PERIYAR UNIVERSITY

NAAC 'A++' Grade - State University - NIRF Rank 56 - State Public University Rank 25

SALEM - 636 011, Tamil Nadu, India.

CENTRE FOR DISTANCE AND ONLINE EDUCATION (CDOE)

MASTER OF BUSINESS ADMINISTRATION

SEMESTER - I

CORE COURSE: QUANTITATIVE TECHNIQUES AND RESEARCH METHODS INBUSINESS

(Candidates admitted from 2024 onwards)

PERIYAR UNIVERSITY

CENTRE FOR DISTANCE AND ONLINE EDUCATION (CDOE)

M.B.A 2024 admission onwards

CORE COURSE – II

Quantitative Techniques and Research Methods in Business

Prepared by:

CENTRE FOR DISTANCE AND ONLINE EDUCATION (CDOE)

Periyar University, Salem - 636011

SYLLABUS

Quantitative Techniques and Research Methods in Business

Unit I: Introduction: Probability - Rules of probability- Probability distribution; Binomial, Poisson and Normal Distributions, their applications in Business and Industrial Problem- Baye's Theorem and its applications - Decision Making under risk and uncertainty; Maximax, Maximin, Regret Hurwitz and Laplace Criteria in Business and Decision Making - Decision tree.

Unit – II: Research Methods: Research - Definition - Research Process - Research Design – Definition- Types Of Research Design - Role of Theory in Research - Variables in Research – Objectives - Hypothesis - Types of Data; Preliminary Vs Secondary- Methods of Primary Data Collection; Survey, Observation, Experiments - Construction Of Questionnaire - Questionnaire Schedule- Validity and Reliability of Instruments - Types of Scales; Nominal, Ordinal, Interval - Types of Attitude Measurement Scales – Sampling Techniques; Probability And Non probability Techniques- Optimal Sample Size determination

Unit III: Data Preparation and Analysis: Data Preparation - Editing –Coding- Data Entry- Data Analysis- Testing Of Hypothesis Univariate and Bivariate

Analysis Parametric And Nonparametric Tests and Interpretation of Test Results- Chi-Square Test- Correlation; Karl Pearson's Vs Correlation Coefficient and Spearman's Rank Correlation- Regression Analysis - One Way and Two Way Analysis Of Variance.

Unit IV: Multivariate Statistical Analysis: Exploratory and Confirmatory Factor Analysis -Discriminant Analysis- Cluster Analysis -Conjoint Analysis - Multiple Regression- Multidimensional Scaling- Their Application In Marketing Problems - Application of Statistical Software For Data Analysis- SEM Analysis

Unit V: Report Writing and Ethics in Business Research: Research Reports- Different Types -Report Writing Format- Content of Report- Need For Executive Summary- Chapterisation -Framing the Title of the Report- Different Styles Of Referencing -Academic Vs Business Research Reports - Ethics In Research.

LIST OF CONTENTS

UNIT I

PROBABILITY DISTRIBUTIONS & DECISION THEORY

UNIT 1 - PROBABILITY DISTRIBUTIONS & DECISION THEORY

Introduction: Probability - Rules of probability- Probability distribution; Binomial, Poisson and Normal Distributions, their applications in Business and Industrial Problem- Baye's Theorem and its applications - Decision Making under risk and uncertainty; Maximax, Maximin, Regret Hurwitz and Laplace Criteria in Business and Decision Making - Decision tree.

Section	Topic
UNIT - I	
Unit Objectives	
Section 1.1	PROBABILITY DISTRIBUTIONS
1.1	Probability distribution – An Introduction
1.2	Probability – definition and relevant concepts
1.3	Rules and theorems of probability
1.4	Bayes theorem
1.5	Probability distribution
1.6	Binomial probability distribution
1.7	Poisson distribution
1.8	Normal distribution
	Let Us Sum Up
	Check Your Progress – Quiz 1
Section 1.2	DECISION THEORY
2.1	Decision Theory
2.2	Decision Making Environments
2.3	Essential Components of Decision Making
2.4	Decision making under uncertainty
2.5	Decision Tree Analysis
	Let Us Sum Up
	Check Your Progress – Quiz 2

UNIT OBJECTIVES:

The objectives of unit on Probability, Probability distributions and decision theory are as follows:

- To understand the basic concepts and theorems of probability
- To learn the meaning and characteristics of probability distributions
- To learn and solve simple problems of probability and probability distributions for business decision making
- To study the decision theory and decision-making environments and make decision using various decision-making criteria
- To learn the decision tree concept and construct decision trees

SECTION I – PROBABILITY AND PROBABILITY DISTRIBUTIONS

1.1 INTRODUCTION

The concept of Probability which originated in the seventeenth century has become one of the most fascinating and debatable subjects in the recent years. Probability concept has assumed great importance and the mathematical theory of probability has become the basis for statistical applications in both social and decision-making research. Probability has become part of our everyday lives. We livein a world in which we are unable to forecast the future with complete certainty. Our need to cope with uncertainty leads us to the study and use probability theory. Knowledge of probabilistic methods has become increasingly essential in quantitative analysis of business and economic problems.

In particular, probability theory is a basic component of the formal theory of decision-making under risk and uncertainty. A through understanding of the fundamentals of probability theory will permit a businessman to deal wit uncertaintyin business situations in such a way that he can assess systematically the risks involved in each alternative and consequently act to minimize risk. In this unit the basic concepts of probability, rules and theorems of probability, probabilitydistributions, their applications in business and industrial problems, decision making under uncertainty and risk and example problems are explained with suitable illustrations.

1.2 PROBABILITY – DEFINITION AND RELEVANT CONCEPTS

Random Experiment: A random experiment is a well-defined process of observing agiven chance phenomena through a series of trials (finite or infinite) each of whichleads to a single outcome.

Event: Event is a possible outcome of an experiment or a result of a trial or an observation.

Sample Space: A sample space is the set of all possible results or outcomes of a random experiment. Suppose, if we have thrown a dice, randomly, then the sample space for this experiment will be all possible outcomes of throwing a dice, such as;

Sample Space = {1,2,3,4,5,6}

Probability: A probability is a number that reflects the chance or likelihood that a particular event will occur in a random experiment. In simple terms, the probability of an event is a number indicating how likely that event will occur.

Probabilities can be expressed as proportions that range from 0 to 1, and they can also be expressed as percentages ranging from 0% to 100%.

A classic example of a probabilistic experiment is a fair coin toss, in which the two possible outcomes are heads or tails. In this case, the probability of flipping a head or a tail is 1/2. In an actual series of coin tosses, we may get more or less than exactly 50% heads.

Elementary Events: An elementary even or a simple even is a single possible outcome of an experiment. It is thus an event which cannot be further subdivided into a combination of other events. Example — Tossing a coin one time will give either head or tail as the outcome.

Compound Events: When two or more event s occur in connection with each other, then their simultaneous occurrence is called a compound event. Example — Tossing a coin two times successively may result the outcome of compound of the same outcomes (HH) or (HT) or (TH) or (TT). The successiveness is not limited to two times only.

Mutually Exclusive events: Two events are said to be mutually exclusive when both cannot happen simultaneously in a single trial or, in other words, the happening of one prevents the happening of the other and vice versa.

> ***Example —*** If a single coin is tossed either, head or tail is the possible outcome; but both cannot be the outcomes at the same time.
>
> Similarly, when a mother gives birth for a single child, the child will be either a male or female and both genders will not be possible.

Collectively Exhaustive Events: In the example of fair coin tossing, there are two possible outcomes: head and tail. The list of these outcomes collectively exhaustive since the result of any toss must be either head or tail. Collective exhaustive events are those which include all possible outcomes. The sum of the probabilities must be one for mutually exclusive and collectively exhaustive events.

Complementary Events: When a favourable event happens in an experiment then the unfavourable events are called the complementary events.

> ***Example*** - Let A be an event of the number of favourable outcomes in the experiment, then Ā is called the complementary event of A. It is clear that events A and Ā are mutually exclusive and collectively exhaustive.

Equally Likely Events: Events are said to be equally likely when one does not occur more often than the others.

> ***Example —*** If an unbiased coin or die is thrown, each face may be expected toobserved approximately the same number of times in the long run. Similarly, the cards of a pack of playing cards are so closely alike that we expect each card to appear equally often when a large number of draws are made with replacement.

Independent events: Independent events are those events whose occurrence is not dependent on any other event.

> ***Example*** - if we flip a coin in the air and get the outcome as Head, then again if we flip the coin but this time we get the outcome as Tail. In both cases, the occurrence of both events is independent of each other.

Dependent Events: In probability, if one event affects the outcome of the other event, is called a dependent.

> **Example** - let's say three cards are to be drawn from a pack of cards. Then the probability of getting a king is highest when the first card is drawn, while the probability of getting a king would be less when the second card is drawn.
> In the draw of the third card, this probability would be dependent upon the outcomes of the previous two cards. We can say that after drawing one card, there will be fewer cards available in the deck, therefore the probabilities after each drawn card changes.

Either or Probability: This indicates the happening of any of the events and happening of both events in an experiment.

Conditional Probability: According to this probability the measure of the probability of an event occurring given that another event has already occurred = P(A|B). In other words, among those instances where B has occurred, P(A|B) is the proportion of cases in which event A occurs.

1.3 RULES AND THEOREMS OF PROBABILITY

1.3.1 The Addition Rule for Either or Probabilities

To find P (A or B), we start by adding the individual probabilities, P(A) and P(B). But this means that the overlap between the two events A and B is counted **twice**: once by P(A) and once by P(B). To correct for this double counting, we need to subtract P (A and B) and, the probability of both events occurring. This gives us the addition rule to find P (A or B):

Additional Theorem of Probability
P (A or B) =P(A) + P(B) − P (A and B)
P (A ∪ B) =P(A) + P(B) − P (A ∩ B)

In the case of mutually exclusive events, the addition rule is P(A or B)=P(A)+P(B)

Illustration Problem 1:

At a local language school, 40% of the students are learning Tamil, 20% of the students are learning Malayalam, and 8% of the students are learning both Tamil and Malayalam. What is the probability that a randomly selected student is learning Tamil or Malayalam?

Solution:

P (Tamil or Malayalam) = P (Tamil) + P (Malayalam) − P (Tamil and Malayalam)

= 0.4 + 0.2 − 0.08 = **0.52**

Illustration Problem 2:

There are 50 students enrolled in the second year of a business degree program. During this semester, the students have to take some elective courses. 18 students decide to take an elective in marketing, 27 students decide to take an elective in finance, and 10 students decide to take an elective in both marketing and finance. What is the probability that a student takes an elective in marketing or finance?

Solution:

Probability of taking marketing = 18/50

Probability of taking finance = 27/50

Probability of taking either marketing or finance = 10/50

P (marketing or finance) =P (marketing) +P(finance) − P (marketing and finance)

=18/50 + 27/50 – 10/50 = **0.7**

1.3.2 Multiplication Rule of Probability

Multiplication for Independent Events
P (A ∩ B) = P(A) x P(B)

In the case of dependent events, the probability of any event is conditional, or depends upon the occurrence or non-occurrence of other events. From definitions of conditional probabilities, we can give the multiplication rule:

Multiplication Rule for Dependent Events

P (A ∩ B) = P(A/B) x P(B)

(or)

P (B ∩ A) = P(B/A) x P(A)

Illustration Problem 3:

Suppose a bag contains 20 balls. 10 of the balls are white, 7 of the balls are red, and 3 of the balls are blue. Suppose one ball is selected at random from the bag.

1.3.2.1 Are the events "selecting a white ball" and "selecting a red ball" mutuallyexclusive? Why?

1.3.2.2 What is the probability of selecting a white or red ball?

Solution:

a) The events "selecting a white ball" and "selecting a red ball" are mutuallyexclusive because the events cannot happen at the same time. It is not possible for the selected ball to be both white and red.

b) P (white) = 10/20

P (red) = 7/20

P (white or red) = P (white) + P (red) = 10/20 + 7/20 = **0.85**

Illustration Problem 4:

Selvam plays college soccer. He makes a goal 65% of the time he shoots. Selvam is going to attempt two goals in a row in the next game. A = the event Selvam is successful on his first attempt. P(A)=0.65. B=the event Selvam is successful on his second attempt. P(B)=0.65. Selvam tends to shoot in streaks. The probability that he makes the second goal given that he made the first goal is 0.90.

a) What is the probability that he makes both goals?

b) What is the probability that Selvam makes either the first goal or the second goal?

Solution:

a) P (A ∩ B) = P (B ∩ A)

Since P(B|A) = 0.90:

P (B ∩ A) =P(B|A) x P(A) = (0.90) x (0.65) = 0.585.

Selvam makes the first and second goals with probability 0.5850

b) P (A ∪ B) =P(A) + P(B) − P (A ∩ B) = 0.65 + 0.65 − 0.585 = 0.715.

Selvam makes either the first goal or the second goal with probability 0.715.

Illustration Problem 5:

A candidate is selected for interview of management trainees for 3 companies. For the first company there are 12 candidates, for the second there are 15 candidatesand for the third there are 10 candidates. What are the chances of his getting job at least one of the companies?

Solution:

The probability that the candidate gets the job at least at one company = 1 – probability that the candidate does not get the job in any company

Probability that the candidate does not get the job in the first company = 1 — 1/12 = 11/12

Probability that the candidate does not get the job in the second company = 1 – 1/15 = 14/15

Probability that the candidate does not get the job in the third company = 1 — 1/10 = 9/10

Since the events are independent, therefore, the probability that the candidate does not get any job in any of the three companies

= (11/12) x (14/15) x (9/10) = 231/100 = 0.77

Hence, the probability that the candidate gets the job at least at one company = 1- 0.77 = **0.23**

1.4 BAYES THEOREM:

Bayes theorem, in simple words, determines the conditional probability of event A given that event B has already occurred based on the following:

- Probability of B given A
- Probability of A
- Probability of B

Bayes Law is a method to determine the probability of an event based on the occurrences of prior events. It is used to calculate conditional probability. Bayes theorem calculates the probability based on the hypothesis.

Bayes rule states that the conditional probability of an event A, given the occurrence of another event B, is equal to the product of the likelihood of B, given A and the probability of A divided by the probability of B. It is given as Bayee's theorem formula

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

Here,

P(A) = how likely A happens (Prior knowledge) - The probability of a hypothesis is true before any evidence is present.

P(B) = how likely B happens (Marginalization) - The probability of observing the evidence.

P(A/B) = how likely A happens given that B has happened (Posterior) - The probability of a hypothesis is true given the evidence.

P(B/A) = how likely B happens given that A has happened (Likelihood) - The probability of seeing the evidence if the hypothesis is true.

The statement of Bayes Theorem is as follows: Let E1, E2, E3, ..., En be a set of events associated with a sample space S, where all events E1, E2, E3, ..., En have non-zero probability of occurrence and they form a partition of S. Let A be any event which occurs with E1 or E2 or E3 ... or En, then according to Bayes Theorem,

$$P(Ei/A) = \frac{P(Ei) \times P(A/Ei)}{\sum_{k=1}^{n} P(Ek) \times P(A/Ek)}, \quad i=1, 2, 3, ..., n$$

- Here $Ei \cap Ej = \varphi$, where $i \neq j$. (i.e) They are mutually exhaustive events
- The union of all the events of the partition, should give the sample space.
- $0 \leq P(Ei) \leq 1$

1.4.1 BAYES THEOREM APPLICATIONS

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. Bayesian inference has found application in various activities, including medicine, science, philosophy, engineering, sports, law, etc. For example, we can use Bayes' theorem to define the accuracy of medical test results by considering how likely any given person is to have a disease and the test's overall accuracy. Bayes' theorem relies on consolidating prior probability distributions to generate posterior probabilities. In Bayesian statistical inference, prior probability is the probability of an event before new data is collected.

Illustration problem 6:

A study of driving violations and drivers who use cell phones produced the following data:

	Driving violations	No driving violations	Total
Cell Phone user	25	280	305

Not a cell phone user	45	405	450
Total	70	685	755

a) What is the probability that a randomly selected person is a cell phone user or hasno speeding violations in the last year?

b) What is the probability that a randomly selected person had a speeding violation in the last year or does not use a cell phone?

Solution:

P (cell phone) = 305/755

P(violations) = 70/755

P (no violations) = 685/755

P (no cell phone) = 450/755

P (cell phone and no violations) = 280/755

P (violations and no cell phone) = 45/755

a) P (cell phone or no violations) = P (cell phone) + P (no violations) – P (cell phoneand no violations)

= 305/755 + 685/755 − 280/755 = 710/755 = **0.94**

b) P (violations or no cell phone) =P(violations) + P (no cell phone) — P (violationsand no cell phone)

= 70/755 + 450/755 − 45/755 = 475/755 = **0.63**

Illustration Problem 7:

In a post office, three clerks are assigned to process incoming mail. The first clerk, X, processes 40%, the second clerk Y, processes 35% and the third clerk Z, processes

25% of the mail. The first clerk has an error rate of 0.04, the second has an error rate of 0.06 and the third has an error rate of 0.03. A mail selected at random from a day's output is found to have an error. The post Master wishes to know the probability that the mail was processed by the first, second or third clerk, respectively.

Solution:

Let A denotes the event that a mail containing an error is selected at random and X, Y and Z be the events that the mail was process by the first, second and third clerk respectively.

Using our usual notations and Bayee's theorem we have to compute the conditional probabilities.

From the information

P (X) = 0.40, P (Y) = 0.35, and P (Z) = 0.25

These probabilities, which can be obtained without additional information are called prior probabilities

From the information, the conditional probabilities observing a record with an error, given that it was processed by one of the three clerks are:

P(A/X) = 0.04, P (A/Y) = 0.06 and P (A/Z) = 0.03

From these probabilities we can calculate joint probabilities:

P (A/X) = 0.04, P (A/Y) = 0.06 and P (A/Z) = 0.03

From these probabilities, we can calculate joint probabilities:

P (A ∩ X) = P (A/X) x P (X) = 0.04 x 0.40 =0.016

P (A ∩ Y) = P (A/Y) x P (Y) = 0.06 x 0.35 = 0.021

P (A ∩ Z) = P (A/Z) x P (Z) = 0.03 x 0.25 =

0.0075

Use Bayee's formula to obtain the desired probabilities

$$P(X/A) = \frac{P(A \cap X)}{P(A \cap X) + P(A \cap Y) + P(A \cap Z)} = \frac{0.016}{(0.016 + 0.021 + 0.0075)}$$

= 0.36

$$P(Y/A) = \frac{P(A \cap Y)}{P(A \cap X) + P(A \cap Y) + P(A \cap Z)} = \frac{0.021}{(0.016 + 0.021 + 0.0075)}$$

= 0.47

$$P(Z/A) = \frac{P(A \cap Z)}{P(A \cap X) + P(A \cap Y) + P(A \cap Z)} = \frac{0.0075}{(0.016 + 0.021 + 0.0075)}$$

= 0.17

These probabilities are called posterior probabilities because they were calculated after it was known that the mail was on containing an error.

Illustration Problem 8:

In a University 30% of the students doing a course in statistics use the book authored by A1, 45% use the book authored by A2, and 25% use the book authored by A3. The proportion of students who learnt about each of these books through theirteachers are: A1 = 0.5, A2 = 0.3 and A3 = 0.2. One of the students selected at random revealed that he learnt about the book he is using through his teachers. Find the probabilities that the book used is authored by A1, A2 and A3 respectively.

Solution:

Let A be the event that a student learnt about the book he is using through histeachers.

Let B1, B2 and B3 be the events that the book is authored by M, N and Q respectively.

Thus, we have P (B1) = 0.30, P (B2) = 0.45 and P (B3) = 0.20.

Also given are the conditional probabilities P (A/B1) = 0.50, P (A/B2) = 0.30 and P (A/B3) = 0.20.

From these, we may compute the following joint probabilities:

P (A ∩ B1) = P (B1) x P (A/B1) = (0.30) x (0.50) =

0.150

P (A ∩ B2) = P (B2) x P (A/B2) = (0.45) x (0.30) = 0.135
P (A ∩ B3) = P (B3) x P (A/B3) = (0.25) x (0.20) = 0.050P (A) = 0.335

Then the desired probabilities are:

$$P(B1/A) = \frac{P(B1) \times P\,A/B1)}{P(A)} = \frac{0.15}{0.355} = 0.45$$

$$P(B2/A) = \frac{P(B2) \times P\,A/B2)}{P(A)} = \frac{0.135}{0.355} = 0.40$$

$$P(B3/A) = \frac{P(B3) \times P\,A/B3)}{P(A)} = \frac{0.05}{0.355} = 0.15$$

1.5 PROBABILITY DISTRIBUTIONS

A probability distribution is an idealized frequency distribution. A frequency distribution describes a specific sample or dataset. It's the number of times each possible value of a variable occurs in the dataset.

The number of times a value occurs in a sample is determined by its probability of occurrence. Probability is a number between 0 and 1 that says how likely something is to occur:

- 0 means it's impossible
- 1 means it's certain

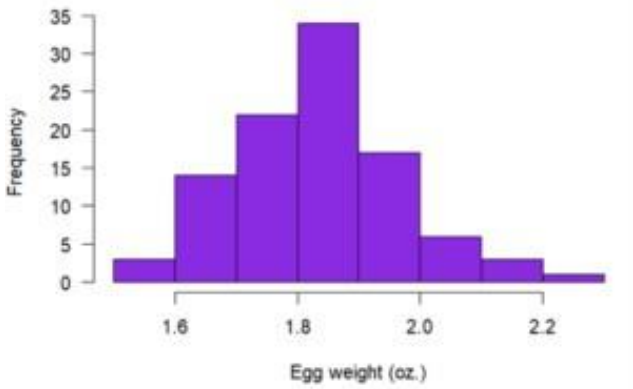
The higher the probability of a value, the higher its frequency in a sample. More specifically, the probability of a value is its relative frequency in an infinitely large sample. Infinitely large samples are impossible in real life, so probability distributions are theoretical. They're idealized versions of frequency distributions that aim to describe the population the sample was drawn from.

Probability distributions are used to describe the populations of real-life variables, like coin tosses or the weight of chicken eggs. They're also used in hypothesis testing to determine p values.

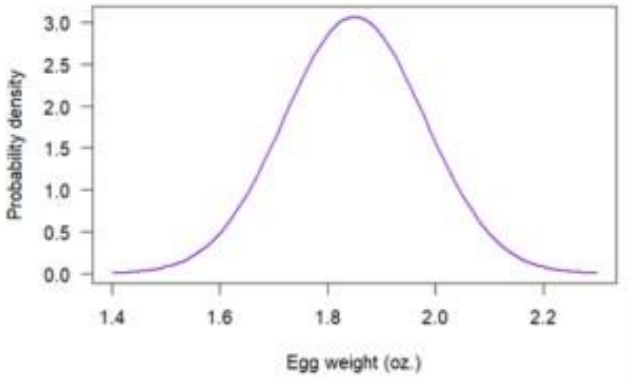
Example: Probability distributions are idealized frequency distributions

Imagine that an egg farmer wants to know the probability of an egg from his farm being a certain size. The farmer weighs 100 random eggs and describes their frequency distribution using a histogram:

He can get a rough idea of the probability of different egg sizes directly from this frequency distribution. For example, he can see that there's a high probability of an egg being around 1.9 oz., and there's a low probability of an egg being bigger than 2.1 oz. Suppose the farmer wants more precise probability estimates. One option is to improve his estimates by weighing many more eggs.



