditions for results to be valid. The ethical researcher reports findings in ways that minimize the drawing of false conclusions. The ethical researcher also uses charts, graphs, and tables to show the data objectively, despite the sponsor's preferred outcomes.

Sponsor's Ethics

Occasionally, research specialists may be asked by sponsors to participate in unethical behavior. Compliance by the researcher would be a breach of ethical standards. Some examples to be avoided are:

- Violating respondent confidentiality.
- Changing data or creating false data to meet a desired objective.
- Changing data presentations or interpretations.
- Interpreting data from a biased perspective.
- Omitting sections of data analysis and conclusions.
- Making recommendations beyond the scope of the data collected.

Let's examine the effects of complying with these types of coercion. A sponsor may offer a promotion, future contracts, or a larger payment for the existing research contract; or the sponsor may threaten to fire the researcher or tarnish the researcher's reputation. For some researchers, the request may seem trivial and the reward high. But imagine, for a moment, what will happen to the researcher who changes research results. Although there is a promise of future research, can the sponsor ever trust that researcher again? If the researcher's ethical standards are for sale, which sponsor might be the highest bidder next time? Although the promise of future contracts seems enticing, it is unlikely to be kept. Each coercive reward or punishment has an equally poor outcome. The "greater than" contracted payment is a payoff. The threats to one's professional reputation cannot be carried out effectively by a sponsor who has tried to purchase you. So the rewards for behaving unethically are illusory.

What's the ethical course? Often, it requires confronting the sponsor's demand and taking the following actions:

• Educate the sponsor to the purpose of research.
• Explain the researcher's role in fact finding versus the sponsor's role in decision making.
• Explain how distorting the truth or breaking faith with respondents leads to future problems.
• Failing moral suasion, terminate the relationship with the sponsor.

Researchers and Team Members

Another ethical responsibility of researchers is their team's safety as well as their own. In addition, the responsibility for ethical behavior rests with the researcher who, along with assistants, is charged with protecting the anonymity of both the sponsor and the respondent.

Safety

It is the researcher's responsibility to design a project so the safety of all interviewers, surveyors, experimenters, or observers is protected. Several factors may be important to consider in ensuring a researcher's **right to safety.** Some urban areas and undeveloped rural areas may be unsafe for research assistants. If, for example, the researcher must personally interview people in a high-crime district, it is reasonable to provide a second team member to protect the researcher. Alternatively, if an assistant feels unsafe after visiting a neighborhood by car, an alternate researcher should be assigned to the destination. It is unethical to require staff members to enter an environment where they feel physically threatened. Researchers who are insensitive to these concerns face both research and legal risks—the least of which involves having interviewers falsify instruments.

Ethical Behavior of Assistants

Researchers should require ethical compliance from team members just as sponsors expect ethical behavior from the researcher. Assistants are expected to carry out the sampling plan, to interview or observe respondents without bias, and to accurately record all necessary data. Unethical behavior, such as filling in an interview sheet without having asked the respondent the questions, cannot be tolerated. The behavior of the assistants is under the direct control of the responsible researcher or field supervisor. If an assistant behaves improperly in an interview or shares a respondent's interview sheet with an unauthorized person, it is the researcher's responsibility. Consequently, all assistants should be well trained and supervised.

Protection of Anonymity

As discussed previously, researchers and assistants protect the confidentiality of the sponsor's information and the anonymity of the respondents. Each researcher handling data should be required to sign a confidentiality and nondisclosure statement.

Professional Standards

Various standards of ethics exist for the professional researcher. Many corporations, professional associations, and universities have a **code of ethics.** The impetus for these policies and standards can be traced to two documents: the Belmont Report of 1979 and the *Federal Register* of 1991.12 Society or association guidelines include ethical standards for the conduct of research. One comprehensive source contains 51 official codes of ethics issued by 45 associations in business, health, and law. The business section of this source consists of ethics standards for

Accounting—American Institute of Certified Public Accountants.
Advertising—American Association of Advertising Agencies; Direct Marketing Association.
Banking—American Bankers Association.
Engineering—American Association of Engineering Societies; National Society of Professional Engineers.

Financial planning—Association for Investment Management and Research; Certified Financial Planner Board of Standards/Institute of Certified Financial Planners; International Association for Financial Planning.
Human resources—American Society for Public Administration; Society for Human Resource Management.
Insurance—American Institute for Chartered Property Casualty Underwriters; American Society of Chartered Life Underwriters and Chartered Financial Consultants.
Management—Academy of Management; The Business Roundtable.
Real estate—National Association of Realtors. Other professional associations' codes have detailed research sections: the American Marketing Association, the American Association for Public Opinion Research, the American Psychological Association, the American Political Science Association, the American Sociological Association, and the Society of Competitive Intelligence Professionals. These associations update their codes frequently.

We commend professional societies and business organizations for developing standards. However, without enforcement, standards are ineffectual. Effective codes (1) are regulative, (2) protect the public interest and the interests of the profession served by the code, (3) are behavior-specific, and (4) are *enforceable.* A study that assessed the effects of personal and professional values on ethical consulting behavior concluded:

The findings of this study cast some doubt on the effectiveness of professional codes of ethics and corporate policies that attempt to deal with ethical dilemmas faced by business consultants. A mere codification of ethical values of the profession or organization may not counteract ethical ambivalence created and maintained through reward systems. The results suggest that unless ethical codes and policies are consistently reinforced with a significant reward and punishment structure and truly integrated into the business culture, these mechanisms would be of limited value in actually regulating unethical conduct

Federal, state, and local governments also have laws, policies, and procedures in place to regulate research on human beings. The U.S. government began a process that covers all research having federal support. Initially implemented in 1966, the Institutional Review Boards (IRBs) engage in a risk assessment and benefit analysis review of proposed research. The Department of Health and Human Services (HHS) translated the federal regulations into policy. Most other federal and state agencies follow the HHS-developed guidelines.

Since 1981, the review requirement has been relaxed so research that is routine no longer needs to go through the complete process.15 Each institution receiving funding from HHS or doing research for HHS is required to have its own IRB to review research proposals. Many institutions require all research, whether funded or unfunded by the government, to undergo review by the local IRB. The IRBs concentrate on two areas. First is the guarantee of obtaining complete, informed consent from participants. This can be traced to the first of 10 points in the

Nuremberg Code.16 Complete informed consent has four characteristics:
1. The respondent must be competent to give consent.
2. Consent must be voluntary, free from coercion, force, requirements, and so forth.
3. Respondents must be adequately informed to make a decision.
4. Respondents should know the possible risks or outcomes associated with the research.

The second item of interest to the IRB is the risk assessment and benefit analysis review. In the review, risks are considered when they add to the normal risk of daily life. Significantly, the only benefit considered is the immediate importance of the knowledge to be gained. Possible long-term benefits from applying the knowledge that may be gained in the research are not considered.