A better option is to recognize that egg size appears to follow a common probability distribution called a normal distribution. The farmer can make an idealized version of the egg weight distribution by assuming the weights are normally distributed:



Since normal distributions are well understood by statisticians, the farmer can calculate precise probability estimates, even with a relatively small sample size.

Variables that follow a probability distribution are called random variables. There's special notation you can use to say that a random variable follows a specific distribution:

- Random variables are usually denoted by X.
- The ~ (tilde) symbol means "follows the distribution."

- The distribution is denoted by a capital letter (usually the first letter of the distribution's name), followed by brackets that contain the distribution's parameters.

For example, the following notation means "the random variable X follows a normal distribution with a mean of μ and a variance of σ2."

$$X \sim N(\mu, \sigma^2)$$

There are two types of probability distributions:

- Discrete probability distributions
- Continuous probability distributions

Example: Probability table

A robot greets people using a random greeting. The probability distribution of the greetings is described by the following probability table:

Greeting	Probability
"Greetings, human!"	0.6
"Hi!"	0.1
"Good day"	0.2
"How do you do!"	0.1

Notice that all the probabilities are greater than zero and that they sum to one.

1.6 BINOMIAL PROBABILITY DISTRIBUTION

In binomial probability distribution, the number of 'Success' in a sequence of n experiments, where each time a question is asked for yes-no, then the Boolean-valued outcome is represented either with success/yes/true/one (probability p) or failure/no/false/zero (probability q = 1 − p). A single success/failure test is also called a Bernoulli trial or Bernoulli experiment, and a series of outcomes is called a **Bernoulli process**. For n = 1, i.e. a single experiment, the binomial distribution is a **Bernoulli distribution**.

1.6.1 Applications of Binomial Distribution:

The Binomial Distribution is used for

- Finding the quantity of raw and used materials while making a product.
- Taking a survey of positive and negative reviews from the public for any specific product or place.
- By using the YES/ NO survey, we can check whether the number of personsviews the particular channel.
- To find the number of male and female employees in an organisation.
- The number of votes collected by a candidate in an election is counted based on 0 or 1 probability.

1.6.2 Characteristics of Binomial Distribution

The properties or characteristics of the binomial distribution are:

- There are two possible outcomes: true or false, success or failure, yes or no.
- There is ‘n’ number of independent trials or a fixed number of n times repeated trials.
- The probability of success or failure remains the same for each trial.
- Only the number of success is calculated out of n independent trials.
- Every trial is an independent trial, which means the outcome of one trial does not affect the outcome of another trial.

1.6.3 Formula of Binomial Distribution

The binomial distribution formula is for any random variable X, given by;

$$P(x:n,p) = nCx\ px\ (1-p)\ n-x$$

(or)

$$P(x:n,p) = {}^{n}C_{x}\, p^{x} (q)^{n-x}$$

Where,

n = the number of experiments

x = 0, 1, 2, 3, 4, . . .

p = Probability of Success in a single experiment

q = Probability of Failure in a single experiment = 1 – p

The binomial distribution formula can also be written in the form of n-Bernoulli trials, where ${}^{n}C_{x}$ = n!/x!(n-x)!. Hence,

$P(x: n, p) = n!/[x!(n-x)!].p^{x}.(q)^{n-x}$

Illustration Problem 9:

If a coin is tossed 5 times, find the probability of:

a) Exactly 2 heads

b) At least 4 heads

c) Getting at most 2 heads

Solution:

a) The repeated tossing of the coin is an example of a Bernoulli trial. According to the problem:

Number of trials: n=5

Probability of head: p= 1/2 and hence the probability of tail, q =1/2

For exactly two heads: x=2

$P(x=2) = {}^{5}C2\ p^{2}\ q^{5-2} = 5!\ /\ 2!\ 3! \times (½)^{2} \times (½)^{3}$

P (x=2) = 5/16

b) For at least four heads,

$x \geq 4$, $P(x \geq 4) = P(x = 4) + P(x=5)$

Hence,

$P(x = 4) = {}^{5}C4\ p^{4}\ q^{5-4} = 5!/4!\ 1! \times (½)^{4} \times (½)^{1} = 5/32$

P (x = 5) = $^5C5\ p^5\ q^{5-5} = (½)^5 = 1/32$

Therefore, **P (x ≥ 4) = 5/32 + 1/32 = 6/32 = 3/16**

c) Getting at most 2 heads

P (at most 2 heads) = P (X ≤ 2) = P (X = 0) + P (X = 1) + P (X =

2)P (X = 0) = $(½)^5 = 1/32$

P(X=1) = $^5C_1\ (½)^{5.} = 5/32$

P(x=2) = $^5C2\ p^2\ q^{5-2} = 5!\ /\ 2!\ 3! \times (½)^2 \times (½)^3 = 5/16$

Therefore, **P (X ≤ 2) = 1/32 + 5/32 + 5/16 = 1/2**

Illustration Problem 10:

A fair coin is tossed 10 times, what are the probability of getting a) exactly 6 heads and b) at least six heads.

Solution:

Let x denote the number of heads in an experiment.

Here, the number of times the coin tossed is 10. Hence, n=10.

The probability of getting head, p ½

The probability of getting a tail, q = 1-p = 1-(½) = ½.

The binomial distribution is given by the formula:

P (X= x) = $^nC_x p^x q^{n-x}$, where = 0, 1, 2, 3,

. . . Therefore, **P (X = x) = $^{10}C_x(½)^x(½)^{10-x}$**

a) The probability of getting exactly 6 heads is:

P (X = 6) = $^{10}C_6(½)^6(½)^{10-6}$

P (X = 6) = $^{10}C_6(½)^{10}$

P (X = 6) = 105/512.

Hence, the probability of getting exactly 6 heads is 105/512.

b) The probability of getting at least 6 heads is P (X ≥ 6)

P (X ≥ 6) = P(X=6) + P(X=7) + P (X= 8) + P (X = 9) + P(X=10)

P (X ≥ 6) = $^{10}C_6(½)^{10} + {}^{10}C_7(½)^{10} + {}^{10}C_8(½)^{10} + {}^{10}C_9(½)^{10} + {}^{10}C_{10}(½)^{10}$

P (X ≥ 6) = 193/512.

Hence, the probability of getting at least 6 heads is 193/512.

1.7 POISSON DISTRIBUTION

The Poisson distribution is a discrete probability function that means the variable can only take specific values in a given list of numbers, probably infinite. A Poisson time. In other words, we can define it as the probability distribution that results from the Poisson experiment. A Poisson experiment is a statistical experiment that classifies the experiment into two categories, such as success or failure. Poisson distribution is a limiting process of the binomial distribution.

A Poisson random variable “x” defines the number of successes in the experiment. This distribution occurs when there are events that do not occur as the outcomes of a definite number of outcomes. Poisson distribution is used under certain conditions. They are:

- The number of trials “n” tends to infinity
- Probability of success “p” tends to zero
- np = 1 is finite

1.7.1 Characteristics of Poisson Distribution

The characteristics of the Poisson distribution are as follows:

- The occurrence of the events is independent. That is the occurrence of an interval of space or time has no effect on the probability of a second occurrence of the event in the same, or any other interval.

- Theoretically, an infinite number of occurrences of the event must be possible in the interval.
- The probability of single occurrence of the event in a given interval is proportional to length of the interval
- In any infinitesimal (extremely small) portion of interval, the probability of two or more occurrences of the event is negligible.

Difference between Binomial and Poisson Distribution:

Poisson distribution differs from the binomial distribution in two important aspects:

- ❖ Rather than consisting of discrete trials, the distribution operates continuously over some given amount of time, distance, area, etc…
- ❖ Rather than producing a sequence of successes and failures, the distribution produces successes, which occur at random points in the specified time, distance, area. These successes are commonly referred to 'occurrences'.

1.7.2 Applications of Poisson Distribution

There are various applications of the Poisson distribution. The random variables that follow a Poisson distribution are as follows:

- To count the number of defects of a finished product
- To count the number of deaths in a country by any disease or natural calamity
- To count the number of infected plants in the field
- To count the number of bacteria in the organisms or the radioactive decay in atoms
- To calculate the waiting time between the events.

1.7.3 Formula for Poisson Distribution

The formula for the Poisson distribution function is given by:

$$f(x) = (e^{-\lambda} \lambda^x)/x!$$

Where,

e is the base of the logarithm

x is a Poisson random variable

λ is an average rate of value

Illustration Problem 11:

Telephone calls arrive at an exchange according to the Poisson process at a rate λ= 2/min. Calculate the probability that exactly two calls will be received during each of the first 5 minutes of the hour.

Solution:

Assume that "N" be the number of calls received during a 1-minute period.

Therefore,

$P(N= 2) = (e^{-2}. 2^2)/2!$

$P(N=2) = 2e^{-2}$.

Now, "M" be the number of minutes among 5 minutes considered, during which exactly 2 calls will be received. Thus "M" follows a binomial distribution with parameters n=5 and $p= 2e^{-2}$.

$P(M=5) = 32 \times e^{-10}$

P (M =5) = 0.00145, where "e" is a constant, which is approximately equal to 2.718.

Illustration Problem 12:

If 3% of electronic units manufactured by a company are defective. Find the probability that in a sample of 200 units, less than 2 bulbs are defective.

Solution:

The probability of defective units p = 3/100 = 0.03

Give n = 200.

We observe that p is small and n is large here. Thus, it is a Poisson distribution.

Mean λ= np = 200 × 0.03 = 6

P (X= x) is given by the Poisson Distribution Formula as $(e^{-\lambda} \lambda^x)/x!$

P (X < 2) = P(X = 0) + P(X= 1)

= $(e^{-6} 6^0)/0! + (e^{--6}6^1)/1!$

= $e^{-6} + e^{--6} \times 6$

= 0.00247 + 0.0148

P (X < 2) = 0.01727

Answer: The probability that less than 2 bulbs are defective is 0.01727

Illustration problem 13:

If in an industry there is a chance that 5% of the employees will suffer by corona. What is the probability that in a group of 20 employees, more than 3 employees will suffer from the corona?

Solution:

Here we have, n = 20, p = 0.05, λ = np = 1

X = Number of employees who will suffer corona

Using Poisson's Distribution

P (X > 3) = 1 - [P (X = 0) + P (X = 1) + P (X = 2) + P (X =

3)]P(X = 0) = = 1/e

P (X = 1) = = 1/e

P (X = 2) = =1/2e

P (X = 3) = =1/6e

P (X > 3) = 1 — [1/e + 1/e + 1/2e + 1/6e]

= 1 – [8/3e] = 0.018988

Illustration Problem 14:

If 1% of the total screws made by a factory are defective. Find the probability that less than 3 screws are defective in a sample of 100 screws.

Solution:

Here we have, n = 100, p = 0.01, λ = np = 1

X = Number of defective screws

Using Poisson's Distribution

P (X < 3) = P (X = 0) + P (X = 1) + P (X = 2)

P (X = 0) = = 1/e

P (X = 1) = =1/e

P (X = 2) = =1/2e

P (X < 3) = 1/e + 1/e +1/2e

= 2.5/e = 0.919698

Illustration Problem 15:

A factory produces nails and packs them in boxes of 200. If the probability that a nail is substandard is 0.006, find the probability that a box selected at random contains at most two nails that are substandard. Use the Poisson distribution formula.

Solution:

If X is the number of substandard nails in a box of 200, then X ~ B (200,0.006)

Since n is large and p is small, the Poisson approximation can be used. The appropriate value of λ is given by

λ= np = 200 × 0.006 = 1.2

So, X ~ Po (1.2) and

$P(X \leq 2) = e^{-1.2} + e^{-1.2} \times 1.2 + (e^{-1.2} \times 1.2^2)/ 2!$

$= 2.92\, e^{-1.2} = 0.8795$

1.8 NORMAL DISTRIBUTION

1.8.1 Definition

The Normal Distribution is defined by the probability density function for a continuous random variable in a system. Let us say, f(x) is the probability density function and X is the random variable. Hence, it defines a function which is integrated between the range or interval (x to x + dx), giving the probability of random variable X, by considering the values between x and x+dx.

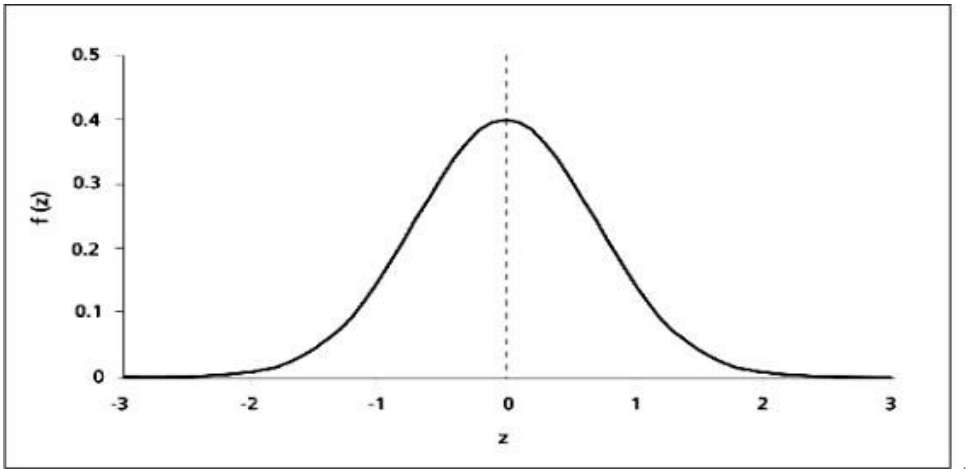
$f(x) \geq 0 \; \forall \; x \in (-\infty, +\infty)$

and $\int_{-\infty}^{+\infty} f(x) = 1$

1.8.2 Normal Distribution Curve

The random variables following the normal distribution are those whose values can find any unknown value in a given range. For example, finding the height of the students in the school. Here, the distribution can consider any value, but it will be bounded in the range say, 0 to 6ft. This limitation is forced physically in our query.

Whereas, the normal distribution doesn't even bother about the range. The range can also extend to $-\infty$ to $+\infty$ and still we can find a smooth curve. These random variables are called Continuous Variables, and the Normal Distribution then provides here probability of the value lying in a particular range for a given experiment. Also, use the normal distribution calculator to find the probability density function by just providing the mean and standard deviation value.



1.8.3 Normal Distribution Properties

Some of the important properties of the normal distribution are listed below:

- In a normal distribution, the mean, median and mode are equal. (i.e., Mean = Median= Mode).
- The total area under the curve should be equal to 1.
- The normally distributed curve should be symmetric at the centre.
- There should be exactly half of the values are to the right of the centre andexactly half of the values are to the left of the centre.
- The normal distribution should be defined by the mean and standard deviation.
- The normal distribution curve must have only one peak. (i.e., Unimodal)
- The curve approaches the x-axis, but it never touches, and it extends farther away from the mean.

1.8.4 Normal Distribution Standard Deviation

Generally, the normal distribution has any positive standard deviation. We know that the mean helps to determine the line of symmetry of a graph, whereas the standard deviation helps to know how far the data are spread out. If the standard deviation is smaller, the data are somewhat close to each other and the graph becomes narrower. If the standard deviation is larger, the data are dispersed more, and the graph

becomes wider. The standard deviations are used to subdivide the area under the normal curve. Each subdivided section defines the percentage of data, which falls into the specific region of a graph.

Using the standard deviation in the normal distribution, the Empirical Rule states that,

- Approximately 68% of the data falls within one standard deviation of themean. (i.e., Between Mean- one Standard Deviation and Mean + one standard deviation)
- Approximately 95% of the data falls within two standard deviations of the mean. (i.e., Between Mean- two Standard Deviation and Mean + two standard deviations)
- Approximately 99.7% of the data fall within three standard deviations of the mean. (i.e., Between Mean- three Standard Deviation and Mean + three standard deviations)

1.8.5 Central limit theorem

The central limit theorem is the basis for how normal distributions work in statistics.

In research, to get a good idea of a population mean, ideally you'd collect data from multiple random samples within the population. A sampling distribution of themean is the distribution of the means of these different samples.

The central limit theorem shows the following:

- Law of Large Numbers: As you increase sample size (or the number of samples), then the sample mean will approach the population mean.
- With multiple large samples, the sampling distribution of the mean is normally distributed, even if your original variable is not normally distributed.

Parametric statistical tests typically assume that samples come from normally distributed populations, but the central limit theorem means that this assumption isn't necessary to meet when you have a large enough sample.

You can use parametric tests for large samples from populations with any kind of distribution as long as other important assumptions are met. A sample size of 30 or more is generally considered large.

For small samples, the assumption of normality is important because the sampling distribution of the mean isn't known. For accurate results, you have to be sure that the population is normally distributed before you can use parametric tests with small samples.

Illustration Problem 16:

The average daily sales of 500 branch offices was Rs.150 thousand and the standard deviation was Rs.15 thousand. Assuming the distribution to be normal, indicate how many branches have sales between:

a) Rs.120 thousand and Rs.145 thousand
b) Rs.140 thousand and Rs. 165 thousand

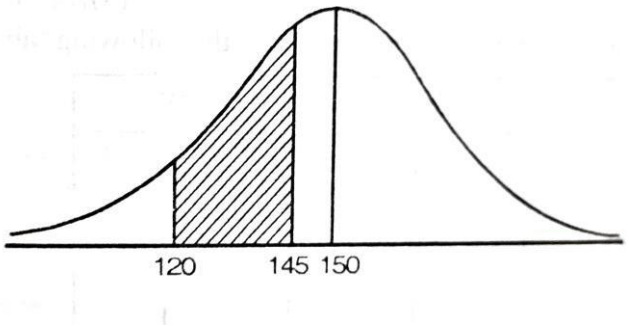
Solution:

a) Standard normal variate corresponding to 120 is

$z = x - \mu/\sigma = 120 — 150/15 = - 2$ and

corresponding to 145, the standard normal variate is

$z = x - \mu/\sigma = 145 — 150/15 = - 0.33$



From the table we find the areas corresponding to the values of z are 0.4772

and 0.1293

Therefore, the desired area between Rs.120 and Rs.145

= 0.4772 – 0.1293 = 0.3479

Hence the expected number of branches having sales between Rs.120 thousand and Rs.145 thousand

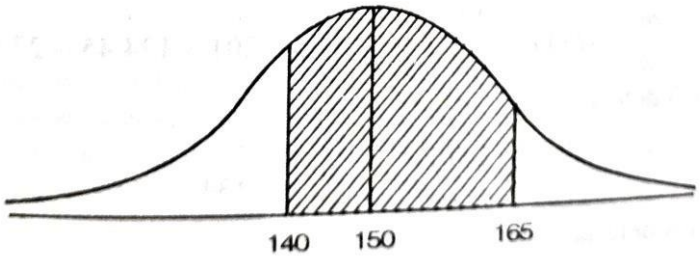
= 0.3479 x 500 = 173.95 = 174

b) Standard normal variate corresponding to 140

isz = 140 — 150/15 = -0.67

and corresponding to 164, the standard normal variate is

z = 165 — 150/15 = 1



From the table, the areas corresponding to the z values are 0.2486 and 0.3413

Therefore, the area is 0.2486 + 0.3413 = 0.5899

Hence, the expected number of branches having sales between Rs.140 thousand and Rs.165 thousand are

0.5899 x 500 = 294.5 = 295

Illustration Problem 17:

X is a normally distributed variable with mean μ = 30 and standard deviation σ = 4.

Find the probabilities

a) P (X < 40)

b) P (X > 21)

c) P (30 < X < 35)

Solution:

a) For x = 40

z = (40 - 30) / 4 = 2.5

Hence P (x < 40) = P (z < 2.5)

= [area to the left of 2.5] = **0.9938**

b) For x = 21, z = (21 - 30) / 4 = -2.25

Hence P(x > 21) = P(z > -2.25)

= [total area] - [area to the left of -2.25]

= 1 - 0.0122 = **0.9878**

c) For x = 30, z = (30 - 30) / 4 = 0 and for x = 35,

z = (35 - 30) / 4 = 1.25

Hence P (30 < x < 35) = P (0 < z < 1.25)

= [area to the left of z = 1.25] - [area to the left of 0]

= 0.8944 - 0.5 = **0.3944**

Illustration Problem 18:

For certain types of computers, the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers. What is the probability that the length of time will be between 50 and 70 hours?

Solution:

Let x be the random variable that represents the length of time. It has a mean of 50 and a standard deviation of 15.

We have to find the probability that x is between 50 and 70

or P (50< x < 70)

For x = 50, z = (50 - 50) / 15 = 0

For x = 70, z = (70 - 50) / 15 = 1.33

P (50< x < 70) = P (0< z < 1.33)

= [area to the left of z = 1.33] - [area to the left of z = 0]

= 0.9082 - 0.5 = 0.4082

The probability that John's computer has a length of time between 50 and 70 hours is equal to 0.4082.

Illustration Problem 19:

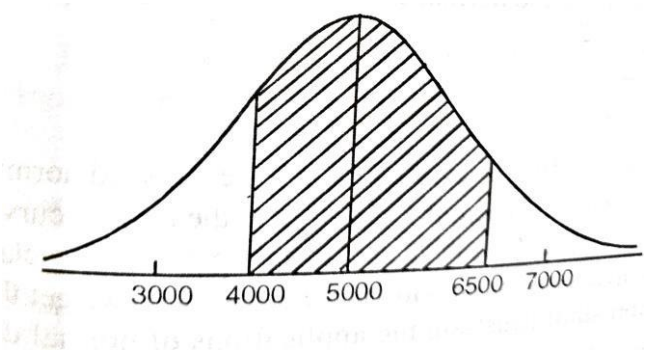
How many workers have a salary between Rs.4000 and Rs.6500, if the arithmetic mean is Rs.5000, standard deviation is Rs.1000 and number of workers is 15000. Assume the salary of the workers to follow the normal distribution

Solution:

The workers salary has a mean of 5000 and a standard deviation of 1000.

z1 = 4000 – 5000 / 1000 = - 1 and z2 = 6500 – 5000 / 1000 =

1.5



From the table, we find that 34.13% of workers fall between Rs.4000 and Rs.5000 and 43.32% fall between Rs.5000 and Rs.6500.

Therefore, 34.13 + 43.32 = 77.45 % of workers have a salary between Rs.4000 and Rs.6500.

Number of workers getting salary between Rs.4000 and Rs.6500 is given by

= 0.7745 x 15000 = 11618

Let's Sum Up

Dear learners, in this section on Probability and Probability Distributions, we have learnt about the basic concepts of Probability with their meaning and definitions with suitable examples. You have also provided the additional, multiplication theorems of probability and Bayes' Theorem and their applications in solving problems to real life and business decision making. Few solved problems have been illustrated for better understanding of the theorems. Further you have learnt about the Probability Distributions with their characteristics. The Binomial Distribution, Poisson Distribution and Normal Distribution have been taught with their formula and suitable illustrations. The problems given in the illustrations were explained with their procedure.

Check Your Progress -Quiz 1

1. **When two events cannot happen simultaneously in a single trial they are called _____**

   a. Compound events

   b. Mutually exclusive events

c. Equally likely events

d. Elementary events

2. In Bayes' theorem, the revised probabilities are referred as _____

a. Prior probabilities

b. Continuous probabilities

c. Discrete probabilities

d. Posterior probabilities

3. It is suitable to use Binomial Distribution only for _____

a. Large values of 'n'

b. Fractional values of 'n'

c. Small values of 'n'

d. Any value of 'n'

4. Poisson distribution is applied for _____

a. Continuous Random Variable

b. Discrete Random Variable

c. Irregular Random Variable

d. Uncertain Random Variable

5. Normal Distribution is symmetric is about

a. Variance

b. Mean

c. Standard deviation

d. Covariance

SECTION II - DECISION THEORY

2.1 DECISION THEORY

Decision theory deals with methods for determining the optimal course of action when a number of alternatives are available and their consequences cannot be forecast with certainty.

It is difficult to imagine a situation which does not involve such decision problems, but we shall restrict ourselves primarily to problems occurring in business, with consequences that can be described in dollars of profit or revenue, cost or loss. For these problems, it may be reasonable to consider as the best alternative that which results in the highest profit or revenue, or lowest cost or loss, on the average, in the long run. This criterion of optimality is not without shortcomings, but it should serve as a useful guide to action in repetitive situations where the consequences are not critical.

2.2 DECISION MAKING ENVIRONMENTS

The quality of the decisions made in an organization will dictate the success or failure of the said business. So, all the available information and alternatives must be studied before arriving at an important decision. The process of decision making will help a great deal.

Another factor that affects these decisions is the environment in which they are taken. There are a few different types of environments in which these decisions are made. And the type of decision-making environment has an impact on the way the decision is taken. Broadly there are three basic types of decision-making environment. Let us take a brief look at each of them.

2.2.1 DECISION MAKING UNDER CERTAINTY

Such type of environment is very sure and certain by its nature. This means that all the information is available and at hand. Such data is also easy to attain and not very expensive to gather. So the manager has all the information he may need to make an informed and well thought out decision. All the alternatives and their outcomes can also be analyzed and then the manager chooses the best alternative.

Another way to ensure an environment of certainty is for the manager to create a closed system. This means he will choose to only focus on some of the alternatives. He will get all the available information with respect to such alternatives he is analyzing. He will ignore the other factors for which the information is not available. Such factors become irrelevant to him altogether.

2.2.2 DECISION MAKING UNDER UNCERTAINTY

In the decision-making environment of uncertainty, the information available to the manager is incomplete, insufficient and often unreliable. In an uncertain environment, everything is in a state of flux. Several external and random forces mean that the environment is most unpredictable.

In these times of chaos, all the variables change fast. But the manager has to make sense of this mayhem to the best of his ability. He must create some order, obtain some reliable data and make the best decision as per his judgment.

2.2.3 DECISION MAKING UNDER RISK

Under the condition of risk, there is the possibility of more than one event taking place. Which means the manager has to first ascertain the possibility and probability of the occurrence or non-occurrence of the event. The manager will generally rely on past experiences to make this deduction.

In this scenario too, the manager has some information available to him. But the availability and the reliability of the information is not guaranteed. He has to chart a few alternative courses of actions from the data he has.

2.3 ESSENTIAL COMPONENT OF DECISION MAKING

Decision Alternatives: There is finite number of decision alternatives available to the decision-maker at each point in time when a decision is made. The number and type of such alternatives may depend on the previous decisions made and their outcomes. Decision alternatives may be described numerically, such as stocking 100 units of a particular item, or non-numerically, such as conducting a market survey to know the likely demand of an item.

States of Nature: A state of nature is an event or scenario that is not under the control of decision makers. For instance, it may be the state of economy (e.g.inflation), a weather condition, a political development etc.

The states of nature may be identified through Scenario Analysis where a section of people is interviewed — stakeholders, long-time managers, etc., to understand states of nature that may have serious impact on a decision.

The states of nature are mutually exclusive and collectively exhaustive with respect to any decision problem. The states of nature may be described numerically such as, demand of 100 units of an item or non-numerically such as employees strike, etc.

Payoff: It is a numerical value (outcome) obtained due to the application of each possible combination of decision alternatives and states of nature. The payoff values are always conditional values because of unknown states of nature. Payoffs resulting from each possible combination of decision alternatives and states of natures are displayed in a matrix called payoff matrix.

2.4 DECISION MAKING UNDER UNCERTAINTY

The following criteria of decision-making under uncertainty are given below:

a) Optimism (Maximax or Minimin) criterion
b) Pessimism (Maximin or Minimax) criterion
c) Equally Likely (Laplace) criterion
d) Coefficient of optimism (Hurwicz) criterion
e) Minimax Regret (salvage) criterion

2.4.1 Optimism (Maximax or Minimin) criterion

In this criterion the decision-maker ensures that he should not miss the opportunity to achieve the largest possible profit (maximax) or the lowest possible cost (minimin). Thus, he selects the decision alternative that represents the maximum of the maxima (or minimum of the minima) payoffs (consequences or outcomes).

Procedure:

- *Locate the maximum (or minimum) payoff values corresponding to each decision alternative*
- *Select a decision alternative with best payoff value (maximum for profit and minimum for cost)*

2.4.2 Pessimism (Maximin or Minimax) criterion

In this criterion the decision-maker ensures that he would earn no less (or pay no more) than some specified amount. Thus, he selects the decision alternative that represents the maximum of the minima (or minimum of the minima in case of loss) payoff in case of profits.

Procedure:

- *Locate the minimum (or maximum in case of profit) payoff value in case of loss (or cost) data corresponding to each decision alternative*
- *Select a decision alternative with the best payoff value (maximum for profit and minimum for lost or cost)*

2.4.3 Equally Likely (Laplace) criterion

Since the probabilities of state of nature are not known, it is assumed that all states of nature will occur with equal probability, i.e. each state of nature is assigned an equal probability. As state of nature are mutually exclusive and collectively exhaustive, so the probability of each of these must be: 1/(number of states of nature).

Procedure:

- *Assign equal probability value to each state of nature by using the formula:*

  1 ÷ (number of states of nature)

- *Compute the expected (or average) payoff for each alternative (course of action) by adding all the payoffs and dividing by the number of possible states of nature, or by applying the formula:*

(probability of state of nature j) x (payoff value for the combination of alternative i and state of nature j)

- *Select the best expected payoff value (maximum for profit and minimum for cost)*

2.4.4 Coefficient of optimism (Hurwicz) criterion

This criterion suggests that a decision-maker should be neither completely optimistic nor pessimistic and, therefore, must display a mixture of both. We introduce the idea of a coefficient of optimism (denoted by α) to measure the decision-maker's degree of optimism. This coefficient lies between 0 and 1, where 0 represents a complete pessimistic attitude about the future and 1 complete optimistic attitude about the future. Thus, if α is the coefficient of optimism, then (1 — α) will represent the coefficient of pessimism. Then

H (criterion of realism) = α (Maximum in column) + (1 – α) (Minimum in column)

Procedure:

a. *Decide the coefficient of optimism α (alpha) and then coefficient of pessimism (1 – α)*
b. *For each decision alternative select the largest and lowest payoff value and multiply these with α and (1 – α) values, respectively. Then calculate the weightage average, H by using above formula.*
c. *Select an alternative with best weighted average payoff value.*

2.4.5 Regret (Savage) Criterion

The criterion is also known as opportunity loss decision criterion or minimax regret decision criterion because decision-maker regrets for choosing wrong decision alternative resulting in an opportunity loss of payoff. Thus, he always intends to minimize this regret.

Procedure:

a. *From the given payoff matrix, develop an opportunity-loss) or regret) matrix as*

follows:

- *Find the best payoff corresponding to each state of nature*
- *Subtract all other payoff values in that row from this value*

b. *For each decision alternative identify the worst (or maximum regret) payoff value. Record this value in the new row.*
c. *Select a decision alternative resulting in a smallest anticipated opportunity- loss value.*

Illustration Problem 20:

The states of nature and strategies for a company's decision process and the payoff are given in the following matrix:

Strategies	States of Nature		
	N1	N2	N3
S1	700000	300000	150000
S2	500000	450000	0
S3	300000	300000	300000

Which strategy should be concerned on the basis of

a) Maximin Criterion
b) Maximax criterion
c) Minimax regret criterion
d) Laplace criterion
e) Hurwicz criterion (α = 0.7)

Solution:

The payoff matrix is rewritten as follows:

a) Maximin criterion

States of Nature	Strategies		
	S1	S2	S3
N1	700000	500000	300000

N2	300000	450000	300000
N3	150000	0	300000

Column Minimum	150000	0	**300000** **Maximum Payoff**

The maximum of column minima is Rs.300000. Hence Strategy S3 is selected

b) Maximax criterion

States of Nature	Strategies		
	S1	S2	S3
N1	700000	500000	300000
N2	300000	450000	300000
N3	150000	0	300000

Column Maximum	**700000** **Maximum Payoff**	500000	300000

The maximum of column maximum is Rs.700000. Hence Strategy S1 is selected

c) Regret criterion

Regret table or opportunity loss table:

Subtract the values in each row from the highest value of that row

States of Nature	Strategies		
	S1	S2	S3
N1	700000 – 700000 =0	700000 – 500000 = 200000	700000 – 300000 = 400000
N2	450000 – 300000 = 150000	450000 – 450000 = 0	450000 – 300000 = 150000

N3	300000 – 150000 = 150000	300000 – 0 = 300000	300000 – 300000 = 0

Column Maximum	**150000** **Minimum Regret**	300000	400000

The minimum opportunity loss is Rs.150000. Hence Strategy S1 is selected

d) Laplace criterion:

Strategies	Expected Return
S1	700000 + 300000 + 150000 / 3 = **383333.33 → Largest Payoff**
S2	500000 + 450000 + 0 / 3 = 316666.66
S3	300000 + 300000 + 300000 / 3 = 300000

The largest payoff is Rs.383333.33 and strategy S1 is selected.

e) Hurwicz criterion (α = 0.7)

Strategies	Expected Return
S1	700000 x 0.7 + 150000 x 0.3 = **535000 → Largest Payoff**
S2	500000 x 0.7 + 0 x 0.3 = 350000
S3	300000 x 0.7 + 300000 x 0.3 = 300000

The largest payoff is Rs.535000 and strategy S1 is selected

2.5 DECISION TREE ANALYSIS

A decision tree analysis is a graphical representation of a decision-making process. It is used to clarify whether the decision is optimal or not. The analysis compares multiple courses of action against each other and displays the risks and benefits related to them visually. It is both a technique and a tool for risk management, used to support every probable alternative in action. When there are several ways to go for a company, the analysis of which one to follow is required. The alternative paths are considered, and the right one is chosen.

Every branch in a decision tree has an endpoint, which is the outcome of selecting a specific path in a decision-making process. The tree is evaluated by expected monetary or quantitative analysis performed for each branch. It shows the benefits of certain decisions or actions and the probability of the downsides of the project happening along with the actual outcome, positive or negative. The basic concept consists of nodes (where choices are made), branches (possible alternatives available at the nodes), and values. It is represented with text boxes, circles, squares, and triangles.

2.5.1 Elements of Decision Tree Chart

- Circles indicate probability nodes, which are a set of possible outcomes.
- Squares indicate decision nodes, which are the choices made by a decision-maker.
- Triangles (or bars) indicate the endpoint on a branch, which is an outcome.
- Alternative branches are lines, which indicate the available alternatives.

2.5.2 Advantages of Decision Tree Analysis

- Decision analysis is applicable in almost every area where there is more than one solution available (project management, organization strategies, budget planning, etc.).
- The analysis assists not only with current solutions and problems. It estimates the decision after it has been made, providing a view of what to expect in the future under similar conditions.
- The chart is easy to create and follow. If a decision-maker structures the diagram correctly, each element flows logically into the other.
- It is a part of a project, constantly growing alongside and accommodating any changes and updates.
- It requires little time and resources.

2.5.3 Disadvantages of Decision Tree Analysis

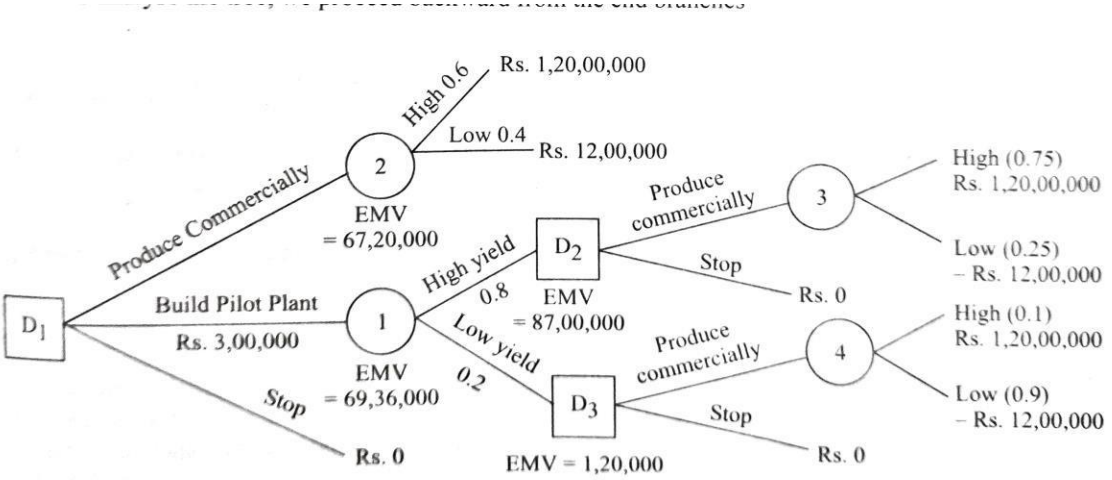
- If there are too many possible actions to be taken, the process becomes overly complicated.
- If a person who chooses between several courses does not involve every possible situation, a false choice can be made.

Illustration Problem 21:

A large steel manufacturing company has three with regard to production: i) produce commercially ii) build pilot plant iii) stop producing steel. The management has estimated that their pilot plant, if build, has 0.8 chance of high yield and 0.2 chance of low yield. If the pilot plant does show a high yield, management assigns a probability of 0.75 that the commercial plant will also have a high yield. If the pilot plant shows a low yield, there is only a 0.1 chance that the commercial plant will show a high yield. Finally, management's best assessment of the yield on a commercial-size plant without building a pilot plant first has a 0.6 chance of high yield. A pilot plant will cost Rs.3,00,000. The profits earned under high and low yield conditions are Rs.1,20,00,000 and Rs.12,00,000 respectively. Find the optimum decision for the company using decision tree analysis.

Solution:

A decision tree representing possible courses of action and states of nature is shown below:



EMV (Node 3) = 0.75 x 12000000 – 0.25 x 1200000 = Rs.87,00,000

EMV (Node 4) = 0.1 x 12000000 – 0.9 x 1200000 = Rs.1,20,000

EMV (Node 1) = 0.8 x 8700000 – 0.2 x 120000 = Rs.69,36,000

EMV (Node 2) = 0.6 x 12000000 – 0.4 x 1200000 = Rs.67,20,000

EMV (Node D2) = Rs.87,00,000

EMV (Node D3) = Rs.1,20,000

EMV (Node D1) = Rs.69,36,000 – 3,00,000

= Rs.66,36,000

Since at decision node D1 the production cost of Rs.67,20,000, associated with course of action – Build pilot plant is least the company should build pilot plant.

Let's Sum Up

This section of learning material provides a frame work for rational decision making in different decision-making environments. You are provided with the various decision-making criterion for finding out the optimal solution in uncertainty conditions.You have learnt that the uncontrollable factors (also called states of nature) affectthe payoff likely to be obtained from a course of action. As a learner you are also exposed the decision tree analysis which is used to display graphically the progression of

sequential decisions and random events. Problems involved the above area have been illustrated for your better understanding.

Check Your Progress -Quiz 2

1. Decision making under uncertainty refers to ..... situation.

A. Deterministic

1. Probabilistic

B. Competitive

C. None of these

2. The coefficient of realism (ά) in the Hurwitz criterion represents _____

A. represents the degree of pessimism

B. represents the degree of optimism

C. is the probability of a state of nature

D. none of these

3. Essential characteristics of a decision model are _____

A. states of nature

B. pay off

C. decision alternatives

D. all of the above

4. Which of the following criterion is not used for decision making under uncertainty

A. maximin

B. maximax

C. minimax

D. minimize expected loss

5. Decision Tree diagram helps in calculating the_____of the states of nature

A. Expected pay off

B. Strategies

C. Loss

D. None of the above

UNIT SUMMARY

This unit focuses on probability and its application in business and industrial problems. It covers fundamental rules of probability and probability distributions, including binomial, Poisson, and normal distributions, which are used to model and predict various outcomes. Bayes' Theorem is explored for its use in updating probabilities based on new information. Decision-making under risk and uncertainty is addressed through criteria like Maximax, Maximin, Regret, Hurwitz, and Laplace, which help in evaluating different outcomes. Additionally, the decision tree technique is introduced as a visual tool to map and analyze decisions in uncertain conditions.

GLOSSARY

KEYWORDS	MEANING
Probability Distribution	A mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.
Binomial Distribution	A discrete probability distribution that describes the number of successes in a fixed number of independent Bernoulli trials, with the same probability of success on each trial.
Poisson Distribution	A discrete probability distribution that expresses the probability

	of a given number of events occurring within a fixed interval of time or space, under the assumption that these events occur independently.
Normal Distribution	A continuous probability distribution characterized by its bell-shaped curve, where most of the values lie around the mean, and the probabilities taper off symmetrically as you move away from the mean.
Decision Tree	A graphical representation of possible solutions to a decision based on different conditions. It is used for decision analysis and helps visualize the consequences of various actions.

Self-Assessment Questions

1.This following table shows the number of athletes who stretch before exercising and how many had injuries within the past year:

	Injury in last year	No injury in last year	Total
Stretches	55	295	350
Does not Stretch	231	219	450
Total	286	514	800

a) What is the probability that a randomly selected athlete stretches before exercising or had an injury last year? (0.72625)

b) What is the probability that a randomly selected athlete does not stretchbefore exercising or had no injuries in the last year? (0.93125)

2. At a local college, 60% of the students are taking a math class, 50% of the students are taking a science class, and 30% of the students are taking both a math and a science class.

a) Are the events “taking a math class” and taking a science class” mutually exclusive? Explain.

b) What is the probability that a randomly selected student is taking a math class or a science class? (0.8)

3. In a manufacturing company, on an average, 2 out of every 100 items coming off an assembly line are defective. The quality control supervisor wants to know the probabilities that in the 2 items randomly selected from the assembly line,

a) both are defective

b) first is defective and second is not

c) first is not defective and second is, and

d) neither of the selected items is defective.

(Ans.: a) 0.0004, b) 0.0196, c) 0.0196, d) 0.9608)

4. Among the sales staff engaged by a company, 60 percent are females. In terms of their professional qualifications, 70 percent of females and 50 percent of males possess a diploma in sales. Find the probabilities that a salesperson selected at random will be a) a male with diploma in sales, and b) a female with diploma in sales

The manufacturer of a certain product has installed three machines, A, B and C, all meant for producing a given product. All the three machines are equally efficient and constitute 25%, 35% and 40% respectively, of a day’s total production. It has been found that, on an average, machine A produces 1% defective items, B produces 2% defective items and C produces 3% defective items. An item is drawn at random from the combined output of all the three machines produced during a specified hour. Find the probabilities that the item selected is produced by a) A, b) B and c) C respectively.

(Ans. A) 0.116, B0 0.326, c) 0.558)

5. Assume that on an overage one telephone number out of fifteen is busy. What is the probability that if six randomly selected telephone numbers are called:

a) not more than three will be busy? (0.9997)

b) at least three of them will be busy? (0.0051)

6. Bottled sweet milk scored in a godown is reported to have gone sour. A test check revealed that milk in 25% of the bottles is unfit for consumption. The salesman at a retail outlet offers 5 bottles for sale on demand, find the probability that milk will be unfit for consumption

a) exactly in 2 bottles (0.2636)

b) at least in 2 bottles (0.3672)

c) at the most in 2 bottles (0.8964)

7. A manufacturer of electric fuses packs fuses in boxes of 10 each and 2000 such boxes were sold. The previous experience shows that 5% of the fuses are defective. Using Poisson distribution, find how many boxes will contain

a) no defective

b) more than one defective

8. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts a) 0, b) 1 and c) at the most 2 bolts will be defective

(Ans.: a) 0.4096, b) 0.4096 and c) 0.9728)

9. The time taken to assemble a car in a certain plant is a random variable having a normal distribution of 20 hours and a standard deviation of 2 hours. What is the probability that a car can be assembled at this plant in a period of time
a) less than 19.5 hours? (0.4013)

b) between 20 and 22 hours? (0.3413)

10. As a result of tests on 20000 electric fans manufactured by a company, it was found that lifetime of the fans was normally distributed with an average life of 2040 hours and standard deviation of 60 hours. On the basis of the information, estimate the number of fans that is expected to run for

a) more than 2150 hours (672)

b) Less than 1960 hours (1836)

10. Mr.Sundaram has Rs.10000 to invest in one of three options. A, B or c. The return on his investment depends on whether the economy experiences inflation, recession or no change at all. The expected returns under each economic condition are given below:

Strategy	States of Nature		
	Inflation	Recession	No change
A	2000	1200	1500
B	3000	800	1000
C	2500	1000	1800

What should he decide using the pessimistic criterion, optimistic criterion, equally likely criterion, regret criterion and Hurwicz criterion (α = 0.6)

11. The research department of Hindustan Ltd. has recommended to pay marketing department to launch a noodle of three different types. The marketing types of noodles to be launched under the following estimated pay- offs for various level of sales.

Type of noodle	Estimated Sales in Units		
	25000	15000	10000
Masala	30	10	10
Chicken	40	15	5
Mixed Veg	55	20	3

What will be the marketing manager's decision if?

a) Maximin criterion
b) Minimax criterion
c) Equally likely criterion
d) Laplace criterion (α= 0.7)
e) Minimax regret criterion?

12. XYZ company manufactures guaranteed tennis balls. At present approximately, 10 percent of the tennis balls are defective. A defective ball leaving the factory costs the company Rs.0.50 to honour its guarantee. Assume that all defective balls are returned. At a cost of Ts.0.10 per ball, the company can conduct a test that always correctly identifies both good and bad tennis balls

    a) Draw a decision tree and determine the optimal course of action and itsexpected cost

    b) At what test cost should the company be indifferent to testing?

ANSWERS FOR CHECK YOUR PROGRESS

CHECK YOUR PROGRESS – QUIZ 1

1) b) Mutually exclusive events
2) d) Posterior Probabilities
3) c) Small values of ‘n’
4) b) Discrete Random Variable
5) b) Mean

CHECK YOUR PROGRESS – QUIZ 2

1) a) Deterministic
2) b) Represents the degree of optimism
3) d) all of the above
4) d) minimize expected loss
5) d) none of the above

Activities:

1. Collect coins of your choice. Toss a single coin for one time, two times and multiple times. Observe the outcomes and classify the type of events, probabilities and distributions formed.

2. Similarly, toss two coins and observe the frequencies, events and distributions made. Classify the type of events and probability distributions

3. List few activities on your day-to-day program. Categorize them based on the type of decision-making environment viz. Certainty, uncertainty and risk. Identify their possible state of nature and courses of action.

Suggested Readings:

1. J.K.Sharma, "Operations Research — Theory and Applications", MacMillan Publishers India Ltd., Revised 5th Edition, 2013

2. S.P.Gupta and M.P.Gupta, "Business Statistics", Sultan Chand & Sons, 2019

3. Amir D Aczel & Jayavel Sounderpandian, "Complete Business Statistics" McGraw Hill Education, 2017

4. G.S.Monga, "Mathematics and Statistics for Economics', Vikas Publishing House, 2001

UNIT II

RESEARCH METHODS

Unit – II: Research Methods
Introduction: Probability - Rules of probability- Probability distribution; Binomial, Poisson and Normal Distributions, their applications in Business and Industrial Problem- Baye's Theorem and its applications - Decision Making under risk and uncertainty; Maximax, Maximin, Regret Hurwitz and Laplace Criteria in Business and Decision Making - Decision tree.