Resources for Ethical Awareness

There is optimism for improving ethical awareness. According to the Center for Business Ethics at Bentley College, over a third of Fortune 500 companies have ethics officers, a substantial rise. Almost 90 percent of business schools have ethics programs, up from a handful several years ago.

5.12 Subjectivity and objectivity in Research:

- ✓ Subjectivity is the condition of being a subject: ie.Quality of possessing perspectives,experiences,feelings,believes,desires and power.

- ✓ Subjectivity is used as an explanation for what influences and inform people judgements about truth or Reality.
- ✓ It is the collection of the Perception,experiences,expectations,personal or cultural understanding,and believes specific to a person.
- ✓ It is often used in contrast to the term objectivity,Which is described as a view of truth or reality.
- ✓ Which is free of any Individuals influence.

Objectivity:

- ➢ The idea of objectivity assumes that a truth or independent reality exists outside of any investigation or observation.
- ➢ The researcher task in this model is to uncover this reality without contaminating it in any way.

Difficulties in Achieving Objectivity of Research:

- o Personal prejudices and Bias
- o Value Judgments
- o Ethical dilemma
- o Complexity of Social Phenomenon

Objectivity is personal neutrality; it allows the facts to speak for themselves and not be influenced by the personal values and biases of the researcher. (Macionis, 20). It can also be interpreted as mind-independent, because it is information that is not being altered based on an opinion. ("Proof that Reality is Mind-Independent"). For example, when looking at a painting, a person would use objectivity to describe the texture, color, and form. These are all facts that are common and cannot be changed.

Subjectivity is judgment based on individual personal impressions and feelings and opinions rather than external facts. (wordnetweb.com) This can be considered mind-dependent, because one is not using a fact, they are using their personal opinion. ("Proof that Reality is Mind-Independent"). For example, if your favorite color is blue, then you are more likely to buy a blue sweater versus a purple sweater.

I believe that objectivity exists in some ways, and also does not exist in others. When doing a social research experiment, it is good to have unbiased results because it will create a more accurate study. However, most of the time this is impossible to do because every person has some kind of subjectivity in them based on their life experiences and opinions. So, in order to research a topic, social researches turn to experiments, surveys, participant observations, and the use of existing sources as guides in research methods.By having a variety of ways to get results, social researches can use these methods to draw conclusions on a certain topic

Types of Statistical Tests

Types of statistical tests: There are a wide range of statistical tests. The decision of which statistical test to use depends on the research design, the distribution of the data, and the type of variable. In general, if the data is normally distributed you will choose from parametric tests. If the data is non-normal you choose from the set of non-parametric tests. Below is a table listing just a few common statistical tests and their use

Type of Test:	Use:
Correlation	These tests look for an association between variables
Pearson correlation	Tests for the strength of the association between two continuous variables
Spearman correlation	Tests for the strength of the association between two ordinal variables (does not rely on the assumption of normal distributed data)
Chi-square	Tests for the strength of the association between two categorical variables
Comparison of Means: look for the difference between the means of variables	
Paired T-test	Tests for difference between two related variables
Independent T-test	Tests for difference between two independent variables
ANOVA	Tests the difference between group means **after** any other variance in the outcome variable is accounted for
Regression: assess if change in one variable predicts change in another variable	
Simple regression	Tests how change in the predictor variable predicts the level of change in the outcome variable
Multiple regression	Tests how change in the combination of two or more predictor variables predict the level of change in the outcome variable
Non-parametric: are used when the data does not meet assumptions required for parametric tests	
Wilcoxon rank-sum test	Tests for difference between two independent variables - takes into account magnitude and direction of difference
Wilcoxon sign-rank test	Tests for difference between two related variables - takes into account magnitude and direction of difference
Sign test	Tests if two related variables are different – ignores magnitude of change, only takes into account direction

Chi-Square Test for Independence

The test is applied when you have two categorical variables from a single population. It is used to determine whether there is a significant association between the two variables.

For example, in an election survey, voters might be classified by gender (male or female) and voting preference (Democrat, Republican, or Independent). We could use a chi-square test for independence to determine whether gender is related to voting preference.

When to Use Chi-Square Test for Independence

The test procedure described in this lesson is appropriate when the following conditions are met:

- The sampling method is simple random sampling.
- Each population is at least 10 times as large as its respective sample.
- The variables under study are each categorical.

- If sample data are displayed in a contingency table, the expected frequency count for each cell of the table is at least 5.

Steps to Solve a Chi-square Test problem

1) Determine the ***null hypothesis*** and ***alternate hypothesis***

- *Null(*H0*):two variables are independent
- Alternate(*H1*):two variables are not independent

2) Calculate the ***Expected Value*** matrix

*Exp.Value=(rowtotal)(columntotal)/samplesize

*The expected value of each cell must be 5 or greater for ChiSquare test. To be valid

3)Determine the***Chi-Squarestatistic***forthedata

*$\chi 2 = \Sigma(Observed - Expected)2Expected$

4)Find the ***DegreesofFreedom***

*D.F.=(rows–1)(columns–1)

5) Compare the ***Chi-Square statistic***

- (your calculated value) to the ***critical value*** at the appropriate ***level of significance*** (the value from the chart in the book in the
- Table onp.A26)(*default value is.050)*

6) If ***Chi- Square statistic<critical value***, accept the ***Null hypothesis*** (variablesareindependent)

- If ***Chi-Square statistic>criticalvalue***,reject the ***Null hypothesis*** and accept the ***alternate hypothesis*** (variables Are not independent)

Problem No:1

A public opinion poll surveyed a simple random sample of 1000 voters. Respondents were classified by gender (male or female) and by voting preference (Republican, Democrat, or Independent)..

	Voting Preferences			Row total
	Republican	Democrat	Independent	
Male	200	150	50	400
Female	250	300	50	600
Column total	450	450	100	1000

Is there a gender gap? Do the men's voting preferences differ significantly from the women's preferences? Use a 0.05 level of significance.

Solution

The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

H_0: Gender and voting preferences are independent.
H_a: Gender and voting preferences are not independent.