RESEARCH METHODS

Section	Topic
	UNIT - II
	Unit Objectives
	RESEARCH METHODS
2.1	Meaning of research
2.2	Definition of research
2.3	Objectives of research
2.4	Characteristics of research
2.5	Research process
2.6	Role of theory in research:
2.7	Research design:
2.8	Types of research design:
2.9	Variables in research
2.10	Hypothesis:
2.11	Types of data:
2.12	Measurement scales:
2.13	Sampling methods
	Let Us Sum Up
	Check Your Progress – Quiz 1
	Unit summary
	Glossary
	Self- Assessment Questions
	Activities / Exercises / Case Studies
	Answers for Check your Progress
	References and Suggested Readings

UNIT OBJECTIVES

- To study the meaning, objectives of research, research process andresearch design
- To understand the types of data, methods of primary data collectionmethods
- To exhibit the construction of questionnaire and measuring validity and reliability
- To learn the meaning and types of scales, types of attitude measurement scales and sampling techniques

2.1 MEANING OF RESEARCH

It is the process of search for knowledge. It is also defined as scientific and systematic search for information on any branch of knowledge. It is the journey of discovery. The curiosity to know more about something drives one person to search more which ultimately brings out more new facts. The technique and methods one adopts to search, obtain, and record the gained information systematically in known as the process of Research.

2.2 DEFINITION OF RESEARCH

D. Slesinger and M. Stephenson in the Encyclopedia of Social Science define research as "the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art.

According to the American sociologist Earl Robert Babbie, "research is a systematic inquiry to describe, explain, predict, and control the observed phenomenon. Itinvolves inductive and deductive methods.

Research is defined as the creation of new knowledge and/or the use of existing knowledge in a new and creative way so as to generate new concepts, methodologies and understandings. This could include synthesis and analysis of previous research to the extent that it leads to new and creative outcomes.

2.3 OBJECTIVES OF RESEARCH

1. To gather new information,
2. To gain familiarity with/ develop new insights into some phenomenon,
3. To describe/ depict the true characteristics of a person, object, situation, or a particular group,
4. To test a hypothesis on the basis of available information,
5. To identify patterns of frequency of occurrence or trends related to them,
6. To investigate the cause of a particular situation or a problem.

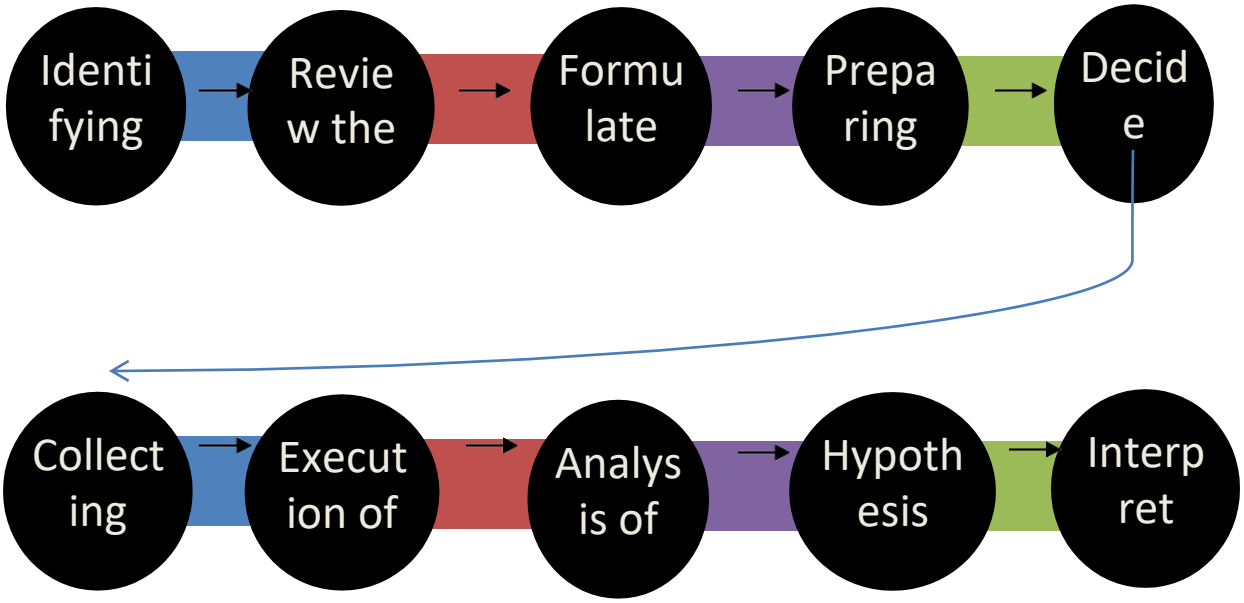
2.4 CHARACTERISTICS OF RESEARCH

1. Good research follows a systematic approach to capture accurate data. Researchers need to practice ethics and a code of conduct while making observations or drawing conclusions.
2. The analysis is based on logical reasoning and involves both inductive and deductive methods.
3. Real-time data and knowledge is derived from actual observations in natural settings.
4. There is an in-depth analysis of all data collected so that there are no anomalies associated with it.
5. It creates a path for generating new questions. Existing data helps create more research opportunities.
6. It is analytical and uses all the available data so that there is no ambiguity in inference.
7. Accuracy is one of the most critical aspects of research. The information must be accurate and correct. For example, laboratories provide a controlled

environment to collect data. Accuracy is measured in the instruments used, the calibrations of instruments or tools, and the experiment's final result.

2.5 RESEARCH PROCESS

The research process in the following chart provides course of actions to be followed step by step. At the same time, one should understand that the process consists of related activities and does not have to strictly follow the sequence and may plan and work for the subsequent steps in advance.



2.5.1 STEPS IN RESEARCH PROCESS

2.5.1.1 DENTIFYING THE RESEARCH PROBLEM:

Finding an issue or formulating a research question/problem is the first step**.** Research Problem can be categorized into the following two types,

1. Problems relating to the nature of an object, place, person or an event,
2. Problems relating to the relationship between variables.

A researcher must identify his area of interest and narrow down the general topic into specific research question/problem. He must make himself familiar with the chosen research question and proceed with discussing the colleagues, experts from government or concerned business units, and academic guides etc.

2.5.1.2 LITERATURE REVIEW:

The researcher must study all the literature available pertaining to research question. He may check out the concepts and theories available or he may scrutinize the empirical literature made earlier related to the topic. He may refer to the government reports, academic journals, books, conference proceedings. Listing the referred literature and indexing them is carried out in this stage. This careful consideration of facts and information will give better understanding of the problems and gaps.

2.5.1.3 HYPOTHESIS FOMULATION:

Hypothesis is an idea or mere assumption of something which is not yet found to be true or not. Testing the assumption by suitable methods to check its logical relationship or empirical consequences is called the working hypothesis. The hypothesis helps the researcher focus on the research area and guides him to be on the right track. It directs the researches to pay attention to the important factors of the problem. Sometimes we may not need to develop a working hypothesis as the research does not need to test the hypothesis as in case of exploratory research, though it is the basic step in most research process. Working hypothesis is drafted as a result of prior examination of available information from related studies and assumptions derived after discussion with relevant experts.

2.5.1.4 PREPARING/FORMULATING THE RESEARCH DESIGN:

Research Design is considered to be blueprint of the structure with which the researcher tries to conduct the research. Based on the purpose of the research, the research design is shaped to facilitate the collection of maximum information/facts/evidence with minimum effort, time and money. Research purpose is categorized in to four types, they are

1. Descriptive
2. Exploratory
3. Experimental
4. Diagnostic

A good research design offers appropriate methods to collect reliable data with minimum bias for studying a particular situation or relationship between variables. It usually involves the following:

1. The means of collecting the information,
2. The skills of the researcher and the support staff,
3. The availability of time,
4. Cost to be involved.

2.5.1.5 SAMPLE DESIGN:

Sampling is the process of selecting few predetermined number of items from a larger population, considering that the selected few items represent the population characteristics. It allows the researchers to do study about the targeted population using the smaller portion of it.

A complete study of each item in a population which is known as census method may provide greater accuracy but for a larger population it is not feasible as it involves large amount of time, energy and cost. Even in census method, to check the element of bias either resurvey is done completely or sample checks is done. The numbers of items thus selected for the study forms the sample.

The plan for selecting a sample from a given population is known as sample design. Based on the way of selection, it can be determined as probability or non-probability samples. In probability samples, the chance of each element getting selected is equal. In non-probability samples, the chance of any element getting selected is unknown and quite arbitrary based on personal judgements.

2.5.1.6 DATA COLLECTION:

It is necessary to collect appropriate data for the research study. However the data may not be readily available and it is the responsibility of the researcher to gathert the

required data. There are many ways of collecting data considering the availability of time, money and related resources. The data which is collected directly with respect to the problem under study is known as primary data. The data collected is original in character. The data collected during the course of experimentation in experimental research is also primary data. The other methods of data collection are done through observation, personal or telephone interviews, and questionnaires.

Secondary data is the data which is already collected and processed by people other than the researcher concerned. Secondary data are usually collected from a) magazines and newspapers, b) academic books, c) government reports, d) historical documents, e) research publications, f) conference proceedings g) trade journals, h) reports and publications from various associations connected with business and industry etc.

2.5.1.7 CONDUCTANCE OF RESEARCH:

It is a very important step in the conducting the research process. He must ensure the timely execution of each step in the process. The researcher should make appropriate arrangements for proper selection and training of the interviewers. He must do the field checks to ascertain whether the project is right on track. He must be ready to face unanticipated and contingent situations and keep the research as realistic as possible.

2.5.1.8 ANALYSIS OF DATA:

Once the data is collected, it is processed and analyzed for making statistical inferences. The collected data is edited, coded, tabulated and analyzed for making useful interpretations. The huge dataset usually is processed and computed using computers to understand about the variables affecting a problem statement.

2.5.1.9 HYPOTHESIS-TESTING:

After analyzing the data, the researcher has to test the formulated hypotheses. He has to check the facts collected whether they support the hypotheses or not. Various tests of measures of relationship, variance and independence are conducted depending on the

nature of research inquiry. Various tests such as Chi square test, t-test, F- test have been developed by various statisticians for the purpose. Appropriate hypothesis testing method is adopted to check whether to accept the hypotheses or reject it.

2.5.1.10 GENERALISATIONS AND INTERPRETATIONS:

After hypothesis is tested several times, the researcher arrives at conclusion to construct theories and generalizations. The process of interpretation helps in answering the research question and thus aids in gaining new insights and conclusions.

2.5.1.11 PREPARATION OF THE REPORT:

Final step in the research process is the preparation of the report containing the details of the entire work done by him. The report writing must be done with utmost care. The format of a research report as follows,

1. Title page
2. Acknowledgements and foreword
3. Contents page
4. List of tables
5. List of graphs and charts
6. Executive Summary
7. Introduction
8. The objectives of the study
9. The scope of the study
10. The limitations of the study
11. Body
12. Conclusions and Recommendations
13. Bibliography
14. Appendix

A good report must be written in a simple language and it should be in a concise manner.

2.6 ROLE OF THEORY IN RESEARCH:

Application of theory has various roles to play in research. The most important roles of theory are as follow:

1) Understanding of the phenomenon for which they account,
2) Providing a basis for prediction,
3) Guiding the direction for research.

2.6.1 UNDERSTANDING:

Theories represent a particular way to understand the phenomena with which they deal. Theories make the understanding of an underlying reality deep and meaningful. Theories provide deep insights and give us the better understanding of a particular field. It provides broad description of particular characteristics of element of discussion.

2.6.2 PREDICTION:

Even when theories do not provide a fundamental insight into the mechanisms of a system, they at least can provide a way to provide the behavior of the system under different values of its controlling variables. A good theory can predict the empirical outcome with reasonable degree of precision.

2.6.3 GUIDANCE:

A theory can provide a sound framework for organizing and interpreting research results. It provides the needed guidance. It helps the researcher to organize the results to be within the existing structure of confirmatory and dis-confirmatory results. It helps in interpreting the results in the light of the theory.

2.6.4 CHARACTERISTICS OF A GOOD THEORY:

Characteristics of a good theory are as follows,

1. Ability to account for data
2. Explanatory relevance
3. Testability
4. Prediction of novel events

5. Parsimony

2.7 RESEARCH DESIGN:

After defining the research problem, the preparation of the design of the research project is known as the research design. Research is the blueprint of the research study. It includes an outline of what the researcher intends to do in the aspects of deciding what experiments to be conducted, tools used to collect the data, methods of sampling and how to handle other operational implications till final analysis of data.

A research design facilitates the efficient research operation, thus yielding better results at much less money and effort. This helps in advance planning of methods to be followed, tools for collection of data, training of staff and monitoring availability of resources for conducting the research smoothly.

2.8 TYPES OF RESEARCH DESIGN:

Research design can be categorized as follow,

1. Exploratory research design,
2. Descriptive research design,
3. Hypothesis-testing research design

2.8.1 EXPLORATORY RESEARCH DESIGN:

The main purpose of this research is to formulate a problem for more precise investigation and on the discovery of new ideas and insights. This involves converting broad, vague problem statements into small and precise problem statements to formulate specific hypothesis. In the early stages of research, one usually lacks from sufficient understanding of the problem to formulate a hypothesis. Thus, a flexible research design is required to consider different aspects of the problem under study.

The following are the areas in which the exploratory study is ideally suited:

1. To gain deep insights into the problem.
2. To generate new ideas.

3. To list all possibilities of the problem.
4. To gain familiarity with a new problem area.
5. To establish priorities so that further research can be conducted.
6. To develop a hypothesis.

2.8.1.1 Characteristics of Exploratory Research:

- Exploratory research is flexible and very versatile.
- Experimentation is not a requirement.
- Cost incurred to conduct study is low.
- This type of research allows very wide exploration of view.
- Research is interactive in nature and also it is open ended.

2.8.1.2 Advantages of Exploratory Research

Investigating new alternatives and possibilities is crucial in light of our constantly evolving environment. In order to accomplish this, exploratory research is a fantastic tool. This strategy has numerous advantages, including creativity and innovation.

- You can be receptive to new concepts and opportunities by conducting exploratory research. This may lead to more creative solutions to problems.
- It promotes problem-solving, i.e., when you investigate novel concepts, you are likely to come up with answers to issues using exploratory research. Therefore, you can tackle challenging challenges more effectively.
- The major benefits of doing exploratory research are that it is adaptable and enables the testing of several hypotheses, which increases the flexibility of your study. It implies that you may test out several strategies to find the most effective.
- Using exploratory research techniques will increase the likelihood that you will produce reliable, valid research findings. Using this data, you can make more reliable inferences.

- The more you conduct research using the exploratory research approach, the more proficient you become. For example, you can learn to distinguish between excellent and terrible questions or ask them effectively.
- Using exploratory research techniques can make it simpler to make judgments based on more information than what you already know about the issue.
- When you employ exploratory research techniques, it will be simpler to present your facts accurately and truthfully. Adopting these techniques makes eliminating biases that might result from reporting on prior hypotheses and facts easier.

2.8.1.3 Disadvantages of Exploratory Research

Exploratory research has its own disadvantages:

- Exploratory research frequently has unclear objectives since it is exploratory. There is a possibility that the researcher won't have all the information needed to do the study. Ultimately, the researcher and participants may experience dissatisfaction and misunderstanding.
- Exploratory research can be challenging and time-consuming. Determining which questions to ask, how to gather data, and how to evaluate it might take a lot of work. Therefore, it can be challenging for researchers to finish their work within the allotted time or budget.
- Exploratory research has many uses but doesn't always yield accurate or valid conclusions. Rather than using actual data, exploratory study frequently relies on theories. The research's findings might be deceptive or invalid if the hypothesis is unreliable or unsubstantiated.
- A problem or issue may not always be discovered via exploratory investigations. The reason is that open-ended questions, frequently used in exploratory research, cannot elicit all the necessary data to resolve a problem.

2.8.1.4 FORMULATION OF HYPOTHESIS IN EXPLORATORY RESEARCH:

The best way to formulate a hypothesis in exploratory research is by using the following four methods,

1. Literature Survey,
2. Experience Survey,
3. Focus Group,
4. Case Studies

2.8.2 DESCRIPTIVE RESEARCH DESIGN:

As the name suggests, it is essentially research to describe something. Studies concerning the description of characteristics of individual, group, organization, markets or situation are the examples of descriptive research. Description research provides association between two or more variables like income and place of shopping, age and preferences. The study aims at gathering complete and accurate information in the said studies, the procedure must be carefully decided. The research design must provide protection against bias and provide more reliability. The research design must be rigid in descriptive research as opposed to be flexible as in case of exploratory research.

The following are the areas where descriptive research is best suited:

1. To determine the characteristics of market such as:
    a. Size of the market
    b. Buying power of the consumer
    c. Product usage pattern
    d. To find out the market share for the product
    e. To track the performance of a brand
2. To determine the association of the two variable such as income and customerpreference,
3. To estimating the proportion of people in a specific population, who behave in a particular way,
4. To make a prediction or to make forecast.

2.8.2.1 Characteristics of descriptive research:

1. Research design is rigid,
2. Objectives are clearly defined,
3. Data collection methods are predefined,
4. Structured instruments are used,
5. More reliability and less bias.

2.8.2.2 Advantages of Descriptive Research

a. Descriptive research has many advantages over other research methods. These advantages include the ability to collect data from a large number of participants, the ability to explore different aspects of a topic, and the ability to understand complex topics.
b. One of the main advantages of descriptive research is that it is able to collect data from a large number of participants. This can be helpful in understanding complex topics. This makes it an important tool for researchers who want to understand complicated issues.
c. Another advantage of descriptive research is that it can explore different aspects of a topic. This can be helpful in understanding how people use products or how people feel about a particular situation. For example, descriptive research could explore how people use a product in different ways or how people feel about a situation after hearing about it.
d. Finally, descriptive research is that it is usually cheaper, easier and quicker to carry out than other methods of research. It also increases the reliability of results, as participants are less likely to lie or select inappropriate answers.

The strengths of descriptive research are that it can ask a large number of participants about a topic. The researcher only needs to contact people who might be interested in the topic and ask them questions about it. This means especially that qualitative researchers can gather information from a large number of participants, if they have an interest in the topic.

2.8.2.3 Disadvantages of Descriptive Research

Though, descriptive research has quite great advantages, however, it is often criticized for its disadvantages.

a. One of the main disadvantages is that it can be difficult to design a study that is relevant to the researcher's specific interests. This can make it difficult to get accurate information from participants.
b. Another disadvantage of descriptive research is that it can be difficult to generalize the findings from the study. The results of the study may not apply to all people or situations. Additionally, descriptive research does not allow for any conclusions to be drawn about cause and effect.
c. The main weakness of descriptive research is that there is no control group. It cannot tell us whether one treatment works better than another. Instead, it must rely on the subjective opinions of the participant or participants in the study, which can make it more difficult for researchers to draw conclusions about why certain treatments work better than others.
d. In Descriptive research, studies that focus on people's perceptions about a particular topic may not provide useful information on how the general population feels about it. This can be especially important for policy makers who must decide which views should be taken into consideration when making decisions about certain subjects or issues.

Check your Progress – Quiz I

1. The first step of research is_____

A. Selecting a problem
B. Searching a problem
C. Identifying a problem.
D. Finding a Problem

2. Research can be classified as _____

A. Basic, Applied and Action Research
B. Philosophical, Historical, Survey and Experimental Research
C. Quantitative and Qualitative Research

D. All the above.

3. The research that applies the laws at the time of field study to draw more and more clear ideas about the problem is _____

A. Applied research.

B. Action research

C. Experimental research

D. None of these

4. Which of the following would be true regarding exploratory research?

A. Exploratory research is highly structured

B. Exploratory research is very formal

C. Exploratory research determines causality

D. Exploratory research is both structured and informal

5. What is the primary focus of descriptive research?

A. Identifying cause-effect relationships

B. Describing characteristics and traits of a phenomenon

C. Exploring new areas of research

D. Evaluating the effectiveness of an intervention

2.9 VARIABLES IN RESEARCH

In research, Variables are any concept/ characteristics that can take on different quantitative values such as height, age, temperature, and some qualitative characteristics such as gender, satisfaction level, customer preference, consumer behavior, attitude etc which can be measured based on presence or absence, and can be manipulated or controlled. They are the factors that researchers observe or manipulate to understand the relationship between them and study their respected outcomes. Following are the types of variables in research,

1. Independent Variable,
2. Dependent Variable,
3. Continuous Variable,
4. Categorical Variable,
5. Discrete Variable,

6. Confounding Variable,
7. Extraneous Variable.

INDEPENDENT VARIABLE: A concept/ characteristic which are manipulated by the researcher is known as the Independent variable. It is also called as predictor variables (they can be used to predict the value of a dependent variable) or explanatory variables (they explain an event or outcome). It is not influenced by any other variables in the study.

DEPENDENT VARIABLE: This is the variable which is measured or observed to determine the effects of other variable, in specific the independent variable. It is also called as response variables (they respond to a change in other variables) or outcome variables (they represent the outcome which you want to measure).

CONTINUOUS VARIABLE: This is the variable that can take on any value within a certain range. Age, height, weight and temperature are an example of continuous variable.

CATEGORICAL VARIABLE: This is a variable that can take on a limited number of values or categories. Categorical variables can be nominal or ordinal. Examples of categorical variable include gender, educational level, designation etc.

DISCRETE VARIABLE: These variables are non-continuous variables and can only be expressed in integer values. Examples of discrete variables include the number of times a person visits a shop and the number of students in a classroom.

CONFOUNDING VARIABLE: A confounding variable is an unmeasured third variable that influences both the cause factor and the effect factor. These unchecked variables cause research biases to your work and may interpret the results.

EXTRANEOUS VARIABLE: Independent Variables that are not related to the purpose of **the** study, but may affect the dependent variable are termed as extraneous variables. A study must always be so designed that the effect upon the dependent variable is attributed entirely to independent variable and not to some extraneous variable.

2.10 HYPOTHESIS:

Hypothesis is an unproven assumption or proposition that explains certain facts or phenomena. It is mere idea or supposition which needs to be proved or disproved. It is a set of statements about which the research wants to test the occurrence of the specified phenomena. By collecting the scientific evidence, one must be able to decide whether the theoretical hypothesis is accepted or not.

2.10.1 THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS:

2.10.1.1 Null Hypothesis

A null hypothesis is a general assertion that there is no relationship or effect between two observed phenomena. Null hypothesis is expected to be negated or disproved through experimentations. As a result, alternate hypothesis is considered to be accepted.

For example, the null hypothesis proposes that a new training method does not yield increased productivity among employees. The researcher tries to study the effect of the new training method on productivity levels. Through proper test and measures, on must see if null hypothesis is proved or disproved. We conclude that rejecting null hypothesis means accepting alternate hypothesis. Hence a researcher must decide the null hypothesis and alternate hypothesis before sample is drawn. Researcher must not make the mistake of deciding the hypotheses after collecting the data and observing it.