- **Formulate an analysis plan**. For this analysis, the significance level is 0.05. Using sample data, we will conduct a chi-square test for independence.
- **Analyze sample data**. Applying the chi-square test for independence to sample data, we compute the degrees of freedom, the expected frequency counts, and the chi-square test statistic. Based on the chi-square statistic and the degrees of freedom, we determine the P-value.

$$DF = (r - 1) * (c - 1) = (2 - 1) * (3 - 1) = 2$$

$$E_{r,c} = (n_r * n_c) / n$$
$$E_{1,1} = (400 * 450) / 1000 = 180000/1000 = 180$$
$$E_{1,2} = (400 * 450) / 1000 = 180000/1000 = 180$$
$$E_{1,3} = (400 * 100) / 1000 = 40000/1000 = 40$$
$$E_{2,1} = (600 * 450) / 1000 = 270000/1000 = 270$$
$$E_{2,2} = (600 * 450) / 1000 = 270000/1000 = 270$$
$$E_{2,3} = (600 * 100) / 1000 = 60000/1000 = 60$$

$$X^2 = \Sigma [(O_{r,c} - E_{r,c})^2 / E_{r,c}]$$
$$X^2 = (200 - 180)^2/180 + (150 - 180)^2/180 + (50 - 40)^2/40$$
$$+ (250 - 270)^2/270 + (300 - 270)^2/270 + (50 - 60)^2/60$$
$$X^2 = 400/180 + 900/180 + 100/40 + 400/270 + 900/270 + 100/60$$
$$X^2 = 2.22 + 5.00 + 2.50 + 1.48 + 3.33 + 1.67 = 16.2$$

where DF is the degrees of freedom, r is the number of levels of gender, c is the number of levels of the voting preference, n_r is the number of observations from level *r* of gender, n_c is the number of observations from level *c* of voting preference, n is the number of observations in the sample, $E_{r,c}$ is the expected frequency count when gender is level *r* and voting preference is level *c*, and $O_{r,c}$ is the observed frequency count when gender is level *r* voting preference is level *c*.

The P-value is the probability that a chi-square statistic having 2 degrees of freedom is more extreme than 16.2.

We use the Chi-Square Distribution Calculator to find $P(X^2 > 16.2) = 0.0003$.

- **Interpret results**. Since the P-value (0.0003) is less than the significance level (0.05), we cannot accept the null hypothesis. Thus, we conclude that there is a relationship between gender and voting preference..

Correlation Coefficient:

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables.

Correlation is a bivariate analysis that measures the strengths of association between two variables. In statistics, the value of the correlation coefficient varies between +1 and -1. When the value of the correlation coefficient lies around ± 1, then it is said to be a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. Usually, in statistics, we measure three types of correlations: Pearson correlation, Kendall rank correlation and Spearman correlation.

***Kendall rank correlation*:** Kendall rank correlation is a non-parametric test that measures the strength of dependence between two variables. If we consider two samples, a and b, where each sample size is n, we know that the total number of pairings with a b is $n(n\text{-}1)/2$. The following formula is used to calculate the value of Kendall rank correlation:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Simple Correlation problems:

- **Examples:**
  Calculate and analyze the correlation coefficient between the number of study hours and the number of sleeping hours of different students.

Number of Study hours	**2**	**4**	**6**	**8**	**10**
Number of sleeping hours	**10**	**9**	**8**	**7**	**6**

- **Solution:**
  The necessary calculation is given below:

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
2	**10**	**-4**	**2**	**-8**	**16**	**4**
4	**9**	**-2**	**1**	**-2**	**4**	**1**
6	**8**	**0**	**0**	**0**	**0**	**0**
8	**7**	**2**	**-1**	**-2**	**4**	**1**
10	**6**	**4**	**-2**	**-8**	**16**	**1**
$\Sigma X = 30$	$\Sigma Y = 40$	$\Sigma(X-\bar{X}) = 0$	$\Sigma(Y-\bar{Y}) = 0$	$\Sigma(X-\bar{X})(Y-\bar{Y}) = -20$	$\Sigma(X-\bar{X})^2 = 40$	$\Sigma(Y-\bar{Y})^2 = 10$

- $\bar{X} = \frac{\Sigma X}{n} = \frac{30}{5} = 6 \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{40}{5} = 8$ and

$$r_{XY} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2 \Sigma(Y-\bar{Y})^2}} = \frac{-20}{20} = -1$$

  There is perfect negative correlation between the number of study hours and the number of sleeping hours….

- 2) From the following data, compute the coefficient of correlation between X and Y:

	X Series	Y Series
Number if Items	**15**	**15**

Arithmetic Mean	**25**	**18**
Sum of Square Deviations	**136**	**138**

Summation of products of deviations of *X* and *Y* series from their arithmetic means = **122**.

Solution:

Here $n = 15,\ \bar{X} = 25,\ \bar{Y} = 18,\ \Sigma(X-\bar{X})^2 = \Sigma(Y-\bar{Y})^2 = 138$

$\Sigma(X-\bar{X})^2(Y-\bar{Y})^2 = 122$ and hence

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\Sigma(Y-\bar{Y})^2}} = \frac{122}{\sqrt{(136)(138)}} = \frac{122}{137} = 0.89$$

Pearson's Correlation Coefficient:

There are several types of correlation coefficient: Pearson's correlation or Pearson correlation is a **correlation coefficient** commonly used in linear regression.

Sample question: Find the value of the correlation coefficient from the following table:

Subject	Age x	Glucose Level y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Step 1:*Make a chart.* Use the given data, and add three more columns: xy, x^2, and y^2.

Subject	Age x	Glucose Level y	xy	x^2	y^2
1	43	99			
2	21	65			
3	25	79			
4	42	75			
5	57	87			
6	59	81			

Step 2::*Multiply x and y together to fill the xy column. For example, row 1 would be 43 × 99 = **4,257**.*

Subject	Age x	Glucose Level y	xy	x^2	y^2
1	43	99	4257		
2	21	65	1365		
3	25	79	1975		
4	42	75	3150		
5	57	87	4959		
6	59	81	4779		

Step 3: *Take the square of the numbers in the x column, and put the result in the x^2 column.*

Subject	Age x	Glucose Level y	xy	x^2	y^2
1	43	99	4257	1849	
2	21	65	1365	441	
3	25	79	1975	625	
4	42	75	3150	1764	
5	57	87	4959	3249	
6	59	81	4779	3481	

Step 4: *Take the square of the numbers in the y column, and put the result in the y^2 column.*

Subject	Age x	Glucose Level y	xy	x^2	y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561

Step 5: *Add up all of the numbers in the columns and put the result at the bottom.^{2} column.* The Greek letter sigma (Σ) is a short way of saying "sum of."

Subject	Age x	Glucose Level y	xy	x^2	y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

Step 6:*Use the following correlation coefficient formula.*

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

The answer is: **2868 / 5413.27 = 0.529809**

From our table:

- $\Sigma x = 247$
- $\Sigma y = 486$
- $\Sigma xy = 20{,}485$
- $\Sigma x^2 = 11{,}409$
- $\Sigma y^2 = 40{,}022$
- n is the sample size, in our case = 6

The correlation coefficient =

- $6(20{,}485) – (247 \times 486) / [\sqrt{[[6(11{,}409) – (247^2)] \times [6(40{,}022) – 486^2]]}]$

  $=0.5298$

The range of the correlation coefficient is from -1 to 1. Our result is 0.5298 or 52.98%, which means the variables have a moderate positive correlation.