2.10.1.2 Alternative Hypothesis

Alternative hypothesis is usually the one which a researcher wants to prove and the null hypothesis is the one which he wants to disprove. It is statement that a researcher thinks it is true and ultimately leads you to reject the null hypothesis.

The null hypothesis is generally symbolized as Ho.

The alternative hypothesis is symbolized as Ha.

The purpose of hypothesis testing is to check whether which of the two hypotheses is correct.

2.11 TYPES OF DATA:

2.11.1 Primary and Secondary Data

Type of data may be primary data or secondary data. The data that is collected fresh for the first time and happen to be original in character is called primary data. The data that is already collected and processed is called secondary data. The researcher must decide on the type of data suited for the study.

2.11.1.1 Primary Data:

Primary data is the data that is collected for the first time through personal experiences or evidence, particularly for research. It is also described as raw data or first-hand information. The mode of assembling the information is costly, as the analysis is done by an agency or an external organisation, and needs human resources and investment. The investigator supervises and controls the data collection process directly.

2.11.1.2 Secondary Data:

Secondary data is a second-hand data that is already collected and recorded by some researchers for their purpose, and not for the current research problem. It is accessible in the form of data collected from different sources such as government publications, censuses, internal records of the organisation, books, journal articles, websites and reports, etc.

2.11.1.3 Primary Data Vs. Secondary Data

Primary Data	**Secondary Data**
Definition	
Primary data are those that are collected for the first time.	Secondary data refer to those data that have already been collected by some other person.
Originality	
These are original because these are collected by the investigator for the first time.	These are not original because someone else has collected these for his own purpose.
Nature of Data	
These are in the form of raw materials.	These are in the finished form.
Reliability and Suitability	

These are more reliable and suitable for the enquiry because these are collected for a particular purpose.	These are less reliable and lesssuitable as someone else has collectedthe data which may not perfectly match our purpose.
Time and Money	
Collecting primary data is quite expensive both in the terms of time and money.	Secondary data requires less time and money; hence it is economical.
Precaution and Editing	
No particular precaution or editing is required while using the primary data as these were collected with a definite purpose.	Both precaution and editing are essential as secondary data were collected by someone else for his own purpose.

Collection of primary data is original in character whereas the collection of secondary data is just the compilation of already available data. There are different methods of data collection, each one of which has its own advantages and disadvantages.

2.11.2 PRIMARY DATA COLLECTION METHODS

1. Observation method
2. Interview Method
3. Questionnaire Method
4. Collection of data through schedules

2.11.2.1 OBSERVATION METHOD:

There are several methods of observation of which any one or some of the combinations can be used. They are as follows,

- Structure or unstructured method
- Participant or non-participant observation method
- Disguised or undisguised method
- Controlled or uncontrolled observation method
- Direct-indirect observation
- Human-mechanical observation

Structured or Unstructured Method:

Under structured observation method, the careful definition of the units is ensured, the style of selecting, observing and recording the observed information are standardized. The accuracy of observation is ensured.

But when the above characteristics are not adopted and not planned well in advance, then the observation method is considered to be unstructured observation. Structured observation is appropriate in descriptive studies, whereas unstructured observation is chosen for the exploratory study.

Participant or Non-Participant Method:

When the observer in the context of research shares the life of the group he is observing, then the observation method is called participant observation. It means when the observer makes himself part of the member of the group he is observing, so that he experiences the same as what the group experiences and records the experiences and information better.

But when the observer is detached from the group and does not want to experience what the group feels through participation is called non-participant observation.

Disguised or Undisguised Method:

In disguised observation, the respondents are not aware that they are being observed. In undisguised observation, the respondents are well aware that they are being observed.

Controlled or Uncontrolled Observation Method:

If the observation happens in natural environment, it is termed as uncontrolled observation. If the observation is carried out in definite and pre-arranged settings, it is termed as controlled observation. In uncontrolled observation, the information collected has tendency of naturalness of the behavior of an individual or situation. In controlled observation, we use accurate measurement is done using instruments as the entire

observation is carried out under controlled circumstances. Example: laboratory experiments.

Direct and Indirect Observation: In direct observation, the actual behavior while expressed is observed. In indirect observation, the results of the behavior are observed. Here the researcher collects the details of the consequences and try to relate the actual behavior indirectly.

Human-Mechanical Observation:

In human observation, the trained observers perform the recording of information of their observation. In mechanical observation, the mechanical devices such as cameras, scanners, audiometers, or other sensors are used to record the information pertaining to the study.

2.11.2.2 INTERVIEW METHOD:

An interview is called personal interview when the interviewer/researcher questions the person of interest face-to-face. The place of interview can be home, college, on the street, or any agreed upon place. The interviewer asks the questions directly and the respondent answers and that is how he collects the information. This method is very much suitable for intensive investigation. Under this interview method, the interviewer has to contact the persons concerned whom the data has to be collected which is not possible all the time.

Other indirect methods can be used for oral examination of the person of interest. Telephonic interview is the methods for indirect collection on data. This method involves the respondent on phone and records his response.

2.11.2.3 COLLECTION OF DATA THROUGH QUESTIONNAIRES:

This is one of widely used method of data collection methods for making big enquiries. It is used by researchers, individuals, private and public organizations. In this method, a list of questions is prepared and sent to the persons concerned with a request to answer and send them back. These questions were usually sent by post. At a later

stage, the questions were typed and send as mail to be answered. The questions were read and appropriately responded with answers in the spaces meant for them. The respondents have to answers on their own and usually done at their convenience.

2.11.2.4 COLLECTION OF DATA THROUGH SCHEDULES:

This method of data collection is similar to the collection of data through questionnaire, but in this method the list of questions is carried to the respondents and being filled by the designated person sent for collection of data. Here the designated person is known as the schedule and does the job of an enumerator. The enumerator explains the objects of the investigation and removes the barriers which any respondents may feel in understanding the questions and makes the effort of explaining the concepts and definitions wherever needed. It is a bit costly method as it involves the training and deployment of the enumerators at right place and time.

2.11.3 SECONDARY DATA:

The data collected by the researcher which is already available and possibly processed by others. The data collected and compiled by anybody other than the researcher is called secondary data. They are the statistical data that are collected not for immediate use.

Some of the common sources of secondary data are as follows, census, large surveys, government reports, academic journals, conference proceedings, organizational records etc.

a. Government reports,
b. Reports of various association associated with industry, banks, and various other agency bodies,
c. Academic reports of universities, economists and scholars,
d. Technical and trade journals,
e. Library books,
f. Journals,
g. Census,

h. Conference proceedings,
i. Newspapers etc

2.11.4 VALIDITY AND RELIABILITY OF INSTRUMENTS:

Validity is the criteria which say about the degree to which an instrument should measure what it is supposed to measure. It tells about the extent to which an instrument measures what we want to measure. If an instrument measures differences between two object, it should actually reflect the true difference. But the question here is, how validity is determined or what sort of confirmation is needed to check the validity of the instrument. Validity is checked based on the below three types,

1) Content validity
2) Criterion- related validity
3) Construct validity

Content validity: Content validity says about the coverage of the topic under study. There is no quantitative way of expressing it. Content validity is determined by the judgments of experts pertaining to the subject. It applies to any context where you create a test or questionnaire for a particular research problem and ensure that the questions asked actually measures what you intend to measure.

> Example: If you want to do research on health needs of hostel students in a university through a survey, then one must cover all the aspects of health needs which includes physical, mental, social and environmental elements. Then your questionnaire will have high content validity. If some aspects are not covered, then the results may not provide accurate representation of health needs of hostel students.

Criterion-related validity:

It relates to the ability to predict some of the future outcomes based on some current criteria. The criteria should possess the following qualities:

1) Relevance

2) Freedom from bias
3) Reliability
4) Availability

It is divided into two types: 1) Predictive validity and Concurrent validity.

Construct validity: A measurement tool is supposed to have construct validity if the predicted relationships confirm with other theoretical propositions. The set of obtained results are compared with the sound theories and if it is found to align with the theoretical concepts, then the measure is considered to be have predicted correlations and valid. A measuring instrument tool is valid if above stated criteria are met with and tends to provide correct results.

2.12 MEASUREMENT SCALES:

The measurement scales are broadly classified as follows:

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

2.12.1 NOMINAL SCALE:

In nominal scale, numbers are used to identify or categorize the objects. Here the numbers are used just for the sake of labeling them. The assigned numbers plays the role of just identifying a particular object or group of objects. The labels given are just to identify and count the subjects. Giving numbers does not establish an order of ranking but to simply identify events or objects or individuals and count their presence or absence or count their number of occurrences.

The numbers assigned has no quantitative value. Nominal scale provides ways oftracking the events or objects or people. Example can be seen in the numbers assigned to sports players in a particular game. The numbers don't carry quantitative value

whereas to identify or label them. The counting of numbers in each group is the only arithmetic operation that can be performed when nominal scale is used. Measures of central tendency are limited to mode only.

Characteristics:

1. It is the least powerful level of measurement.
2. It has no arithmetic origin.
3. It shows no order or distance relationship.
4. It distinguishes or categorizes objects into different groups

Example:

Are you satisfied with the recruitment process?

Do you feel comfortable with prospect of continuing to work from home for one more year?

Answers:

Yes- 1

No- 2

The answer 'Yes' is labeled as '1' and 'No' is labeled as '2'. The number assigned has no meaning or value, but for the mere identification. Even assigning '1' for 'No' and '2' for 'Yes' make no effect on the answers given by the respondents.

2.12.2 ORDINAL SCALE:

The ordinal scale is used for ranking purpose in research studies. The ordinal scale helps in placing events or objects or characteristics in rank order. This scale only allows ranking of elements in highest to lowest or vice versa and there be no equal space between any two ranks. Ranking based on scores of students in exam in one good example, whereas the difference between 1st rank score and 2nd rank score is not necessarily the same as difference between successive ranks. The ranks here only

suggest 'greater than' or 'less than' characteristic rather than stating how much greater or lower than the other element. Measures of central tendency is the median.

Characteristics:

1. It ranks the objects from highest to lowest or vice versa.
2. It implies only 'greater than' or 'less than' characteristics and not how much higher or lesser among the other elements.
3. The difference between adjacent ranks need not be equal always.

Example:

1. Ranking of designation of teachers in an University such as Professor, Associate Professor, Assistant Professor, Lecturers.
2. Survey question asking customers to rank a list of cold beverages in order of theirpreference.

The difference between any two rank orders is not unknown. It is rank ordered from lowest to highest based on the preference.

2.12.3 INTERVAL SCALE:

Interval scale is more powerful than the nominal and ordinal scale. It uses the principle of 'equality of intervals", that is the intervals are made equal. This measurement scale makes it possible for ordering items with equal intervals. The units are made equal with equal distance. The main limitation in this scale is that it there is no true zero. It means that it does not have the capacity to measure the complete absence of a characteristic.

For example, zero degrees Celsius temperature of water in a bowl doesn't mean there in no temperature in the water. Zero degree Celsius is the freezing point of water and the zero here doesn't mean there is absolutely no temperature. The difference in temperature between 30° and 40° is same as the difference between 60° and 70°, but we cannot say that the temperature of 100° is twice as warm as 50°. Adding two bowls of 50°

Celsius water will not produce 100° Celsius as result of mixing. Measures of central tendency are Mean.

Characteristics:

1) Interval scales have no absolute zero.
2) Scale is based on the principle of equality of intervals.
3) Units are equidistant and quantifiable.
4) Scale does not have the capacity to measure the complete absence of a trait orcharacteristic.

Example:

1. How likely is it that you would recommend this product to your family or friends?

1	2	3	4	5	6	7	8	9	10

2. Customer satisfaction is key to organizational success.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2.12.4 RATIO SCALE:

Ratio scale has the absolute zero point. In this scale, height, weight and times are measured. Zero kilograms of weight means absence of weight. All statistical techniques are usable in ratio scale. Characteristics and benefits of the above scales are present in this scale. Measures of central tendency are geometric and harmonic means.

Characteristics:

1. This scale has an absolute zero measurement.
2. All statistical techniques can be used.

Example:

Sales of a product for this year are twice the sales of the same product last year. All statistical operations can be performed on this scale.

2.13 SAMPLING METHODS

The **sampling method** or **sampling technique** is the process of studying the population by gathering information and analyzing that data. It is the basis of the data where the sample space is enormous.

There are several different sampling techniques available, and they can be subdivided into two groups. All these methods of sampling may involve specifically targeting hard or approach to reach groups.

2.13.1 TYPES OF SAMPLING METHOD

In Statistics, there are different sampling techniques available to get relevant results from the population. The two different types of sampling methods are::

- Probability Sampling
- Non-probability Sampling

2.13.2 PROBABILITY SAMPLING

The probability sampling method utilizes some form of random selection. In this method, all the eligible individuals have a chance of selecting the sample from the whole sample space. This method is more time consuming and expensive than the non-probability sampling method. The benefit of using probability sampling is that it guarantees the sample that should be the representative of the population.

2.13.2.1 PROBABILITY SAMPLING TYPES

Probability Sampling methods are further classified into different types, such as simple random sampling, systematic sampling, stratified sampling, and clustered sampling. Let us discuss the different types of probability sampling methods along with illustrative examples here in detail.

2.13.2.2 SIMPLE RANDOM SAMPLING

In simple random sampling technique, every item in the population has an equal and likely chance of being selected in the sample. Since the item selection entirely depends on the chance, this method is known as "**Method of chance Selection**". As the sample size is large, and the item is chosen randomly, it is known as "**Representative Sampling**".

Benefits of Simple Random Sampling:

- Participants have an equal and fair chance of being selected. As the selection method used gives every participant a fair chance, the resulting sample is unbiased and unaffected by the research team. It is perfect for blind experiments.
- This technique also provides randomized results from a larger pool. The resulting smaller sample should be representative of the entire population of participants, meaning no further segmenting is needed to refine groups down.
- Lastly, this method is cheap, quick, and easy to carry out – great when you want to get your research project started quickly.

Example:

Suppose we want to select a simple random sample of 200 students from a school. Here, we can assign a number to every student in the school database from 1 to 500 and use a random number generator to select a sample of 200 numbers.

2.13.2.3 SYSTEMATIC SAMPLING

In the systematic sampling method, the items are selected from the target population by selecting the random selection point and selecting the other methods after a fixed sample interval. It is calculated by dividing the total population size by the desired population size.

Systematic sampling can be used when researchers want to ensure that the sample is evenly spread across the entire population. For example, a company could select every nth person from the company directory filtered by last name. Other forms of sampling may accidentally cluster similar populations (i.e. too many people from finance are selected based on how the sample is aggregated).

Advantages of Systematic Sampling

- It's simple and convenient to use when creating and analyzing samples.
- The method offers a faster and simpler way to represent a large population without having to identify each member.
- It avoids bias because member selection happens through the use of a sampling interval.
- There's less opportunity to manipulate the data because of the use of a fixed sampling interval in sample selection, which removes the influence of the researchers.
- Systematic sampling enables researchers to come up with an evenly distributed data set, resulting in a more authentic and reliable representation of the overall population.

Example:

Suppose the names of 300 students of a school are sorted in the reverse alphabetical order. To select a sample in a systematic sampling method, we have to choose some 15 students by randomly selecting a starting number, say 5. From number 5 onwards, will select every 15th person from the sorted list. Finally, we can end up with a sample of some students.

2.13.2.4 STRATIFIED SAMPLING

In a stratified sampling method, the total population is divided into smaller groups to complete the sampling process. The small group is formed based on a few characteristics in the population. After separating the population into a smaller group, the statisticians randomly select the sample.

Stratification gives a smaller error in estimation and greater precision than the simple random sampling method. The greater the differences among the strata, the greater the gain in precision.

Example:

There are three bags (A, B and C), each with different balls. Bag A has 50 balls, bag B has 100 balls, and bag C has 200 balls. We have to choose a sample of balls from each bag proportionally. Suppose 5 balls from bag A, 10 balls from bag B and 20 balls from bag C.

Advantages of Stratified Sampling

- **Increased Precision**

One of the primary advantages of stratified sampling is its ability to increase the precision of estimates by ensuring representation from all subgroups or strata within the population. This precision leads to more accurate and reliable results compared to simple random sampling, especially when there are significant differences between subgroups.

- **Reduced Sampling Bias**

Stratified sampling helps mitigate various biases, by ensuring that each subgroup of the population is adequately represented in the sample. This reduces the risk of over-representing or under-representing certain segments of the population, which can skew the results and lead to erroneous conclusions. By sampling proportionally from each stratum, researchers can obtain a more balanced and representative

sample, thereby minimizing bias and increasing the validity of the findings.

- **Enhanced Generalizability**

Because stratified sampling ensures representation from all subgroups within the population, the results are often more generalizable or applicable to the entire population. By capturing the diversity of characteristics and perspectives across different strata, researchers can draw conclusions that are more robust and applicable to a broader range of individuals or entities. This enhanced generalizability makes the findings from stratified sampling more valuable for informing decision-making and guiding actions within the population of interest.

- **Detection of Subgroup Differences**

Another advantage of stratified sampling is its ability to detect differences or patterns within specific subgroups of the population. By analyzing the data separately for each stratum, researchers can identify unique trends, preferences, or behaviors that may exist within certain demographic or behavioral segments. This granularity allows for a deeper understanding of the population dynamics and can inform targeted interventions or strategies tailored to the needs of different subgroups.

- **Efficient Resource Allocation**

Stratified sampling can also lead to more efficient resource allocation by focusing data collection efforts on the most relevant subgroups or strata within the population. Instead of using resources indiscriminately across the entire population, researchers can prioritize areas of interest or importance based on the stratification variables. This targeted approach not only saves time and resources but also maximizes the utility of the data collected, resulting in a more cost-effective research process.

2.13.2.5 CLUSTER SAMPLING

In the cluster sampling method, the cluster or group of people are formed from the population set. The group has similar significant characteristics. Also, they have an equal chance of being a part of the sample. This method uses simple random sampling for the cluster of population.

In this sampling technique, researchers analyze a sample that consists of multiple sample parameters such as demographics, habits, background — or any other population attribute, which may be the focus of conducted research. This method is usually conducted when groups that are similar yet internally diverse form a statistical population. Instead of selecting the entire population, cluster sampling allows the researchers to collect data by bifurcating the data into small, more productive groups.

Example:

An educational institution has ten branches across the country with almost the number of students. If we want to collect some data regarding facilities and other things, we can't travel to every unit to collect the required data. Hence, we can use random sampling to select three or four branches as clusters.

Advantages of Cluster Sampling

- **Cost-effectiveness:**

  While it can differ based on the type of cluster sampling you use (one-stage or two-stage), cluster sampling is generally a cost-efficient sampling process. It allows you to gather responses from a certain niche audience without having to pay for the whole sample to come from that audience (which can be expensive, depending on their criteria). Sample cost will of course depend on the size and distribution of your sample population, the number and size of selected clusters, and the sampling method used.

- **Efficiency and Speed:**

It can be very time-consuming to identify and capture every individual that makes up your desired population of respondents. Cluster sampling speeds up this process by creating mini-representations of each subgroup you want to collect data from. So, instead of having to survey all college-level students in the US, perhaps you survey a 'cluster' of college students from each state. As you can imagine, this is a much faster approach that still provides trustworthy and representative results.

- **Natural groupings:**

As alluded to above with our college student example, cluster sampling can be based on already naturally formed groupings in the population (i.e. neighborhoods, school districts, etc.). These natural groupings represent what the population at large tends to look like and reduces potential sampling bias by ensuring each cluster has some form of homogeneity that makes them one representative piece of the whole picture.

2.13.3 NON-PROBABILITY SAMPLING

The non-probability sampling method is a technique in which the researcher selects the sample based on subjective judgment rather than the random selection. In this method, not all the members of the population have a chance to participate in the study.

2.13.3.1 NON-PROBABILITY SAMPLING TYPES

Non-probability Sampling methods are further classified into different types, such as convenience sampling, consecutive sampling, quota sampling, judgmental sampling, snowball sampling. Here, let us discuss all these types of non-probability sampling in detail.

2.13.3.2 CONVENIENCE SAMPLING

In a convenience sampling method, the samples are selected from the population directly because they are conveniently available for the researcher. The samples are easy to select, and the researcher did not choose the sample that outlines the entire population. This method is applied when there is a large population or infinite population.

Example:

In researching customer support services in a particular region, we ask your few customers to complete a survey on the products after the purchase. This is a convenient way to collect data. Still, as we only surveyed customers taking the same product. At the same time, the sample is not representative of all the customers in that area.

Advantages of Convenience Sampling

- **Data collection is easier and anyone can do the research:** You don't need to be trained or have a lot of experience to do data collection for convenience sampling. By creating a survey that helps you collect information in quantitative ways, can allow you to quickly analyse trends. In addition, having a smaller sample will save time going through lots of raw data.
- **Quick and low cost:** Being able to conduct your research at speed and with very little cost involved can make this sampling method preferable to investing in full-scale research projects.
- **Great for initial research:** If you're not sure about the thoughts, beliefs, and values of your target audience, or if you just want to do a small-scale initial survey, convenience sampling is a great way to go about it.
- **Pilots can be quicker:** Using convenience sampling for pilot data collection can give you the information your managers need to make decisions quickly with the right data.
- **Fewer rules make easily accessible participants:** Since the sample is made up of present and willing people that are convenient to approach, the process becomes faster and convenient overall for the research carrying it out.

- **Ease of future participants:** If more participants are needed at a later date to create multiple samples in future research (to provide more information results over time or to try and replicate results), this is easier as there are no criteria to assess for.

Disadvantages of Convenience Sampling

- **Sampling bias:** As the sample is based on people who are willing at the time and place that the researcher is present, you won't be gaining a range of people each time you're collecting data. In addition, the research subjectively chooses people to ask if they would like to be a part of the research, so this could influence the final sample as well.
- **Selection bias:** Many researchers might point out that having a convenience sample may end up excluding demographic subsets from the results. Also, the volunteer nature of the participation means that people who are inclined to know about the subject or pro-topic may appear more represented in the data.
- **Unable to generalise data:** As the sample will be unrepresentative of the total population, you will find it hard to generalise about the population as a whole.
- **Low external validity:** If you do base research only on convenience sampling without replicating results or adding in an additional probability-based sampling method, your research findings might lack credibility within the wider research industry.
- **Positivity bias:** You may end up having a positivity bias if the people you recruit are too close to you personally and know you want certain results, while people from your workplace may want to please the researcher in general.

2.13.3.3 QUOTA SAMPLING

In the quota sampling method, the researcher forms a sample that involves the individuals to represent the population based on specific traits or qualities. The researcher chooses the sample subsets that bring the useful collection of data that generalizes the entire population.

Uses of Quota Sampling

Some of the essential uses of quota sampling are as follows:

- In quota sampling, the judgment of sampling is done on the basis of convenience or a fixed quota. It is used when the decision of the researcher is final for choosing the samples
- The quota sampling is used because it allows the researchers to select a subgroup as a sample that is of great interest to study
- This method is used when the researcher wants to determine the relationship between the subgroups
- The quota sampling method should be selected when there is a limited time

Quota Sampling Example

Let us assume that we need to know about the career goals of university students. More particularly, the differences in the career goals among fresher, juniors and seniors are to be examined. Suppose the concerned university contains 10,000 students and can be taken as our population.

Now, we have to divide our population of 10,000 students into categories such as freshers, juniors and seniors. Suppose we find that there are 3000 freshers (30%), 2500 junior students (25%) and 2000 senior students (20%). Our sample must have these proportions. It means that if we sample 1000 students, then we must consider 300 freshers, 250 juniors and 200 seniors. Lastly, we may start collecting samples from these students based on our proportion.