Spearman's Rank Correlation Coefficient

The Spearman's Rank Correlation Coefficient is used to discover the strength of a link between two sets of data. This example looks at the strength of the link between the price of a convenience item (a 50cl bottle of water) and distance from the Contemporary Art Museum in El Raval, Barcelona.

A correlation can easily be drawn as a scatter graph, but the most precise way to compare several **pairs of data** is to use a statistical test - this establishes whether the correlation is really significant or if it could have been the result of chance alone.

Spearman's Rank correlation coefficient is a technique which can be used to summarise the strength and direction (negative or positive) of a relationship between two variables.

The result will always be between 1 and minus 1.

Method - calculating the coefficient

- Create a table from your data.
- Rank the two data sets. Ranking is achieved by giving the ranking '1' to the biggest number in a column, '2' to the second biggest value and so on. The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements.
- Tied scores are given the mean (average) rank. For example, the three tied scores of 1 euro in the example below are ranked fifth in order of price, but occupy three positions (fifth, sixth and seventh) in a ranking hierarchy of ten. The mean rank in this case is calculated as (5+6+7) ÷ 3 = 6.
- Find the difference in the ranks (d): This is the difference between the ranks of the two values on each row of the table. The rank of the second value (price) is subtracted from the rank of the first (distance from the museum).
- Square the differences (d^2) To remove negative values and then sum them (Σd^2).

Convenience Store	**Distance from CAM (m)**	**Rank distance**	**Price of 50cl bottle (€)**	**Rank price**	**Difference between ranks (d)**	**d^2**
1	50	10	1.80	2	8	64
2	175	9	1.20	3.5	5.5	30.25
3	270	8	2.00	1	7	49
4	375	7	1.00	6	1	1
5	425	6	1.00	6	0	0
6	580	5	1.20	3.5	1.5	2.25
7	710	4	0.80	9	-5	25
8	790	3	0.60	10	-7	49
9	890	2	1.00	6	-4	16
10	980	1	0.85	8	-7	49
						$\Sigma d^2 = 285.5$

Data Table: Spearman's Rank Correlation

- Calculate the coefficient (***R***) using the formula below. The answer will always be between 1.0 (a perfect positive correlation) and -1.0 (a perfect negative correlation).

When written in mathematical notation the Spearman Rank formula looks like this :

$$(R) = 1$$

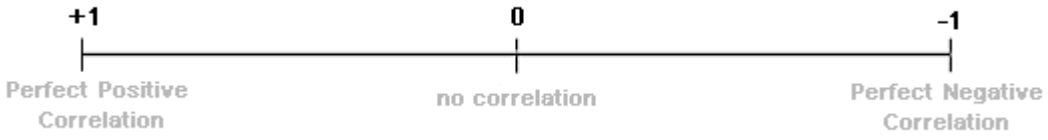
Now to put all these values into the formula.

- Find the value of all the d^2 values by adding up all the values in the Difference^2 column. In our example this is **285.5**. Multiplying this by **6** gives **1713**.
- Now for the bottom line of the equation. The value *n* is the number of sites at which you took measurements. This, in our example is **10**. Substituting these values into $n^3 - n$ we get **1000 - 10**
- We now have the formula: ***R*** = 1 - (1713/990) which gives a value for ***R***:

  1 - 1.73 = -0.73

What does this *R* value of -0.73 mean?

The closer ***R*** is to +1 or -1, the stronger the likely correlation. A perfect positive correlation is +1 and a perfect negative correlation is -1. The ***R*** value of -0.73 suggests a fairly strong negative relationship.



A further technique is now required to test the **significance** of the relationship.

The ***R*** value of **-0.73** must be looked up on the Spearman Rank significance table below as follows:

- Work out the 'degrees of freedom' you need to use. This is the number of pairs in your sample minus 2 (n-2). In the example it is 8 (10 - 2).
- Now plot your result on the table.
- If it is below the line marked 5%, then it is possible your result was the product of chance and you must reject the hypothesis.
- If it is above the 0.1% significance level, then we can be 99.9% confident the correlation has not occurred by chance.
- If it is above 1%, but below 0.1%, you can say you are 99% confident.
- If it is above 5%, but below 1%, you can say you are 95% confident (i.e. statistically there is a 5% likelihood the result occurred by chance).

In the example, the value 0.73 gives a significance level of slightly less than 5%. That means that the probability of the relationship you have found being a chance event is **about 5 in a 100**. You are 95% certain that your hypothesis is correct.

<u>Regression Problems:</u>

1. A researcher wants to know if there is a relationship between the number of shopping centers in a state and the retail sales (in billions $) of that state. A random sample of 8 states is listed below. After determining, via a scatter-plot, that the data followed a linear pattern, the regression line was found. Using the given data and the given regression output answer the following questions.

State	Num	Sales
1	630	15.5
2	370	7.5
3	616	13.9
4	700	18.7
5	430	8.2
6	568	13.2
7	1200	23.0
8	2976	87.3

Output

a = -4.930

b = 0.030

r = 0.991

a. What is the equation of the regression line?

b. Interpret the slope in the words of the problem.

c. Find r^2 and interpret its meaning in the words of the problem.

d. Find the error for predicting the sales of a state with 1200 stores.

e. Use the regression line to predict the sales for a state with 100 stores

Solution:

a. y-hat = -4.930 + 0.030x
b. Slope = 0.030 means that for every increase of 1 shopping center retail sales increases 0.030 billion dollars, on average.
c. r^2 = .982 means that there is a 98.2% reduction in error in predicting retail sales by using number of shopping centers.
d. x = 1200, y = 23.0 y-hat = 31.07 so error = 23.0 – 31.07 = -8.07
e. This is an example of extrapolation so you should not do it.

2. A pharmaceutical company is investigating the relationship between advertising expenditures and the sales of some over-the-counter (OTC) drugs. The following data represents a sample of 10 common OTC drugs. Find the equation of the regression line, using Advertising dollars as the independent variable and Sales as the response variable. Interpret the slope of the line in the words of the problem. Find r^2 and interpret it in the words of the problem. Use the line to predict the Sales if Advertising dollars = $50 million. Note that AD = Advertising dollars in millions and S = Sales in millions $.

ADVERTISEMENT	SALES
22	64

25	74
29	82
35	90
38	100
42	120
46	120
52	142
65	180
88	230

Calculator Output

a = 6.629, b = 2.569, r = .996

Solution :

y-hat = 6.629 + 2.569x
The slope = 2.569 means that for every increase of $1 million Advertising, sales increases $2.569 million on average.
r^2 = .993 so there is a 99.3% reduction in error for predicting Sales using advertising dollars.
y-hat = 6.629 + 2.596(50) = 135.079

ANOVA:

Problem: Susan Sound predicts that students will learn most effectively with a constant background sound, as opposed to an unpredictable sound or no sound at all. She randomly divides twenty-four students into three groups of eight. All students study a passage of text for 30 minutes. Those in group 1 study with background sound at a constant volume in the background. Those in group 2 study with noise that changes volume periodically. Those in group 3 study with no sound at all. After studying, all students take a 10 point multiple choice test over the material.

Their scores follow:

Group	test scores							
1) constant sound	7	4	6	8	6	6	2	9
2) random sound	5	5	3	4	4	7	2	2
3) no sound	2	4	7	1	2	1	5	5

x_1	x_1^2	x_2	x_2^2	x_3	x_3^2
7	49	5	25	2	4
4	16	5	25	4	16
6	36	3	9	7	49
8	64	4	16	1	1
6	36	4	16	2	4
6	36	7	49	1	1
2	4	2	4	5	25
9	81	2	4	5	25
$\Sigma x_1 = 48$	$\Sigma x_1^2 = 322$	$\Sigma x_2 = 32$	$\Sigma x_2^2 = 148$	$\Sigma x_3 = 27$	$\Sigma x_3^2 = 125$
$(\Sigma x_1)^2 = 2304$		$(\Sigma x_2)^2 = 1024$		$(\Sigma x_3)^2 = 729$	
$M_1 = 6$		$M_2 = 4$		$M_3 = 3.375$	

$$SS_{total} = (322 + 148 + 125) - \frac{(48 + 32 + 27)^2}{24}$$

= 595 - 477.04

SS_{total} = 117.96

$$SS_{among} = \left[\frac{2304}{8} + \frac{1024}{8} + \frac{729}{8}\right] - 477.04$$

= 507.13 - 477.04

$\underline{\mathbf{SS_{among}}}$ = 30.08

$\underline{\mathbf{SS_{within}}}$ = 117.96 - 30.08 = 87.88

Source	SS	df	MS	F
Among	30.08	2	15.04	3.59
Within	87.88	21	4.18	

*(according to the F sig/probability table with df = (2,21) F must be at least 3.4668 to reach $p < .05$, so F score **is** statistically significant)

Interpretation: Susan can conclude that her hypothesis **may** be supported. The means are as she predicted, in that the constant music group has the highest score. However, the signficant *F* only indicates that at least two means are signficantly different from one another, but she can't know which specific mean pairs significantly differ until she conducts a post-hoc analysis.

One-Way Analysis of Variance

You select four different, independent, groups of six patients and record the number of days it takes to teach them a particular task, with each group receiving one of four types of rewards: Reward 1, Reward 2, Reward 3, and Reward 4. The number of days are given in the following table.

Reward 1	Reward 2	Reward 3	Reward 4
3	6	9	12
5	7	10	13
6	9	15	15
2	7	12	18
1	11	11	15
2	6	10	13

Use the data above to conduct a one-way analysis of variance.

I. State your hypotheses

Null hypothesis: The type of reward does not make a difference in the number of days required for Down's Syndrome patients to learn a task.

Research hypothesis: The type of reward makes a difference in the number of days required for Down's syndrome patients to learn a task

II. After stating the hypotheses, always begin an analysis of variance problem by computing all required sums.

Reward 1	Reward 2	Reward 3	Reward 4	
3	6	9	12	
5	7	10	13	
6	9	15	15	
2	7	12	18	
1	11	11	15	
2	6	10	13	
$\sum X_1 = 19$	$\sum X_2 = 46$	$\sum X_3 = 67$	$\sum X_4 = 86$	$\sum\sum X = 218$
$\sum X_1^2 = 79$	$\sum X_2^2 = 372$	$\sum X_3^2 = 771$	$\sum X_4^2 = 1256$	$\sum\sum X^2 = 2478$
$n_1 = 6$	$n_2 = 6$	$n_3 = 6$	$n_4 = 6$	$N_{total} = 24$
$\bar{X}_1 = 3.167$	$\bar{X}_2 = 7.667$	$\bar{X}_3 = 11.167$	$\bar{X}_4 = 14.333$	$k = 4$

III. Compute SS_{total}

$$SS_{total} = \sum\sum X^2 - \frac{\left(\sum\sum X\right)^2}{N_{total}} = 2478 - \frac{218^2}{24} = 2478 - \frac{47524}{24}$$

$$SS_{total} = 2478 - 1980.167 = 497.833$$

IV. Compute SS_{bg}

$$SS_{bg} = \left[\frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} + \frac{\left(\sum X_3\right)^2}{n_3} + \frac{\left(\sum X_4\right)^2}{n_4}\right] - \frac{\left(\sum\sum X\right)^2}{N_{total}}$$

$$SS_{bg} = \left[\frac{19^2}{6} + \frac{46^2}{6} + \frac{67^2}{6} + \frac{86^2}{6}\right] - \frac{218^2}{24} = \left[\frac{361 + 2116 + 4489 + 7396}{6}\right] - \frac{47524}{24}$$

$$SS_{bg} = \left[\frac{14362}{6}\right] - 1980.167 = 2393.667 - 1980.167 = 413.5$$

V. Compute SS_{wg}

$$SS_{wg} = SS_{total} - SS_{bg} = 497.833 - 413.5 = 84.333$$

VI. Compute df_{total}

$df_{total} = N_{total} - 1 = 24 - 1 = 23$

VII. Compute df_{bg}

$df_{bg} = k - 1 = 4 - 1 = 3$

VIII. Compute df_{wg}

$df_{wg} = (n_1 - 1) + (n_2 - 1) + (n_3 - 1) + (n_4 - 1) = (6-1) + (6-1) + (6-1) + (6-1)$
$df_{wg} = 5 + 5 + 5 + 5 = 20$

IX. Compute MS_{bg}

$$MS_{bg} = \frac{SS_{bg}}{df_{bg}} = \frac{413.5}{3} = 137.833$$

X. Compute MS_{wg}

$$MS_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{84.333}{20} = 4.217$$

XI. Compute F

$$F = \frac{MS_{bg}}{MS_{wg}} = \frac{137.833}{4.217} = 32.685$$

XII. Find the critical value of the F ratio in Table F and determine the significance of F

A. df = (3, 20)
B. Critical Value at alpha = .05 is **3.10**
C. Because F > Critical Value we can reject the Null Hypothesis and accept the Research Hypothesis
D. The probablity of an F ratio this large happening just by chance is less than .05 ($p < .05$).