2.13.3.4 PURPOSIVE OR JUDGEMENTAL SAMPLING

In purposive sampling, the samples are selected only based on the researcher's knowledge. As their knowledge is instrumental in creating the samples, there are the chances of obtaining highly accurate answers with a minimum marginal error. It is also known as judgemental sampling or authoritative sampling.

Example:

Consider a scenario where a panel decides to understand what are the factors which lead a person to select ethical hacking as a profession. Ethical hacking is a skill which has been recently attracting youth. More and more people are selecting it as a profession. The researchers who understand what ethical hacking is will be able to decide who should form the sample to learn about it as a profession. That is when judgemental sampling is implemented. Researchers can easily filter out those participants who can be eligible to be a part of the research sample.

Advantages of Judgemental sampling

- **Consumes minimum time for execution:** In this sampling bias approach, researcher expertise is important and there are no other barriers involved due to which selecting a sample becomes extremely convenient.
- **Allows researchers to approach their target market directly:** There are no criteria involved in selecting a sample except for the researcher's preferences. Due to this, he/she can communicate directly with the target audience of their choice and produce desired results.
- **Almost real-time results:** A quick poll or survey can be conducted with the sample using judgemental sampling since the members of the sample will possess appropriate knowledge and understanding of the subject.

2.13.3.5 SNOWBALL SAMPLING

Snowball sampling or chain-referral sampling is defined as a non-probability sampling technique in which the samples have rare traits. This is a sampling technique, in which existing subjects provide referrals to recruit samples required for a research study.

For example, if you are studying the level of customer satisfaction among the members of an elite country club, you will find it extremely difficult to collect primary

data sources unless a member of the club agrees to have a direct conversation with you and provides the contact details of the other members of the club.

This sampling method involves a primary data source nominating other potential data sources that will be able to participate in the research studies. Snowball sampling method is purely based on referrals and that is how a researcher is able to generate a sample. Therefore, this method is also called the chain-referral sampling method.

Snowball sampling is a popular business study method. The snowball sampling method is extensively used where a population is unknown and rare and it is tough to choose subjects to assemble them as samples for research.

Applications of Snowball Sampling:

- **No official list of names of the members:** This sampling technique can be used for a population, where there is no easily available data liketheir demographic information. For example, homeless or list of members of anelite club, whose personal details cannot be obtained easily.
- **Difficulty to locate people:** People with rare diseases are quite difficult to locate. However, if a researcher is carrying out a research study similar in nature, finding the primary data source can be a challenge. Once he/she is identified, they usually have information about more such similar individuals.
- **People who are not willing to be identified:** If a researcher is carrying out a study which involves collecting information/data from sex workers or victims of sexual assault or individuals who don't want to disclose their sexual orientations, these individuals will fall under this category.
- **Secretiveness about their identity:** People who belong to a cult or are religious extremists or hackers usually fall under this category. A researcher will have to use snowball sampling to identify these individuals and extract information from them.

Advantages of Snowball Sampling

1. **It's quicker to find samples:** Referrals make it easy and quick to find subjects as they come from reliable sources. An additional task is saved for a researcher, this time can be used in conducting the study.
2. **Cost effective:** This method is cost effective as the referrals are obtained from a primary data source. It's is convenient and not so expensive as compared to other methods.
3. **Sample hesitant subjects:** Some people do not want to come forward and participate in research studies, because they don't want their identity to be exposed. Snowball sampling helps for this situation as they ask for a reference from people known to each other. There are some sections of the target population which are hard to contact. For example, if a researcher intends to understand the difficulties faced by HIV patients, other sampling methods will not be able to provide these sensitive samples. In snowball sampling, researchers can closely examine and filter members of a population infected by HIV and conduct aresearch by talking to them, making them understand the research objective, andeventually, analyzing the received feedback.

2.13.4 PROBABILITY SAMPLING VS NON-PROBABILITY SAMPLING METHODS

Probability Sampling Methods	Non-probability Sampling Methods
Probability Sampling is a sampling technique in which samples taken from a larger population are chosen based on probability theory.	Non-probability sampling method is a technique in which the researcher chooses samples based on subjective judgment, preferably random selection.
These are also known as Random sampling methods.	These are also called non-random sampling methods.
These are used for research which is conclusive.	These are used for research which is exploratory.

These involve a long time to get the data.	These are easy ways to collect the data quickly.
There is an underlying hypothesis in probability sampling before the study starts. Also, the objective of this method is to validate the defined hypothesis.	The hypothesis is derived later by conducting the research study in the case of non-probability sampling.

Let's Sum Up

In this unit we have studied the meaning, objectives of research in management, research process and its stages and types of research design. We have understood the data types, methods of primary data collection, the merits and demerit of different primary data collection methods and sources of secondary data. We also studied the method of constructing a questionnaire for data collection and measuring the validity and reliability of the questionnaire. Further we studied the importance of hypothesis and types of hypotheses. The meaning of scale construction in research methodology, different types of scales, types of attitude measurement scales and their characteristics have been discussed in this unit. Finally, the probability and non-probability sampling techniques have been exhibited in this unit.

Check your Progress – Quiz II

1. A statement made about a population for testing purpose is called_____

A. Statistic

B. Hypothesis

C. Level of Significance

D. Test-Statistic

2. Which of the following is the main point of difference between primary data and secondary data?

A. The collection of secondary data is costlier and more time consuming whencompared to primary data
B. The secondary data is always original whereas the primary data is not
C. The primary data is much more reliable than secondary data because it iscollected directly from respondents
D. None of the above

3. What is the meaning of cluster sampling?

A. It is a process where the sampling universe is divided into multiple groups
B. It is a process where the samples for a study is obtained through consciousselection
C. It is a process where the samples for a study are selected at regular intervals
D. It is a process through which the sample for a study is divided into multiplegroups

4. The limitations of a personal interview include _____

A. Personal biases
B. Lack of space and time
C. Both of them
D. None of them

5. _____Scale is used for labelling variables into distinct classifications

A. Nominal
B. Ordinal
C. Ratio
D. Interval

6. _____is a numerical scale, where the order of the variable is known as well as the difference between these variables

A. Nominal
B. Ordinal
C. Ratio
D. Interval

7. _____Scaling techniques used for the purpose of categorisation or division

A. Interval
B. Nominal
C. Ordinal

D. Ratio

8. Of the following sampling methods,_____is a probability method.

A. Judgement

B. Quota

C. Simple random

D. Convenience

9. When the available population is_____, we use a stratified sample.

A. too small

B. very large

C. homogeneous

D. heterogeneous

10. Snowball sampling comes under the category of_____.

A. random sampling

B. probability sampling

C. quota sampling

D. nonprobability sampling

UNIT SUMMARY:

Research methods involve systematic approaches to investigate problems, encompassing the research process, design, and data collection methods. Research designs, such as exploratory, descriptive, and experimental, help structure investigations, while theories guide the formulation of hypotheses and objectives. Data is collected through surveys, observations, or experiments, with careful attention to validity, reliability, and appropriate scaling. Sampling techniques (probability and non-probability) ensure representative data, and determining optimal sample size is crucial for accurate results.

Glossary

Glossary	MEANING
Research Design	The framework or plan for conducting research to address specific problems or questions.
Hypothesis	A testable prediction about the relationship between variables in a study.
Primary Data Collection	The process of gathering new, original data directly from sources.
Validity and Reliability	Validity measures accuracy; reliability measures consistency of research instruments.
Sampling Techniques	Methods used to select a subset from a population for study, either randomly or non-randomly.

Self-Assessment Questions

1. Discuss the research process with its steps and examples
2. Enumerate the different scaling techniques applied in social science research
3. List and explain the probability sampling techniques with their advantages
4. Describe the non-probability sampling techniques with their advantages
5. Write the different methods of primary data collection with their merits and demerits

Activities Related to Research Methods

1. Assume of doing research for identifying the employees feedback on welfare measures in an organization. Develop a research process and the research design for this study
2. Write the sampling methods to be adopted with justification for the following:
   a. Survey for Customer feedback of an electric two-wheeler

b. Opinion poll for a general election

c. Employees feedback on job and salary

ANSWERS FOR CHECK YOUR PROGRESS – QUIZ 1

1) **C. Identifying a problem.**
2) **D. All the above.**
3) **A. Applied research.**
4) **D. Exploratory research is both structured and informal.**
5) **B. Describing characteristics and traits of a phenomenon.**

ANSWERS FOR CHECK YOUR PROGRESS – QUIZ 2

1. **B. Hypothesis**
2. **C. The primary data is much more reliable than secondary data because it is collected directly from respondents**
3. **A. It is a process where the sampling universe is divided into multiple groups**
4. **B. Both of them**
5. **A. Nominal**
6. **D. Interval**
7. **B. Nominal**
8. **C. Simple random**
9. **D. Heterogeneous**
10. **D. Nonprobability sampling**

Suggested Readings:

1. Cooper, D.R., Schindler, P. and Sharma, J.K., Business Research Methods,11th Edition, Tata-McGraw Hill, 12 th Edition, 2018.
2. Ranjith Kumar, Research Methodology: A Step-by-Step guide for Beginners, Sage, South Asia, 4th Edition, 2023.
3. Kothari C.R. & Gaurav Garg, Research Methodology: Methods and Techniques, New Age International Publishers, 2023

UNIT III

DATA PREPARATION AND ANALYSIS

Unit III: Data Preparation and Analysis
Data Preparation - Editing –Coding- Data Entry- Data Analysis- Testing Of Hypothesis Univariate and Bivariate Analysis Parametric And Nonparametric Tests and Interpretation of Test Results- Chi-Square Test- Correlation; Karl Pearson's Vs Correlation Coefficient and Spearman's Rank Correlation- Regression Analysis - One Way and Two Way Analysis Of Variance.

Data Preparation and Analysis

UNIT OBJECTIVES:

- To learn the concepts of data editing, coding and data entry process
- To enumerate the meaning of data analysis and its types

- To bring out the procedure of hypothesis testing and methods of parametric and non-parametric tests
- To exhibit the methodology of chi-square test with suitable examples
- To understand the concepts of correlation and regression with solved illustrations
- To study the meaning and applications of ANOVA

3.1 INTRODUCTION OF DATA AND DATA PROCESSING

The data which is collected for the purpose of the study itself cannot reveal everything. This being a raw data, it is required to process and analyze in order to have desired result. The data which is collected cannot be directly used for making analysis. Before analysis, data is required to be processed.

Data processing is an intermediate stage between collection of data and their analysis and interpretation, which include Checking, Editing, Coding and Tabulation. Data processing is a crucial stage in research. After collecting the data

certain conclusions which may confirm or invalidate the hypothesis which he had formulated towards the beginning of research worth. The mass of data collected during the field work is to be processed with a view to reducing them to manageable proportions. Only by such a careful and systematic processing, the data will lend itself for statistical treatment and meaningful interpretation and conclusion.

The processing of data includes editing, coding, classification and tabulation. The collected data should be organized in such a way so that table charts can be prepared for presentation. The processing of data is necessary because, the data

collected should be examined and errors and mistakes are rectified so that at the stage of analysis of data, no difficulty is experienced. Various steps involved in processing of data are Editing, Coding, Classification and Tabulation.

3.2 DATA CODING

Data coding is the process of converting data into a form that can be analyzed. It involves assigning numerical or categorical codes to data items, such as responses to survey questions or demographic information. Coded data can then be analyzed using statistical software or other tools. Some common types of data coding include:

3.2.1 TYPES OF DATA CODING

1. **Nominal coding:** This involves assigning labels or categories to data items. For example, responses to a survey question about marital status might be coded as follows: Single 1, Married 2, Divorced 3, Widowed 4.
2. **Ordinal coding:** This involves assigning categories to data items in a specific order. For example, responses to a survey question about satisfaction level might be coded as follows: Very dissatisfied = 1, Frustrated = 2, Neutral = 3, Satisfied = 4, and Very satisfied = 5.
3. **Dichotomous coding:** This involves assigning a binary code (e.g., 0 or 1) to data items. For example, responses to a survey question about gender might be coded as follows: Male = 0, Female = 1.
4. **Numeric coding:** This involves assigning numerical values to data items. For example, responses to a survey question about age might be coded as follows: 18-24 years old = 1, 25-34 years old = 2, 35-44 years old = 3, and so on.
5. **Derived variables:** This involves calculating new variables based on existing data. For example, a researcher might calculate the mean score for a set of survey questions or create a new variable based on the sum of several other variables.
6. **Truncation:** This involves removing part of a data item. For example, a researcher might truncate a part of the variable (e.g. recording the value as

23, 47 etc., for measurement values 12.23, 12.47 etc.). This could be helpful when the analysis is being performed manually.

Data coding is an important step in data analysis, as it allows researchers to make sense of large amounts of data and draw meaningful conclusions. By assigning codes to data items, researchers can quickly identify patterns and trends that would otherwise be difficult to detect.

3.3 DATA EDITING

3.3.1 EDITING

Editing means to rectify or to set to order or to correct or to establish sequence. Editing is the process of examining the data collected in questionnaire or interview schedule to deduct errors and omissions and to correct those if possible. When the whole data collection is over, a final and thorough check up is made for data processing. It is better if the data collected is verified even before the data analysis is carried out. In this process editing is the first step. Editing is done to assure that the collected data are accurate, consistent with other facts gathered uniformly entered and as complete as possible. For example, imagine if we get the newspaper unedited, how the news will appear? Similarly, an unedited filmwill have no sequence of events, which means the story cannot be understood at all.

3.3.2 TYPES

Editing is performed at two stages and depending on that it could be of two types:

- Field editing
- Central editing / Centralized editing

Field editing

Field editing is the process for completing the information recorded in abbreviated or in illegible form at the time of recording the respondent's response. This sort of editing should be carried out as soon as possible after interview. In field editing completeness of the forms should be checked by person. It may be possible that the investigator might have forgotten to record the information. If investigator

recorded information is incomplete form using abbreviations than it should be completed.

Central editing / Centralized editing

Central editing is done on the return to the office after completing all forms of schedule. This sort of editing is performed by single editor or by a team of editors. The editors are free to correct the obvious errors such as an entry in the wrong place, entry recorded in different units and the like. At central level, editors must correct various mistakes of the investigator. In case of gap in the answers the editor will be required to decide the proper answer to meet out the gap in answer. This can be done by reviewing the other information in questionnaire. Some times in spite of all efforts, if correctness of the answer is impossible than it is safe to strike out such wrong answers. All the wrong answers should be dropped by the editors.

3.3.3 SIGNIFICANCE OF EDITING

- It is pre-requisite for accuracy.
- It is useful in the elimination of the incorrect reply.
- It ensures the consistency of the data collected and avoids contradictions.
- It is useful to convert the answers into uniform units of measurement before coding

3.4 DATA TABULATION

The process of placing classified data into tabular form is known as tabulation. A table is a symmetric arrangement of statistical data in rows and columns. Rows are horizontal arrangements whereas columns are vertical arrangements. It may be simple, double or complex depending upon the type of classification. Tabulation is a systematic & logical presentation of numeric data in rows and columns to facilitate comparison and statistical analysis. It facilitates comparison by bringing related information close to each other and helps in further statistical analysis and interpretation.

3.4.1 TABULATION

Tabulation is the process of summarizing raw data and displaying in compact form of vertical columns and horizontal rows of numbers for further analysis. Analysis of data is made possible through tables. Tabulation may be done manually or mechanically or electronically.

Tabulation is the process of presenting in an orderly manner of the classified data in a table. In other words, it is a method of presenting the summarized data. Tabulation is very important because,

- Its helps to conserve space
- It avoids any need for explanation
- Computation of the data is made easier
- Comparison of data becomes very simple
- Adequacy or inadequacy of the data is clearly visible.

A table contains columns and rows. These columns and rows create small boxes which are called cells. Tables are classified as

- One-way table
- Two-way table and
- Multi-way table

Through one-way table, we would get only one information, while the two-way table can provide minimum two information's and the multi-way table could provide a number of information. In case of large number of items in big study, project strip method is used. Long strip used in this method, where 160 columns and 320 items can be recorded in one strip.

3.4.2 CLASSIFICATION OF TABULATION

Tabulation can be classified as

A. Simple tabulation

B. Complex tabulation

Simple tabulation

It gives information about one or more groups of independent questions. This results, in one way table, provides information of one characteristics of data.

Complex tabulation

In this type of tabulation, the data is divided in two or more categories which gives information regarding more sets of inter-related question. It results in two-way or three-way tables which gives information about several inter related characteristics of data. This complex table is described as cross tabulation.

3.4.3 COMPONENTS OF TABLE

There are set rules for tabulation but due care should be taken on following aspect for constructing table.

1. Each table should have clear number for the purpose of the reference.
2. Every table should also have suitable title and this title should be self-explanatory.
3. There should be proper heading to each column and raw of the table in brief.
4. The body of the table contains the numerical information. Data presented in the body is arranged as per the description.
5. The unit of measurement is frequently written as headnote such as in ‘000’ (in thousand) or million (i.e. 10 lakhs) or Cr. (i.e. crores).

3.4.4 RULES FOR TABULATION

1. Captions and stubs should be arranged in systematic order.
2. Measurement should be clearly defined.
3. Avoid overloading the table with details.
4. Table should be logically arranged.
5. Avoid the use of abbreviations.

6. Expression like etc. is not good in table.

7. Dittoes mark should not be used in table.

8. Not available letter should be used for information not available or even use of (dash) can be made for these explanation.

9. Miscellaneous items should be placed in the last row of the table.

In order to facilitate comparison, the arrangement of the categories should be made according to alphabetical or chorological order.

10.Table must be suitable to the requirements of the study.

3.5 DATA ENTRY

Data entry refers to the process of manually inputting or transferring data from physical or digital sources into electronic formats. It is a fundamental aspect of data management, ensuring that information is accurately recorded, organised, and stored for future use. Accurate and well-organised data is essential for informed decision-making, analysis, reporting, and maintaining the integrity of business records.

3.5.1 METHODS OF DATA ENTRY

- **Manual Data Entry**: This method involves individuals manually entering data using keyboards, keypads, or touchscreens. It is suitable for small-scale data entry tasks or situations where data is received in physical formats like paper documents or forms.
- **Online Data Entry**: Online data entry involves inputting data directly into online forms or systems. It is commonly used for tasks such as online surveys, customer registration, or e-commerce transactions.
- **Automated Data Entry**: Automated data entry relies on technology, such as optical character recognition (OCR) or intelligent character recognition (ICR), to extract data from scanned documents, images, or machine-readable formats. Automated methods help reduce manual errors and speed up the data entry process.

3.5.2 CHALLENGES IN DATA ENTRY

- **Accuracy and Quality Control**: Maintaining accuracy in data entry can be challenging due to human errors, such as typos, omissions, ormisinterpretation of handwritten or poorly scanned documents. Implementing quality control measures, such as double-checking entries, validation rules, and data cleaning processes, can help mitigate these challenges.
- **Data Volume and Time Constraints**: Large volumes of data can overwhelm data entry personnel, leading to potential errors or delays. Adequate resources, efficient workflows, and time management strategies are crucial to ensure timely and accurate data entry.
- **Data Security and Confidentiality**: Data entry involves handling sensitive information, such as personal or financial data. Implementing data security measures, such as secure data transmission, access controls, and data encryption, is vital to protect confidentiality and comply with data protection regulations.

3.5.3 BEST PRACTICES FOR DATA ENTRY

- **Standardised Data Entry Procedures**: Establishing standardised procedures and guidelines for data entry ensures consistency and accuracy. This includes defining data formats, naming conventions, validation rules, and data entry protocols.
- **Training and Skill Development**: Providing comprehensive training to data entry personnel on proper data entry techniques, tools, and quality control processes is essential. Regular skill development programs can help improve efficiency and accuracy.
- **Automation and Technology Integration**: Leveraging technology, such as OCR, ICR, or data entry software, can automate data entry processes and minimise manual errors. Integrating data entry systems with other business applications or databases ensures seamless data flow and reduces duplication of effort.

- **Quality Control Measures**: Implementing quality control checks, such as double-entry verification, validation rules, and data cleaning processes, helps identify and rectify errors before they impact business operations or decision-making.
- **Regular Data Backup and Maintenance**: Regular data backup and maintenance practices ensure data integrity and protect against data loss. Implementing data archiving and version control strategies is crucial for long-term data management.

3.6 DATA ANALYSIS

Data analysis is the process of collecting, modelling, and analyzing data to extract insights that support decision-making. There are several methods and techniques to perform analysis depending on the industry and the aim of the investigation.

3.6.1 IMPORTANCE OF DATA ANALYSIS

Before we go into detail about the categories of analysis along with its methods and techniques, you must understand the potential that analyzing data can bring to your organization.

- **Informed decision-making:** From a management perspective, you can benefit from analyzing your data as it helps you make decisions based onfacts and not simple intuition. For instance, you can understand where to invest your capital, detect growth opportunities, predict your incomes, or tackle uncommon situations before they become problems. Like this, you can extract relevant insights from all areas in your organization, and with the help of dashboard software, present the information in a professional and interactive way to different stakeholders.
- **Reduce costs:** Another great benefit is to reduce costs. With the help of advanced technologies such as predictive analytics, businesses can spot improvement opportunities, trends, and patterns in their data and plan their strategies accordingly. In time, this will help you save money and resources on implementing the wrong strategies. And not just that, by predicting different

scenarios such as sales and demand you can also anticipate production and supply.

- **Target customers better:** Customers are arguably the most crucial element in any business. By using analytics to get a 360° vision of all aspects related to your customers, you can understand which channels they use to communicate with you, their demographics, interests, habits, purchasing behaviors, and more. In the long run, it will drive success to your marketing strategies, allow you to identify new potential customers, and avoid wasting resources on targeting the wrong people or sending the wrong message. You can also track customer satisfaction by analyzing your client's reviews or your customer service department's performance.

3.6.2 UNIVARIATE ANALYSIS:

Univariate analysis involves the examination of a single variable (or feature) at a time. It is the simplest form of data analysis and is often used to describe the characteristics of individual variables.

Purpose: Univariate analysis is primarily used to summarize and visualize the distribution of a single variable, assess its central tendency (mean, median, mode), dispersion (range, variance, standard deviation), and shape (e.g., histogram, box plot). It helps in understanding the characteristics of a single variable in isolation.

3.6.3 BIVARIATE ANALYSIS:

Bivariate analysis involves the analysis of two variables simultaneously to determine if there is a relationship or association between them. It explores how changes in one variable are related to changes in another.

Purpose: Bivariate analysis is used to understand the relationship, correlation, or association between two variables. Common techniques for bivariate analysis include scatter plots, correlation coefficients (e.g., Pearson's correlation), and contingency tables (for categorical variables). It helps answer questions like, "Is there a relationship between a person's age and their income?

3.6 HYPOTHESIS TESTING

Hypothesis testing in statistics refers to analyzing an assumption about a population parameter. It is used to make an educated guess about an assumption using statistics. With the use of sample data, hypothesis testing makes an assumption about how true the assumption is for the entire population from wherethe sample is being taken.

Any hypothetical statement we make may or may not be valid, and it is then our responsibility to provide evidence for its possibility. To approach any hypothesis, we follow these four simple steps that test its validity.

1. First, we formulate two hypothetical statements such that only one of them is true. By doing so, we can check the validity of our own hypothesis.
2. The next step is to formulate the statistical analysis to be followed based upon the data points.
3. Then we analyze the given data using our methodology.
4. The final step is to analyze the result and judge whether the null hypothesis will be rejected or is true.