XIII. Create the Source Table

Source	SS	Df	MS	F	P
Between	413.500	3	137.833	32.685	<.05
Within	84.333	20	4.217		
Total	497.833	23			

XIV. Because F is greater than the critical value we must compute the HSD

A. $MS_{wg} = 4.217$
B. n = 6 (Hint: The number of scores in each group)
C. q = 3.96 (Hint: To enter table Q use k = 4...k is the number of groups...and $df_{wg} = 20$)
D. Compute HSD

$$HSD = q \cdot \sqrt{\frac{MS_{wg}}{n}} = 3.96 \cdot \sqrt{\frac{4.217}{6}} = 3.96 \cdot \sqrt{.703} = 3.96 \cdot .838 = 3.318$$

XV. Compare all pairs of means

$$\left|\bar{X}_1 - \bar{X}_2\right| = \left|3.167 - 7.667\right| = 4.500$$
$$\left|\bar{X}_1 - \bar{X}_3\right| = \left|3.167 - 11.167\right| = 8.000$$
$$\left|\bar{X}_1 - \bar{X}_4\right| = \left|3.167 - 14.333\right| = 11.166$$
$$\left|\bar{X}_2 - \bar{X}_3\right| = \left|7.667 - 11.167\right| = 3.500$$
$$\left|\bar{X}_2 - \bar{X}_4\right| = \left|7.667 - 14.333\right| = 6.666$$
$$\left|\bar{X}_3 - \bar{X}_4\right| = \left|11.167 - 14.333\right| = 3.166$$

XVI. Conclusions

A. Reward 2 is significantly greater than Reward 1
B. Reward 3 is significantly greater than Reward 1
C. Reward 4 is significantly greater than Reward 1
D. Reward 3 is significantly greater than Reward 2
E. Reward 4 is significantly greater than Reward 2
F. Reward 3 and Reward 4 are not significantly different from one another.

Two-Way Analysis of Variance

As a budding psychologist, you wonder whether you can teach old dogs new tricks. So you go to the pound and adopt 15 old dogs and 15 puppies. Then you attempt to teach each of the 30 dogs one of the standard dog tricks, "sit", "stay", and "roll over." Teaching only one trick to each dog, you keep a record of how many days it takes before they learn the tricks. The results of your experiment are listed in the table below. Use that data to conduct a two-way analysis of variance to determine if old dogs can learn new tricks.

		Type of Trick	
	"Sit" (Column 1)	**"Shake" (Column 2)**	**"Roll Over" (Column 3)**
Puppies (Row 1)	2 1 3 1 2	4 5 4 6 7	6 9 7 8 10
Old Dogs (Row 2)	2 5 2 4 3	9 10 11 13 7	13 12 15 17 13

Because there are 2 rows and 3 columns this makes this problem a 2 by 3 two-way analysis of variance. To begin any two-way analysis of variance we must first compute all the means and all the sums for the cells, rows, and columns. To make the table easier to read and use I have replaced the data in each cell with the appropriate sums and means. (Refer to the table at the top of the problem if you want to see the actual data rather than the sums.)

		Type of Trick		
	"Sit" (Column 1)	**"Shake" (Column 2)**	**"Roll Over" (Column 3)**	
Puppies (Row 1)	$\sum X = 9$ $\sum X^2 = 19$ $n = 5$ $\overline{X} = 1.8$	$\sum X = 26$ $\sum X^2 = 142$ $n = 5$ $\overline{X} = 5.2$	$\sum X = 40$ $\sum X^2 = 330$ $n = 5$ $\overline{X} = 8$	$\sum X_{\eta} = 75$ $\sum X_{\eta}^{\ 2} = 491$ $n_{\eta} = 15$ $\overline{X}_{\eta} = 5$

Old Dogs (Row 2)	$\sum X = 16$ $\sum X^2 = 58$ $n = 5$ $\bar{X} = 3.2$	$\sum X = 50$ $\sum X^2 = 520$ $n = 5$ $\bar{X} = 10$	$\sum X = 70$ $\sum X^2 = 996$ $n = 5$ $\bar{X} = 14$	$\sum X_{r_2} = 136$ $\sum X_{r_2}^2 = 1574$ $n_{r_2} = 15$ $\bar{X}_{r_2} = 9.667$
	$\sum X_{c_1} = 25$ $\sum X_{c_1}^2 = 77$ $n_{c_1} = 10$ $\bar{X}_{c_1} = 2.5$	$\sum X_{c_2} = 76$ $\sum X_{c_2}^2 = 662$ $n_{c_2} = 10$ $\bar{X}_{c_2} = 7.6$	$\sum X_{c_3} = 110$ $\sum X_{c_3}^2 = 1326$ $n_{c_3} = 10$ $\bar{X}_{c_3} = 11$	$\sum\sum X = 211$ $\sum\sum X^2 = 2065$ $N_{total} = 30$ $k_{cols} = 3$ $k_{rows} = 2$ $k_{cells} = 6$

I. We always begin by stating our Null and Research Hypotheses for all three F ratios

A. Rows

H_0: There is no difference between the time it takes an old dog to learn a trick and a puppy to learn a trick.

H_1: There is a difference between the time it takes an old dog to learn a trick and a puppy to learn a trick.

B. Columns

H_0: There will be no difference between the time it takes to learn the different types of tricks.

H_1: There will be a difference between the time it takes to learn the different types of tricks.

C. Interaction

H_0: There is no interaction

H_1: There is an interaction

II. The Source Table.....To complete the two-way analysis of variance we will fill out the following source table:

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Square	F	p
Rows					
Columns					
Interaction					
Within					
Total					

III. Compute Sums of Squares

A. Sums of Squares Total (SS_{total})

$$SS_{total} = \sum\sum X^2 - \frac{\left(\sum\sum X\right)^2}{N_{total}} = 2065 - \frac{(211)^2}{30} = 2065 - \frac{44521}{30} = 2065 - 1484.033 = 580.967$$

B. Sums of Squares Rows (SS_r)

$$SS_{rows} = \left[\frac{\left(\sum X_{r_1}\right)^2}{n_{r_1}} = + \frac{\left(\sum X_{r_2}\right)^2}{n_{r_2}}\right] - \frac{\left(\sum\sum X\right)^2}{N_{total}} = \left[\frac{75^2 + 136^2}{15}\right] - 1484.033$$

$$SS_{rows} = \left[\frac{5625 + 18496}{15}\right] - 1484.033 = \left[\frac{24121}{15}\right] - 1484.003 = 1608.067 - 1484.033$$

$$SS_{rows} = 124.034$$

C. Sums of Squares Columns (SS_c)

$$SS_{cols} = \left[\frac{\left(\sum X_{c_1}\right)^2}{n_{c_1}} + \frac{\left(\sum X_{c_2}\right)^2}{n_{c_2}} + \frac{\left(\sum X_{c_3}\right)^2}{n_{c_3}}\right] - \frac{\left(\sum\sum X\right)^2}{N_{total}}$$

$$SS_{cols} = \left[\frac{25^2 + 76^2 + 110^2}{10}\right] - 1484.033 = \left[\frac{625 + 5776 + 12100}{10}\right] - 1484.033$$

$$SS_{cols} = \left[\frac{18501}{10}\right] - 1484.033 = 1850.1 - 1484.033 = 366.067$$

D. Sums of Squares Within Groups (SS_{wg})

$$SS_{wg} = \sum\sum X^2 - \sum\left[\frac{\left(\sum X_{cell}\right)^2}{n_{cell}}\right] = 2065 - \left[\frac{9^2 + 26^2 + 40^2 + 16^2 + 50^2 + 70^2}{5}\right]$$

$$SS_{wg} = 2065 - \left[\frac{81 + 676 + 1600 + 256 + 2500 + 4900}{5}\right] = 2065 - \left[\frac{10013}{5}\right] = 2065 - 2002.6$$

$$SS_{wg} = 62.4$$

E. Sums of Squares Interaction (SS_{rxc})

$$SS_{rxc} = SS_{total} - \left[SS_{rows} + SS_{cols} + SS_{wg}\right] = 580.967 - \left[124.034 + 366.067 + 62.4\right]$$

$$SS_{rxc} = 580.967 - 552.501 = 28.466$$

F. Copy the Sums of Squares to the source table

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Square	F	p
Rows	124.034				
Columns	366.067				
Interaction	28.466				
Within	62.4				
Total	580.967				

IV. Compute degrees of freedom

A. Degrees of Freedom Total (df_{total})
$df_{total} = N_{total} - 1 = 30 - 1 = 29$

B. Degrees of Freedom Rows (df_r)
$df_r = n_r - 1 = 2 - 1 = 1$

C. Degrees of Freedom Columns (df_c)

$df_c = n_c - 1 = 3-1 = 2$

D. Degrees of Freedom Interaction (df_{rxc})

$df_{rxc} = df_r \cdot df_c = 1 \cdot 2 = 2$

E. Degrees of Freedom Within (df_{wg})

$df_{wg} = N_{total}$ - Number of Cells = 30-6 = 24

F. Copy Sums of Squares to the Source Table

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Square	F	p
Rows	124.034	1			
Columns	366.067	2			
Interaction	28.466	2			
Within	62.4	24			
Total	580.967	29			

V. Compute Mean Squares

A. Mean Square Rows (MS_r)

$$MS_{rows} = \frac{SS_{rows}}{df_{rows}} = \frac{124.034}{1} = 124.034$$

B. Mean Square Columns (MS_c)

$$MS_{cols} = \frac{SS_{cols}}{df_{cols}} = \frac{366.067}{2} = 183.034$$

C. Mean Square Interaction (MS_{rxc})

$$MS_{rxc} = \frac{SS_{rxc}}{df_{rxc}} = \frac{28.466}{2} = 14.233$$

D. Mean Square Within (MS_{wg})

$$MS_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{62.400}{24} = 2.6$$

E. Copy Mean Squares to the Source Table

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Square	F	p
Rows	124.034	1	124.034		
Columns	366.067	2	183.034		

Interaction	28.466	2	14.233		
Within	62.4	24	2.600		
Total	580.967	29			

VI. Compute the F ratios

A. F ratio for Rows (F_r)

$$F_{rows} = \frac{MS_{rows}}{MS_{wg}} = \frac{124.034}{2.6} = 47.705$$

B. F ratio for Columns (F_c)

$$F_{cols} = \frac{MS_{cols}}{MS_{wg}} = \frac{183.034}{2.6} = 70.398$$

C. F ratio for Interaction (F_{rxc})

$$F_{rxc} = \frac{MS_{rxc}}{MS_{wg}} = \frac{14.233}{2.6} = 5.474$$

D. Copy F ratios to the Source Table

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Square	F	p
Rows	124.034	1	124.034	47.705	
Columns	366.067	2	183.034	70.398	
Interaction	28.466	2	14.233	5.474	
Within	62.4	24	2.600		
Total	580.967	29			

VII. Conclusions and Significance of the F ratios

A. F_r

1. Critical value with (1,24) degrees of freedom = 4.26
2. F_r = 47.705 this is greater than the critical value. Therefore we reject the H_0 for Rows. The probability of this F ratio happening just by chance is <.05.
3. The puppies learned significantly faster than the old dogs.

B. F_c

1. Critical value with (2,24) degrees of freedom = 3.40
2. F_c = 70.398 this is greater than the citical value. Therefore we reject H_0 for Columns. The probability of this F ratio happening just by chance is <.05.
3. Because there are three different columns we must now compare the means from each column with each of the other column means using the HSD.
   a. Find the value of q in table Q with k, the number of groups being compared, equal to 3 and the degrees of freedom within groups equal to 24.

q = 3.53

b. Compute the HSD

$$HSD = q \cdot \sqrt{\frac{MS_{wg}}{n}} = 3.53 \cdot \sqrt{\frac{2.6}{10}} = 3.53 \cdot \sqrt{.26} = 3.53 \cdot .510 = 1.800$$

c. Compare the column means

$$\left|\bar{X}_{c_1} - \bar{X}_{c_2}\right| = |2.5 - 7.6| = 5.1$$

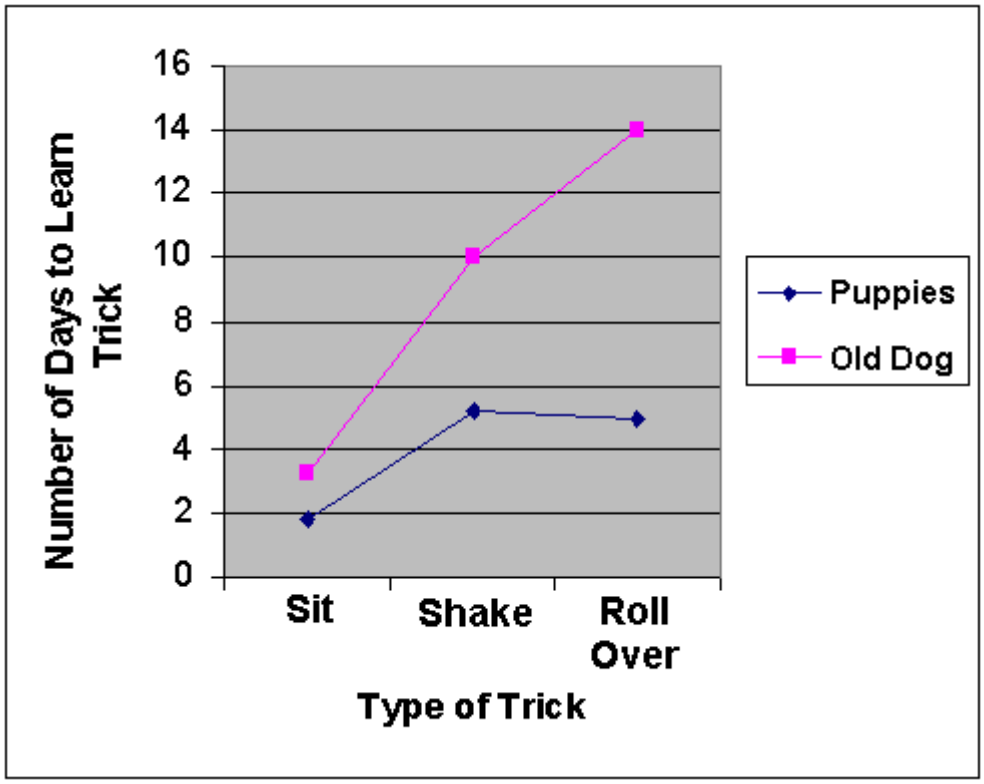
$$\left|\bar{X}_{c_1} - \bar{X}_{c_3}\right| = |2.5 - 11| = 8.5$$

$$\left|\bar{X}_{c_2} - \bar{X}_{c_3}\right| = |7.6 - 11| = 3.4$$

4. Because all the comparisons are greater than the HSD all the different types of tricks are significantly different from one another.