Often after formulating research statements, the validity of those statementsneeds to be verified. Hypothesis testing offers a statistical approach to the researcher about the theoretical assumptions he/she made. It can be understood as quantitative results for a qualitative problem.

Hypothesis testing provides various techniques to test the hypothesis statement depending upon the variable and the data points. It finds its use in almost every field of research while answering statements such as whether this new medicine will work, a new testing method is appropriate, or if the outcomes of a random experiment are probable or not.

3.6.1 BASIC CONCEPTS CONCERNING TESTING OF HYPOTHESES

1. **The level of significance:** This is a very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should be chosen with great care, thought and reason. In case we take the significance

level at 5 per cent, then this implies that H0 will be rejected when

the sampling result (i.e., observed evidence) has a less than 0.05 probability of occurring if H0 is true. In other words, the 5 per cent level of significance means that researcher is willing to take as much as a 5 per cent risk of rejecting the null hypothesis when it (H0) happens to be true. Thus, the significance level is the maximum value of the probability of rejecting H0 whenit is true and is usually determined in advance before testing the hypothesis.

2. **Decision rule or Test of Hypothesis:** A decision rule is a procedure that the researcher uses to decide whether to accept or reject the null hypothesis. The decision rule is a statement that tells under what circumstances to reject the null hypothesis. The decision rule is based on specific values of the test statistic (e.g., reject H0 if Calculated value > table value at the same level of significance)

3. **One- tailed and Two-tailed Tests:** A test of statistical hypothesis, where the region of rejection is on only one side of the sampling distribution, is called a one tailed test. For example, suppose the null hypothesis states that the mean is less than or equal to 10. The alternative hypothesis would be that the mean is greater than 10. The region of rejection would consist of a range of numbers located on the right side of sampling distribution i.e., a set of numbers greater than 10.

   A test of statistical hypothesis, where the region of rejection is on both sides of the sampling distribution, is called a two-tailed test. For example, suppose the null hypothesis states that the mean is equal to 10. The alternative hypothesis would be that the mean is less than 10 or greater than 10. The region of rejection would consist of a range of numbers located on both sides of sampling distribution.

   Types of Error:

   In the context of testing of hypotheses, there are basically two types of errors we can make.

a. Type 1 error: To reject the null hypothesis when it is true is to make what is known as a type I error. The level at which a result is declared significant is known as the type I error rate, often denoted by **α.**

b. Type II error: If we do not reject the null hypothesis when in fact there is a difference between the groups, we make what is known as a type II error. The type II error rate is often denoted as **β.**

3.6.2 PROCEDURE OF HYPOTHESIS TESTING

Procedure for hypothesis testing refers to all those steps that we undertake for making a choice between the two actions i.e., rejection and acceptance of a null hypothesis. The various steps involved in hypothesis testing are stated below:

1. Making a formal statement:

The step consists in making a formal statement of the null hypothesis (H0) and also of the alternative hypothesis (Ha or H1). This means that hypotheses should be clearly stated, considering the nature of the research problem.

2. Selecting a significance level:

The hypotheses are tested on a pre-determined level of significance and as such the same should be specified. Generally, in practice, either 5% level or 1% level is adopted for the purpose.

3. Deciding the distribution to use:

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

4.Selecting a random sample and computing an appropriate value:

Another step is to select a random sample(s) and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution. In other words, draw a sample to furnish empirical data.

4. Calculation of the probability:

One has then to calculate the probability that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact true.

5. Comparing the probability and Decision making:

Yet another step consists in comparing the probability thus calculated with the specified value for α, the significance level. If the calculated probability is equal to or smaller than the α value in case of one-tailed test (and α /2 in case of two-tailed test), then reject the null hypothesis (i.e., accept the alternative hypothesis), but if the calculated probability is greater, then accept the null hypothesis.

3.6.3 HYPOTHESIS TESTING EXAMPLE

Based on the type of data collected, we perform a one-tailed *t*-test to test whether men are in fact taller than women. This test gives:

- an estimate of the difference in average height between the two groups.
- a *p*-value showing how likely to see this difference if the null hypothesis of no difference is true.

The *t*-test shows an average height of 175.4 cm for men and an average height of 161.7 cm for women, with an estimate of the true difference ranging from 10.2 cm to infinity. The *p*-value is 0.002.

In our analysis of the difference in average height between men and women, we find that the *p*-value of 0.002 is below the cutoff of 0.05, so we decide to reject the null hypothesis of no difference.

3.6.4 DIFFERENT TESTS OF HYPOTHESES

Hypothesis testing determines the validity of the assumption (technically described as null hypothesis) with a view to choose between two conflicting hypotheses about the value of a population parameter. Hypothesis testing helps to decide on the basis of a sample data, whether a hypothesis about the population is likely to be true or false. Statisticians have developed several tests of hypotheses (also known as the tests of significance) for the purpose of testing of hypotheses which can

be classified as:

a. Parametric tests or standard tests of hypotheses; and

b. Non-parametric tests or distribution-free test of hypotheses.

3.6.5 PARAMETRIC AND NON-PARAMETRIC TESTS

Parametric tests usually assume certain properties of the parent population from which we draw samples. Assumptions like observations come from a normal population, sample size is large, assumptions about the population parameters like mean, variance, etc., must hold good before parametric tests can be used. But there are situations when the researcher cannot or does not want to make such assumptions. In such situations we use statistical methods for testing hypotheses which are called non-parametric tests because such tests do not depend on any assumption about the parameters of the parent population. Besides, most non-parametric tests assume only nominal or ordinal data, whereas parametric tests require measurement equivalent to at least an interval scale.

The important parametric tests are: 1. z-test; 2. t-test; and 3. F-test. All these tests are based on the assumption of normality i.e., the source of data is considered to be normally distributed.

Non-parametric tests are used when the data isn't normal. Therefore, the key is to figure out if you have normally distributed data. The only non-parametric test you are likely to come across in elementary stats is the chi-square test. However, there are several others. For example: the Kruskal Willis test is the non-parametric alternative to the One-way ANOVA and the Mann Whitney is the non- parametric alternative to the two-sample t test.

3.6.6 LIMITATIONS OF THE TEST OF HYPOTHESES

- Test do not explain the reasons as to why does the difference exist, say between the means of the two samples. They simply indicate whether the difference is due to fluctuations of sampling or because of other reasons but the tests do not tell us as to which is/are the other reason(s) causing the difference.

- Results of significance tests are based on probabilities and as such cannot be

expressed with full certainty.

Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the truth of the hypotheses.

3.7 CHI-SQUARE TEST

The Chi-Square test is a statistical procedure for determining the difference between observed and expected data. This test can also be used to determine whether it correlates to the categorical variables in our data. It helps to find out whether a difference between two categorical variables is due to chance or a relationship between them.

3.7.1 DEFINITION

A chi-square test is a statistical test that is used to compare observed and expected results. The goal of this test is to identify whether a disparity between actual and predicted data is due to chance or to a link between the variables under consideration. As a result, the chi-square test is an ideal choice for aiding in our understanding and interpretation of the connection between our two categorical variables.

A chi-square test or comparable nonparametric test is required to test a hypothesis regarding the distribution of a categorical variable. Categorical variables, which indicate categories such as animals or countries, can be nominal or ordinal. They cannot have a normal distribution since they can only have a few particular values.

For example, a meal delivery firm in India wants to investigate the link between gender, geography, and people's food preferences.

3.7.2 FORMULA FOR CHI-SQUARE TEST

$$x_c^2 = \frac{\Sigma\,(O_i - E_i)^2}{E_i}$$

Where,

c = Degrees of freedom

O = Observed Value

E = Expected Value

The degrees of freedom in a statistical calculation represent the number of variables that can vary in a calculation. The degrees of freedom can be calculated to ensure that chi-square tests are statistically valid. These tests are frequently used to compare observed data with data that would be expected to be obtained if a particular hypothesis were true.

The Observed values are those valued collected.

The expected values are the frequencies expected, based on the null hypothesis.

3.7.3 TYPES OF CHI-SQUARE TESTS

There are two main types of Chi-Square tests namely -

1. Independence
2. Goodness-of-Fit

3.7.3.1 Independence

The Chi-Square Test of Independence is a derivable (also known as inferential) statistical test which examines whether the two sets of variables arelikely to be related with each other or not. This test is used when we have counts of values for two nominal or categorical variables and is considered as non-parametric test. A relatively large sample size and independence of observations are the required criteria for conducting this test.

Example:

In a movie theatre, suppose we made a list of movie genres. Let us consider this as the first variable. The second variable is whether or not the people who came to watch those genres of movies have bought snacks at the theatre. Here the null hypothesis is that th genre of the film and whether people bought snacks or not are unrelatable. If this is true, the movie genres don't impact snack sales.

3.7.3.2 Goodness-of-Fit

In statistical hypothesis testing, the Chi-Square Goodness-of-Fit test determines whether a variable is likely to come from a given distribution or not. We must have a set of data values and the idea of the distribution of this data. We can use this test when we have value counts for categorical variables. This test demonstrates a way of deciding if the data values have a “good enough” fit for our idea or if it is a representative sample data of the entire population.

Example:

Suppose we have bags of balls with five different colours in each bag. The given condition is that the bag should contain an equal number of balls of each colour. The idea we would like to test here is that the proportions of the five coloursof balls in each bag must be exact.

3.7.4 USES OF THE CHI-SQUARED TEST:

- The Chi-squared test can be used to see if your data follows a well-knowntheoretical probability distribution like the Normal or Poisson distribution.

- The Chi-squared test allows you to assess your trained regression model's goodness of fit on the training, validation, and test data sets.

Illustration Problem 1:

A survey on cars had conducted in 2011 and determined that 60% of car owners have only one car, 28% have two cars, and 12% have three or more. Supposing that you have decided to conduct your own survey and have collected the data below, determine whether your data supports the results of the study. Use a significance level of 0.05. Also, given that, out of 129 car owners, 73 had one car and 38 had two cars.

Solution:

Let us state the null and alternative hypotheses.

H_0: The proportion of car owners with one, two or three cars is 0.60, 0.28 and 0.12 respectively.

H_1: The proportion of car owners with one, two or three cars does not match the proposed model.

A Chi-Square goodness of fit test is appropriate because we are examining the distribution of a single categorical variable.

Let's tabulate the given information and calculate the required values.

	Observed (O_i)	**Expected (E_i)**	**$O_i - E_i$**	**$(O_i - E_i)^2$**	**$(O_i - E_i)^2/E_i$**
One car	**73**	**0.60 × 129 = 77.4**	**-4.4**	**19.36**	**0.2501**
Two cars	**38**	**0.28 × 129 = 36.1**	**1.9**	**3.61**	**0.1**
Three or more cars	**18**	**0.12 × 129 = 15.5**	2.5	**6.25**	**0.4032**
Total	**129**				**0.7533**

Therefore, $\chi^2 = \sum(O_i - E_i)^2/E_i = 0.7533$

Let's compare it to the chi-square value for the significance level 0.05.

The degrees for freedom = 3 – 1 = 2

Using the table, the critical value for a 0.05 significance level with df = 2 is 5.99.

That means that 95 times out of 100, a survey that agrees with a sample will have a χ^2 value of 5.99 or less.

The Chi-square statistic is only 0.7533, so we will accept the null hypothesis.

Illustration Problem 2:

From the data given in the following table, find out whether there is any relationship between gender and the preference of colour.

Colour	Male	Female	Total
Red	25	45	70
Green	45	25	70
Blue	50	10	60
Total	120	80	200

(Given : For ν =2, χ 2 0.05 = 5.991)

Solution:

Let us take the following hypothesis:

Null Hypothesis $H0$: There is no relationship between gender and preference of colour.

Alternative Hypothesis Ha: There is relationship between gender and preference of colour.

We have to first calculate the expected value for the observed frequencies.

Colour	**Gender**	**O**	**E**	**O-E**	**(O-E)2**	**(O-E)2 /E**
Red	M	25	42	-17	289	6.88
	F	45	28	17	289	10.32
Blue	M	45	42	3	9	0.21
	F	25	28	-3	9	0.32
Green	M	50	36	14	196	5.44
	F	10	24	-14	196	8.16
						χ **2= 31.33**

The degrees of freedom are (r-1) (c-1) = (3-1) (2-1) = 2.

The critical value of χ 2 for 2 degrees of freedom at 5% level of significance is 5.991.

Since the calculated χ 2 =31.33 exceeds the critical value of χ 2, the null hypothesis is rejected. Hence, the conclusion is that there is a definite relationship between gender and preference of colour.

3.8 CORRELATION

Correlation refers to a process for establishing the relationships between two variables. You learned a way to get a general idea about whether or not two variables are related, is to plot them on a “scatter plot”. While there are many measures of association for variables which are measured at the ordinal or higher level of measurement, correlation is the most commonly used approach.

The correlation coefficient is usually represented using the symbol r, and it ranges from -1 to +1.

A correlation coefficient quite close to 0, but either positive or negative, implies little or no relationship between the two variables. A correlation coefficient close to plus 1 means a positive relationship between the two variables, with increases in one of the variables being associated with increases in the other variable.

A correlation coefficient close to -1 indicates a negative relationship between two variables, with an increase in one of the variables being associated with a decrease in the other variable. A correlation coefficient can be produced for ordinal, interval or ratio level variables, but has little meaning for variables which are measured on a scale which is no more than nominal.

3.8.1 WHAT DOES CORRELATION MEASURE?

In statistics, Correlation studies and measures the direction and extent of relationship among variables, so the correlation measures co-variation, not causation. Therefore, we should never interpret correlation as implying cause and effect relation. For example, there exists a correlation between two variables X andY, which means the value of one variable is found to change in one direction, the value of the other variable is found to change either in the same direction (i.e. positive change) or in the opposite direction (i.e. negative change).

3.8.2 KARL PEARSON’SCORRELATION COEFFICIENT

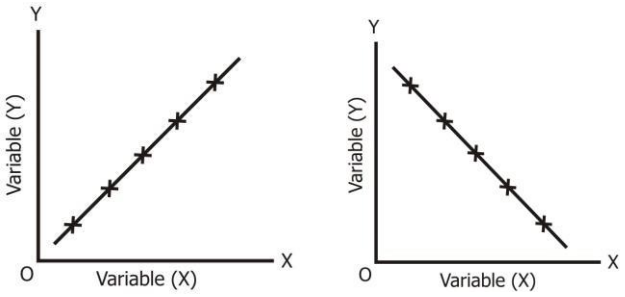
The correlation coefficient, r, is a summary measure that describes the extent of the statistical relationship between two interval or ratio level variables. The correlation coefficient is scaled so that it is always between -1 and +1. When r is close to 0 this means that there is little relationship between the variables and the farther away from 0 r is, in either the positive or negative direction, the greater the relationship between the two variables.

The two variables are often given the symbols X and Y. In order to illustrate how the two variables are related, the values of X and Y are pictured by drawing the scatter diagram, graphing combinations of the two variables. The scatter diagram is given first, and then the method of determining Pearson's r is presented.

3.8.3 SCATTER DIAGRAM

A scatter diagram is used to examine the relationship between both the axes (X and Y) with one variable. In the graph, if the variables are correlated, then thepoint drops along a curve or line. A scatter diagram or scatter plot gives an idea of the nature of relationship.

In a scatter correlation diagram, if all the points stretch in one line, then the correlation is perfect and is in unity. However, if the scatter points are widely scattered throughout the line, then the correlation is said to be low. If the scatter points rest near a line or on a line, then the correlation is said to be linear.



Positive Correlation **Negative Correlation**

Pearson Correlation Coefficient Formula:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2] \times [n\sum y^2 - (\sum y)^2]}$$

Where,
r - Pearson correlation coefficient
x - Values of first variable

y - Values of second variable

n - Total number of values

Illustration Problem 3:

A survey was conducted in your city. Given is the following sample data containing a person's age and their corresponding income. Find out whether the increase in age has an effect on income using the correlation coefficient formula.

Age (Years)	25	30	36	43
Income (Rs.)	30000	44000	52000	7000

Solution:

x - Age (years)

y - Income (Rs.)/1000

x	y	x^2	y^2	xy
25	30	625	900	750
30	44	900	1936	1320
36	52	1296	2704	1872
43	70	1849	4900	3010
$\sum x = 131$	**$\sum y = 196$**	**$\sum x^2 = 4670$**	**$\sum y^2 = 10440$**	**$\sum xy = 6952$**

$r = n(\sum xy) - (\sum x)(\sum y) / \sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$

$r = 4 \times 6952 - (131 \times 196) / \sqrt{[4 \times 4670 \times 131^2] \times [4 \times 10440 - 196^2]}$

r = 0.99

Yes, with the increase in age a person's income increases as well, since the Pearson correlation coefficient between age and income is very close to 1.

Illustration Problem4:

Marks obtained by 5 students in marketing and finance as given below:

Marketing	16	15	12	10	8
Finance	11	18	10	20	17

Calculate the Pearson correlation coefficient.

Solution:

x - Marks in marketing y - Marks in finance

x	y	x^2	y^2	xy
16	11	256	121	176
15	18	225	324	270
12	10	144	100	120
10	20	100	400	200
8	17	64	289	136
$\sum x = 61$	**$\sum y = 76$**	**$\sum x^2 = 789$**	**$\sum y^2 = 1234$**	**$\sum xy = 902$**

$r = n(\sum xy) - (\sum x)(\sum y) / \sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$

$r = 5 \times 902 - (61 \times 76) / \sqrt{[5 \times 789 \times 61^2] \times [5 \times 1234 - 76^2]}$

r = - 0.424

3.8.5 SPEARMAN'S RANK CORRELATION COEFFICIENT

Spearman's rank correlation coefficient, denoted by r_s is a numerical value such that $-1 \leq r_s \leq 1$.It gives a measure of the likelihood of one variable increasing as the other increases (a direct association) or of one variable decreasing as the other increases (an inverse association).

Direct associations are indicated by positive values, and inverse associations are indicated by negative values. No association is indicated by a value of 0. The stronger the association, the closer r_s is to −1 or 1, and the weaker the association, the closer it is to 0. Rank correlation coefficient values of 1 or −1 mean that either theranks agree entirely (r_s=1) or they are direct opposites (r_s=−1).

The formula for Spearman's rank correlation coefficient is

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

d^2- square of the difference in the ranks of the two coordinates

n - number of points in the data set

Nb: Ranks should be from small value to large value

Illustration Problem 5:

Find the Spearman's correlation coefficient between x and y.

x	4	7	8	5	8	12
y	7	6	6	4	6	10

Solution:

x	4	7	8	5	8	12
R_1	1	3	4.5	2	4.5	6
y	7	6	6	4	6	10
R_2	5	3	3	1	3	6
d	-4	0	1.5	1	1.5	0
d^2	16	0	2.25	1	2.25	0

$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$

$\sum d^2 = 16 + 0 + 2.25 + 1 + 2.25 + 0 = 21.5$

$= 1 - [6 (21.5)] / 6 (35)$

$r_s = 0.386$

Illustration Problem 6:

Find the Spearman's rank correlation for the following ranks of two judges:

Rank of Judge A	1	5	4	2	6	3
Rank of Judge B	1	5	4	3	6	2

Solution:

Rank of Judge A	1	5	4	2	6	3
Rank of Judge B	1	5	4	3	6	2
Difference (d)	0	0	0	-1	0	1
d^2	0	0	0	1	0	1

$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$

$= 1 - [6 (2)] / 6 (35) =$ **0.943**

3.9 REGRESSION ANALYSIS

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modelling the future relationship between them.

Regression analysis includes several variations, such as linear, multiple linear, and nonlinear. The most common models are simple linear and multiple linear. Nonlinear regression analysis is commonly used for more complicated data sets in which the dependent and independent variables show a nonlinear relationship.

3.9.1 SIMPLE LINEAR REGRESSION

Simple linear regression is a model that assesses the relationship between a dependent variable and an independent variable. The simple linear model is expressed using the following equation:

$y = a + bx$

Where:

y – Dependent variable

x – Independent (explanatory) variable

a – Intercept

b – Slope

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n \sum xy - (\sum x)(\sum Y)}{n \sum x^2 - (\sum x)^2}$$

Illustration Problem 7:

Find linear regression equation for the following two sets of data:

x	2	4	6	8
y	3	7	5	10

Solution:

Construct the following table:

x	y	x^2	xy
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
$\sum x = 20$	**$\sum y = 25$**	**$\sum x^2 = 120$**	**$\sum xy = 144$**

$$b = \frac{n \sum xy - (\sum x)(\sum Y)}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{(4 \times 144) - (20 \times 25)}{(4 \times 120) - 400}$$

= 0.95

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2}$$

$$= \frac{(25 \times 120) - (20 \times 144)}{(4 \times 120) - 400}$$

= 1.5

Linear regression is given by:

y = a + bx

y = 1.5 + 0.95 x

Illustration problem 8:

Find out the linear regression equation for the following sets of data:

x	2	3	5	8
y	3	6	5	12

Solution:

Construct the following table:

x	y	x^2	xy
2	3	4	6
3	6	9	18
5	5	25	25
8	12	64	96
$\sum x = 18$	**$\sum y = 26$**	**$\sum x^2 = 102$**	**$\sum xy = 145$**

$$b = \frac{n\sum xy - (\sum x)(\sum Y)}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{(4 \times 145) - (18 \times 26)}{(4 \times 102) - 324}$$

$$= 1.33$$

$$a = a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2}$$

$$= \frac{(26 \times 102) - (18 \times 145)}{(4 \times 102) - 145}$$

$$= 0.515$$

Linear regression equation is given by:

$y = 0.515 + 1.33x$

3.10 ANALYSIS OF VARIANCE (ANOVA)

ANOVA, or Analysis of Variance is a parametric statistical technique that helps in finding out if there is a significant difference between the mean of three or more groups. It checks the impact of various factors by comparing groups (samples) based on their respective mean. ANOVA tests the null hypothesis that all group means are equal, against the alternative hypothesis that at least one group mean is different.

3.10.1 ASSUMPTIONS FOR ANOVA

1. The dependent variable is approximately normally distributed within each group. This assumption is more critical for smaller sample sizes.
2. The samples are selected at random and should be independent of one another.
3. All groups have equal standard deviations.
4. Each data point should belong to one and only one group. There should be no overlap or sharing of data points between groups.

In the case of two-way ANOVA, there are additional assumptions related to the interaction between the independent variables.

- The effect of one independent variable on the dependent variable should be consistent across all levels of the other independent variable.
- Combined effect of two independent variables is equal to the sum of their individual effects.

3.10.2 TYPES OF ANOVA

There are two main types of ANOVA:

1. **One-way ANOVA:** This is the most basic form of ANOVA and is used when there is only one independent variable with more than two levels or groups. It assesses whether there are any statistically significant differences among the means of the groups.
2. **Two-way ANOVA:** Extending the one-way ANOVA, two-way ANOVA involves two independent variables. It allows for the examination of the main effects of each variable as well as the interaction between them. The interaction effect explores whether the effect of one variable on the dependent variable isdifferent depending on the level of the other variable.