C. F_{rxc}

1. Critical value with (2,24) degrees of freedom = 3.40
2. F_{rxc} = 5.474 this is greater than the critical value. Therefore we reject H_0 for Interaction. The probability of this F ratio happening just by chance is <.05.
3. Because the interaction is significant we must now graph the cell means. Remember that when creating this graph use the dependent variable as the Y axis label and either rows or columns, which ever has the most groups, as the X axis label.



You can see from the graph that the puppies learn at a much faster rate overall than the old dogs. And you can also see that the sit, shake, and roll over tricks are progressively more difficult for the old dogs, but there seems to be very little difference between shake and roll over for the younger dogs.

VII. The final completed Source Table

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Square	F	P
Rows	124.034	1	124.034	47.705	<.05
Columns	366.067	2	183.034	70.398	<.05
Interaction	28.466	2	14.233	5.474	<.05
Within	62.4	24	2.600		
Total	580.967	29			

Example Problem (Two way Anova)
As a book publisher you want to determine if there are differences in the mean number of books sold between mystery books, romance books and biographies. Additionally you want to determine if the mean number of books sold varies across 3 regions of the country; east coast, west coast and midwest. Finally, you are interested in any unique interaction between book type and geographical region that would explain differences in mean book sales. You review the past sales records of 20 books for each combination of book type and geographical region. You conduct a 2-way ANOVA and the partial results are listed below. Test to see if there are any significant effects for book type (factor A), geographical region (factor B) or their interaction (AB) at alpha = .05.

Source	Degrees of Freedom	SS	MS	F-calc.	F-crit
Factor A			380		
Factor B		1264			
Interaction (AB)		564			
Error					
Total		21024			

SOLUTION
We need to complete the table and then use it to answer the question posed in the problem. The best way to start to complete the table is to fill in the parts you know for sure and then calculate the other missing parts. Use the table on the bottom of page 390 as the guide.

(1) The Degrees of Freedom. We know there are 3 levels on the categorical variable "book type" which is factor A, thus the df for factor A are (a-1) or (3-1) = 2. Factor B (geographical region) also has 3 levels and thus its df are (b-1) or (3-1) = 2. The interaction term we know has (a-1)(b-1) df and thus this is (3-1)(3-1) = (2)(2) = 4. The Error term has ab(n-1) df. a=3, b=3 and n= 20 (given in problem - remember in 2-way ANOVA the "n" is the number of observations in each cell of the factor A by factor B matrix and the number in each cell has to be equal). Thus ab(n-1) = 9(19) = 171. Finally the total df = abn-1. which is ((3)(3)(20)) - 1 = 180 - 1 = 179. This checks out as the df for factor A, B, AB, and Error all add up to 179. Additionally we know that the total df should equal the combined cell sample size minus one. And the combined cell sample size would be 9 (there are 9 cells in this factor A by Factor B matrix) times 20 = 180, and thus 180 - 1 = 179. Our numbers check out both ways. Now that we have df we can move to finishing the SS row.

(2) Sum of Squares. We need to calculate SSA from MSA and the df: (a-1). We know that MSA = SSA/(a-1) thus if we solve for SSA we get SSA = MSA(a-1) = 380(2) = 760. Next we are missing SSE. But we have SSA, SSB, SSAB and SST, and we know SST = SSA+SSB+SSAB+SSE. Again working backward we find SSE = SST-(SSA+SSB+SSAB). Plug in the numbers and SSE = 21,024 - (760+1264+564) = 21,024 - 2588 = 18,436. We now have all the SS row filled in and we can move to MS.

(3)Mean Square. MSA is given. MSB is SSB/(b-1) or 1264/2 = 632. MSAB = SSAB/(a-1)(b-1) = 564/4 = 141, and finally MSE = SSE/ab(n-1) = 18,436/171 = 107.81. Now move to F-calc.

(4) F-Calculated Values. F for factor A = MSA/MSE = 380/107.81 = 3.52. F for factor B = MSB/MSE = 632/107.81 = 5.86. F for the interaction (AB) = MSAB/MSE = 141/107.81 = 1.31. Now we need the F-critical values.

(5) F-Critical Values. For Factor A we want F(2, 171) = 3.07 (round down to 120 for 171), for factor B we look up F(2,171) = 3.07 and for the interaction we want F(4,171) = 2.45. (Notice on the critical values that the numerator df and denominator df match up with the f-calculated num./den. For example, the f-calc. for factor A is MSA/MSE or A/E thus for the critical value we want the dfA as the numerator and dfE as the denominator. Be careful to make sure your numerators and denominators match up for each pair of F-calc and F-crit.) Now lets put together the complete table:

Source	Degrees of Freedom	SS	MS	F-calc.	F-crit
Factor A	2	760	380	3.52	3.07
Factor B	2	1264	632	5.86	3.07
Interaction (AB)	4	564	141	1.31	2.45
Error	171	18,436	107.81		
Total	179	21024			

INTERPRETATION:
We have three pairs of hypotheses we must test.

(1) The best place to start is the **interaction term.** The NULL: there is no interaction, the ALT: there is an interaction. We would accept the null based on a F-calc(1.31) < F-crit(2.45) and thus conclude there is no interaction. We can thus move to interpreting the two main effects without worry of there being an interaction of the two. (If there had been an interaction then we would normally not have interpreted the main effects because we would want to first isolate the interaction (the unique combo of book type and geo. region) to identify the exact mean sales differences shown by that interaction.

(2) **Factor A.** The NULL: all book types have equal mean sales, and the ALT: not all book types have equal mean sales. In this case we would reject the null based on F-calc(3.52) > F-crit(3.07). Thus we conclude that at least one pair of book types have different mean sales.

(3) **Factor B.** The NULL: all geographical regions have equal mean sales, and the ALT: not all gepgraphical regions have equal mean sales. In this case we would reject the null based on F-calc(5.86) > F-crit(3.07). Thus we conclude that at least one pair of geographical regions have different mean sales.

And that is the conclusion of our 2-way ANOVA!