3.10.3 APPLICATIONS OF ONE-WAY ANOVA

The One-Way ANOVA is often used to analyze data from the following types of studies:

- Field studies
- Experiments
- Quasi-experiments

The One-Way ANOVA is commonly used to test the following:

- Statistical differences among the means of two or more groups
- Statistical differences among the means of two or more interventions
- Statistical differences among the means of two or more change scores

Several use cases of one-way ANOVA from different domains are:

- **Medicine:** One-way ANOVA can be used to compare the effectiveness of different treatments on a particular medical condition. For example, it could be used to determine whether three different drugs have significantly different effects on reducing blood pressure.
- **Education:** One-way ANOVA can be used to analyze whether there are significant differences in test scores among students who have been taught using different teaching methods.
- **Marketing:** One-way ANOVA can be employed to assess whether there are significant differences in customer satisfaction levels among products from different brands.
- **Manufacturing:** One-way ANOVA can be utilized to analyze whether there are significant differences in the strength of materials produced by different manufacturing processes.

- **Psychology:** One-way ANOVA can be used to investigate whether thereare significant differences in anxiety levels among participants exposed to different stressors.
- **Agriculture:** One-way ANOVA can be used to determine whether different fertilizers lead to significantly different crop yields in farming experiments.

3.10.4 PROCEDURE FOR ONE-WAY ANOVA

Example:

Fertilizer Effect on Plant Growth

Imagine you are researching the impact of different fertilizers on plant growth. You apply three types of fertilizer (A, B and C) to separate groups of plants. After a set period, you measure the average height of plants in each group. You can use one-way ANOVA to test if there's a significant difference in average height among plants grown with different fertilizers.

Step 1: Null and Alternative Hypotheses

- **Null Hypothesis(H0)**: The means of all groups are equal (there's no significant difference in plant growth due to fertilizer type)
- **Alternative Hypothesis (H1):** At least one group mean is different from the others (fertilizer type has a significant effect on plant growth).

Step 2: Data Collection and Data Organization

After a set growth period, carefully measure the final height of each plant in allthree groups. Now organize your data. Each column represents a fertilizer type (A, B, C) and each row holds the height of an individual plant within that group.

Step 3: Calculate the group Statistics

- Compute the mean final height for plants in each fertilizer group (A, B and C).
- Compute the total number of plants observed (N) across all groups.

- Determine the total number of groups (K) in our case, k=3(A, B, C)

Step 4: Calculate Sum of Square

So Total sum of square, between-group sum of square, within-group sum of square will be calculated.

Here, Total Sum of Square represents the total variation in final height across all plants.

Between-Group Sum of Square reflects the variation observed between the average heights of the three fertilizer groups. And Within-Group Sum of Square captures the variation in final heights within each fertilizer group.

Step 5: Compute Degrees of Freedom

Degrees of freedom define the number of independent pieces of information used to estimate a population parameter.

- **Degrees of Freedom Between-Group:** k-1 (number of groups minus 1) So, here it will be 3-1 =2
- **Degrees of Freedom Within-Group:** N-k (Total number of observations minus number of groups)

Step 6: Calculate Mean Squares

Mean Squares are obtained by dividing the respective Sum of Squares by degrees of freedom.

- **Mean Square Between:** Between- Group Sum of Square/Degrees of Freedom Between-Group
- **Mean Square Within:** Within-Group sum of Square/Degrees of Freedom Within-Group

Step 7: Compute F-statistics

The F-statistic is a test statistic used to compare the variation between groups to the variation within groups. A higher F-statistic suggests a potentially stronger effect of fertilizer type on plant growth.

The F-statistic for one-way ANOVA is calculate by using this formula:

$$F = \frac{MS_{between}}{MS_{within}}$$

Here,

$MS_{between}$ is the mean square between groups, calculated as the sum of squares between groups divided by the degrees of freedom between groups.

MS_{within} is the mean square within groups, calculated as the sum of squares within groups divided by the degrees of freedom within groups.

- **Degrees of Freedom Between Groups (dof between):** dof_between = k-1

  Where k is the number of groups(levels) of the independent variable.

- **Degrees of Freedom Within Groups (dof_within):** dof_within = N-k

Where N is the number of observations and k is the number of groups(levels) of the independent variable.

For one-way ANOVA, total degrees of freedom is the sum of the degrees of freedom between groups and within groups:

dof_total= dof_between + dof_within

Step 8: Determine Critical Value and Decision

Choose a significance level (alpha) for the analysis, usually 0.05 is chosen

Look up the critical F-value at the chosen alpha level and the calculated Degrees of Freedom Between-Group and Degrees of Freedom Within-Group using an F-distribution table.

Step 9: Compare the calculated F-statistic with the critical F-value

- If the calculated F-statistic is greater than the critical F-value, reject the null hypothesis(H0). This indicates a statistically significant difference in average plant heights among the three fertilizer groups.
- If the calculated F-statistic is less than or equal to the critical F-vale, fail to reject the null hypothesis (H0). You cannot conclude a significant difference based on this data.

3.10.5 TWO-WAY ANOVA

A two-way ANOVA is used to estimate how the mean of a quantitative variable changes according to the levels of two categorical variables. Use a two-way ANOVA when you want to know how two independent variables, in combination, affect a dependent variable.

Example:

You are researching which type of fertilizer and planting density produces the greatest crop yield in a field experiment. You assign different plots in a field to a combination of fertilizer type (1, 2, or 3) and planting density (1=low density, 2=high density), and measure the final crop yield in bushels per acre at harvest time.

When to use a two-way ANOVA?

You can use a two-way ANOVA when you have collected data on a quantitative dependent variable at multiple levels of two categorical independent variables.

A quantitative variable represents amounts or counts of things. It can be divided to find a group mean.

You can use a two-way ANOVA to find out if fertilizer type and planting density have an effect on average crop yield.

A two-way ANOVA with interaction tests three null hypotheses at the same time:

- There is no difference in group means at any level of the first independent variable.

- There is no difference in group means at any level of the second independent variable.
- The effect of one independent variable does not depend on the effect of the other independent variable (no interaction effect).

A two-way ANOVA without interaction (an additive two-way ANOVA) only tests the first two of these hypotheses.

3.10.4.1 ASSUMPTIONS OF THE TWO-WAY ANOVA

To use a two-way ANOVA your data should meet certain assumptions. Two-way ANOVA makes all of the normal assumptions of a parametric test of difference:

1. **Homogeneity of variance (a.k.a. homoscedasticity)**

   The variation around the mean for each group being compared should be similar among all groups. If your data don't meet this assumption, you may be able to use a non-parametric alternative, like the Kruskal-Wallis test.

2. **Independence of observations**

   Your independent variables should not be dependent on one another (i.e. one should not cause the other). This is impossible to test with categorical variables – it can only be ensured by good experimental design.

   In addition, your dependent variable should represent unique observations – that is, your observations should not be grouped within locations or individuals.

   If your data don't meet this assumption (i.e. if you set up experimental treatments within blocks), you can include a blocking variable and/or use a repeated-measures ANOVA.

3. **Normally-distributed dependent variable**

   The values of the dependent variable should follow a bell curve (they should be normally distributed). If your data don't meet this assumption, you can try a data transformation.

Let's Sum Up

In this unit we have discussed the concepts of data editing, data coding, data tabulation and data entry after data collection. The data analysis meaning and types have been discussed with examples. The hypothesis testing procedure and methods of parametric and non-parametric data have been explained with suitable examples. The chi-square test has been explained with its formula and illustrations. Correlation concepts, types and applications have been presented with suitable problems and solutions. Regression analysis is discussed with example problems. The One-Way and Two-Way Analysis of Variance have been explained with their applications and assumptions.

Check your Progress – Quiz

1. The unedited responses from a respondent exactly as indicated by that respondent are referred to as _____

   A. Codes

   B. Files

   C. Raw data

   D. Strings

2. The assignment of numbers to edited data is known as _____

   A. Editing

   B. Coding

   C. Test tabulation

   D. Verification

3. Valarie's job at a research firm is to transfer data from survey questionnaires to a computer file. Her job is _____

   A. Data entry

   B. Data coding

   C. Data editing

   D. Data scanning

4. In creating coding categories for open-ended questions, the tallying of the replies to a small number completed questionnaires for a question in order to create coding categories for that question is known as _____

   A. Recording data

B. Test tabulation

C. Data entry

D. Creating a plug value

5. If the Critical region is evenly distributed then the test is referred as?
A. Two tailed

B. One tailed

C. Three tailed

D. Zero tailed

6. Type 1 error occurs when?

A. We reject H_0 if it is False

B. We reject H_0 if it is True

C. We accept H_0 if it is True

D. We accept H_0 if it is False

7. Correlation refers to _____

A. The association between two variables

B. The causal relationship between two variables

C. The proportion of variance that two variables share

D. None of the above

8. The slope of the regression line of Y on X is also referred to as the:

A. Regression coefficient of X on Y
B. The correlation coefficient of X on Y
C. Regression coefficient of Y on X
D. Correlation coefficient of Y on X.

9. _____attempts to model the relationship between two variables by fitting a linear equation to observed data

A. Linear regression
B. Covariance
C. Dependent Variable
D. Standard Deviation

10. Which of the following best describes the purpose of using ANOVA in research?

A. ANOVA is used to compare the means of two groups
B. ANOVA is used to compare the means of more than two groups
C. ANOVA is used to determine the correlation between two variables
D. ANOVA is used to determine the interaction effect between dependent variables

UNIT SUMMARY

Data Preparation and Analysis focuses on preparing, organizing, and processing data through editing, coding, and entry for analysis, followed by hypothesis testing and various statistical methods such as univariate, bivariate analysis, parametric/nonparametric tests, Chi-Square tests, correlation (Karl Pearson's and Spearman's), regression analysis, and one-way/two-way ANOVA to interpret relationships and patterns in the data.

GLOSSARY

Keywords	Meaning
Data Preparation	The process of cleaning, transforming, and organizing raw data for analysis.
Editing	Reviewing and correcting data for consistency and accuracy before analysis
Coding	Assigning numeric or symbolic codes to responses for easier data categorization and analysis.
Parametric Tests:	Statistical tests that assume the underlying data follows a specific distribution, usually normal.
Nonparametric Tests	Statistical tests that do not assume a specific distribution of the data.

Self-Assessment Questions:

1. Janice and Paul did a study on feelings of stress and life satisfaction. Participants completed a measure on how stressed they were feeling (on a 1 to 30 scale) and a measure of how satisfied they felt with their lives (measured on a 1 to 10 scale). The table below indicates the participants' scores.

Participant	1	2	3	4	5	6	7	8	9	10
Stress score	11	25	19	7	23	6	11	22	25	10
Satisfaction	7	1	4	9	2	8	8	3	3	6

Find the Karl Pearson's coefficient of correlation for the above data

2. Calculate the Spearman's rank correlation coefficient from the following data:

X	20	38	30	40	50	55
Y	17	45	30	35	40	25

3. From the following data, find out whether there is any relationship between gender and preference of colour at 5% significance level:

Colour	**Male**	**Female**	**Total**
Green	40	60	100
White	35	25	60
Yellow	25	15	40
Total	**100**	**100**	**200**

4. A student counted the number of words in an essay she has written, recording the total every 10 lines:

No. of lines (x)	10	20	30	40	50	60	70	80
No. of words (y)	75	136	210	291	368	441	519	588

Find the regression equation

5. Explain the process of raw data in the form of data editing, data entry, data coding and data tabulation

6. Discuss the procedure of hypothesis testing with suitable illustration

7. Describe the one-way and two-way ANOVA and their areas of application and procedure of testing of hypothesis

ANSWERS FOR CHECK YOUR PROGRESS

1. C) Raw data
2. B) Coding
3. A) Data entry
4. B) Test tabulation
5. A) Two-tailed
6. B) We reject H0 if it is True
7. A) The association between two variables
8. C) Regression coefficient of Y on X
9. A) Linear regression
10. B) ANOVA is used to compare the means of more than two groups

Reference of further Reading:

1. Amir D Aczel & Jayavel Sounderpandian, "Complete Business Statistics", Tata-McGraw Hill, 7th Edition,2017
2. SP Gupta & MP Gupta, "Business Statistics", Sultan Chand & Sons, 2019

UNIT IV

MULTIVARIATE STATISTICAL ANALYSIS

MULTIVARIATE STATISTICAL ANALYSIS
Exploratory and Confirmatory Factor Analysis -Discriminant Analysis- Cluster Analysis -Conjoint Analysis - Multiple Regression- Multidimensional Scaling- Their Application In Marketing Problems - Application of Statistical Software For Data Analysis- SEM Analysis

MULTIVARIATE STATISTICAL ANALYSIS

UNIT OBJECTIVES

The unit on Multivariate Statistical Analysis has the following objectives:

- To understand the meaning and application of various techniques of multivariate analysis
- To explain the features of various multivariate analysis techniques
- To give the areas of applications of multivariate analysis in marketing problems
- To study the application of statistical software for data analysis
- To learn the meaning and application of SEM analysis

4.1 INTRODUCTION:

Multivariate analysis is used to find patterns and correlations between multiple factors by analyzing two or more variables at once. Experts generally group the various approaches of multivariate analysis into two camps — dependent techniques for those instances when the outlined variables depend on one another, and independent techniques for when they do not. Multivariate analysis allows you to find patterns between variables, helping you better understand the effects that different factors have on each other and the relationships between them. It represents a critical tool for marketers looking for ways to get deeper insight into the outcome of campaign decisions.

We can use multivariate analysis to better determine the impact of complex market and global factors on everything from sales to consumer behaviour. This unit will demonstrate how multivariate analysis can help you determine how different variables interact with each other in complex scenarios, and the power that comes with identifying patterns and the relationships between those variables. It will also detail multiple methods, examples, and possible use cases to help you understand the strengths of each.

4.2 MULTIVARIATE ANALYSIS

Multivariate analysis is a method of gathering multiple sets of data and drawing cause-and-effect conclusions about their constituent parts.

Conducting multivariate analyses can help companies forecast future opportunities, risks, and demand for products. This helps with investment strategies, business decisions, and setting expectations.

Businesses often have large quantities of financial, operational, customer, and purchase data to help inform business decisions based on statistical significance rather than intuition. By relying on this type of analysis, you can decrease your overall risk and chance of failure.

A company may also use Multivariate analysis to gain new insights. This could include uncovering new customer targets or identifying market patterns that exist during certain times of the year or hours of the day. Without Multivariate analysis, opportunities might get buried beneath an avalanche of unorganized data.

4.3 EXPLORATORY FACTOR ANALYSIS

Exploratory factor analysis (EFA) is one of a family of multivariate statistical methods that attempts to identify the smallest number of hypothetical constructs (also known as factors, dimensions, latent variables, synthetic variables, or internal attributes) that can parsimoniously explain the covariation observed among a set of measured variables (also called observed variables, manifest variables, effect indicators, reflective indicators, or surface attributes).

That is, to identify the common factors that explain the order and structure among measured variables. In the social and behavioral sciences, factors are assumed to be unobservable characteristics of people, which are manifested in differences in the scores attained by those people on the measured variables.

A factor is an unobservable variable that influences more than one observed measure and that accounts for the correlations among these observed measures. In other words, the observed measures are interrelated because they share a common cause (i.e., they are influenced by the same underlying construct); if the latent

construct was partitioned out, the intercorrelations among the observed measures will be zero.

Measured variables are selected for their utility as indicators of anticipated factors. That is, their content, convergent, and discriminant validity. Thus, measured variables should adequately represent the domains the factors are thought to tap and not include variables from unrelated domains.

For example, if the broad domain of reading is to be analyzed, multiple variables that measure each dimension of reading (e.g., phonemic awareness, phonics, vocabulary, comprehension) should be selected, but it would be inappropriate to include variables that measure addition and subtraction skills.

In contrast, variables that tap blending, segmentation, rhyme, and deletion of phonemes would be appropriate if the narrow domain of phonemic awareness is to be investigated. Inadequate sampling of the domain may fail to uncover important common factors or produce spurious factors. Of course, investigations of existing measurement instruments have a ready-made list of variables, and the researcher's task is to evaluate the validity of those variables (structural, factorial, or construct validity).

4.4 CONFIRMATORY FACTOR ANALYSIS (CFA)

CFA is a special form of factor analysis, most commonly used in social science research. It is used to test whether measures of a construct are consistent with a researcher's understanding of the nature of that construct (or factor). As such, the objective of confirmatory factor analysis is to test whether the data fit a hypothesized measurement model. This hypothesized model is based on theory and/or previous analytic research.

In confirmatory factor analysis, the researcher first develops a hypothesis about what factors they believe are underlying the measures used (e.g., "Depression" being the factor underlying the Beck Depression Inventory and the Hamilton Rating Scale for Depression) and may impose constraints on the model based on these a priori hypotheses. By imposing these constraints, the researcher is forcing the model to be consistent with their theory.

For example, if it is posited that there are two factors accounting for the covariance in the measures, and that these factors are unrelated to each other, the researcher can create a model where the correlation between factor A and factor B is constrained to zero.

Model fit measures could then be obtained to assess how well the proposed model captured the covariance between all the items or measures in the model. If the constraints the researcher has imposed on the model are inconsistent with the sample data, then the results of statistical tests of model fit will indicate a poor fit, and the model will be rejected. If the fit is poor, it may be due to some items measuring multiple factors. It might also be that some items within a factor are more related to each other than others.

4.4.1 Difference between EFA and CFA

The goal of EFA is to identify factors based on data and to maximize the amount of variance explained. The researcher is not required to have any specific hypotheses about how many factors will emerge, and what items or variables these factors will comprise. If these hypotheses exist, they are not incorporated into and do not affect the results of the statistical analyses.

By contrast, CFA evaluates *a priori* hypotheses and is largely driven by theory. CFA analyses require the researcher to hypothesize, in advance, the number of factors, whether or not these factors are correlated, and which items/measures load onto and reflect which factors. As such, in contrast to exploratory factor analysis, where all loadings are free to vary, CFA allows for the explicit constraint of certain loadings to be zero.

4.5 DISCRIMINANT ANALYSIS

Discriminant analysis (DA) is a multivariate technique used to separate two or more groups of observations (individuals) based on k variables measured on each experimental unit (sample) and find the contribution of each variable in separating the groups.

Discriminant function analysis is a supervised classification technique, wherein distinct observations with predefined group memberships along with associated

variables are analyzed as well as separated and new objects are allocated to the previously defined groups.

Discriminant Analysis may be used for two objectives: either we want to assess the adequacy of classification, given the group memberships of the objects under study; or we wish to assign objects to one of a number of (known) groups of objects.

Discriminant analysis is a classification problem, where two or more groups or clusters or populations are known a priori and one or more new observations are classified into one of the known populations based on the measured characteristics.

4.6 ADVANTAGES OF DISCRIMINANT FUNCTION ANALYSIS

Discriminant analysis offers a potential advantage: it classified ungrouped cases. "The functions are generated from a sample of cases for which group membership is known; the functions can then be applied to new cases with measurements for the predictor variables but unknown group membership."

4.7 TYPES OF CLUSTER ANALYSIS

The clustering algorithm needs to be chosen experimentally unless there is a mathematical reason to choose one cluster method over another.It should be noted that an algorithm that works on a particular set of data will not work on another set of data. There are a number of different methods to perform cluster analysis. Some of them are,

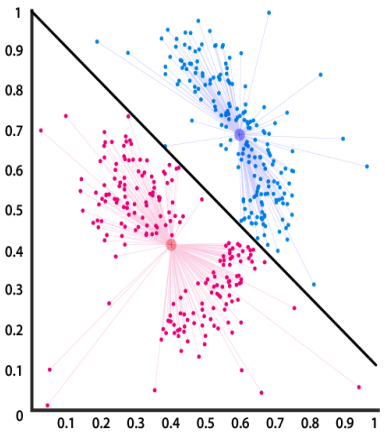
4.7.1 Hierarchical Cluster Analysis

In this method, first, a cluster is made and then added to another cluster (the most similar and closest one) to form one single cluster. This process is repeated until all subjects are in one cluster. This particular method is known as **Agglomerative method**. Agglomerative clustering starts with single objects and starts grouping them into clusters.

The divisive method is another kind of Hierarchical method in which clustering starts with the complete data set and then starts dividing into partitions.

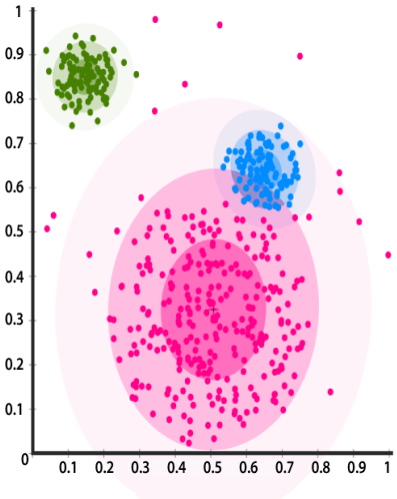
4.7.2 Centroid-based Clustering

In this type of clustering, clusters are represented by a central entity, which may or may not be a part of the given data set. K-Means method of clustering is used in this method, where k are the cluster centers and objects are assigned to the nearest cluster centres.



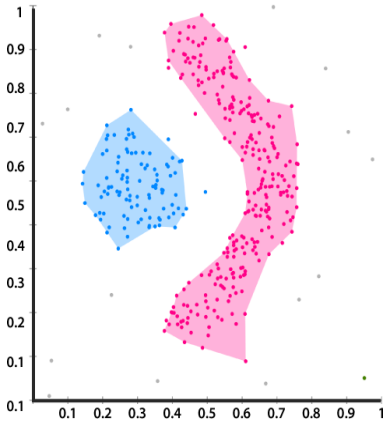
4.7.3 Distribution-based Clustering

It is a type of clustering model closely related to statistics based on the modals of distribution. Objects that belong to the same distribution are put into a single cluster.This type of clustering can capture some complex properties of objects like correlation and dependence between attributes.



4.7.4 Density-based Clustering

In this type of clustering, clusters are defined by the areas of density that are higher than the remaining of the data set. Objects in sparse areas are usually required to separate clusters.The objects in these sparse points are usually noise and border points in the graph.The most popular method in this type of clustering is DBSCAN.



To learn more on the cluster and other statistics-related topics, visit BYJU'S.

4.7.5 Applications and Examples

It is the principal job of exploratory data mining, and a common method for statistical data analysis. It is used in many fields, such as machine learning, image analysis, pattern recognition, information retrieval, data compression, bioinformatics and computer graphics.

It can be used to examine patterns of antibiotic resistance, to incorporate antimicrobial compounds according to their mechanism of activity, to analyse antibiotics according to their antibacterial action.

Cluster analysis can be a compelling data-mining means for any organization that wants to recognise discrete groups of customers, sales transactions, or other kinds of behaviours and things. For example, insurance providing companies use cluster analysis to identify fraudulent claims and banks apply it for credit scoring.

4.8 CONJOINT ANALYSIS

Conjoint analysis is defined as a survey-based advanced market research analysis method that attempts to understand how people make complex choices. We make choices that require trade-offs every day — so often that we may not even realize it. Even simple decisions like choosing a laundry detergent or deciding to booka flight are mental conjoint studies that contain multiple elements that lead us to our choice.

Conjoint analysis is one of the most effective models for extracting consumer preferences during the purchasing process. This data is then turned into a quantitative measurement using statistical analysis. It evaluates products or services in a way no other method can.

Researchers consider conjoint analysis as the best survey method for determining customer values. It consists of creating, distributing, and analyzing surveys among customers to model their purchasing decision based on response analysis.

Over the past 50 years, Conjoint analysis has evolved into a method that market researchers and statisticians implement to predict the kinds of decisions consumers will make about products by using questions in a survey.

The central idea is that consumers evaluate different characteristics of aproduct and decide which are more relevant to them for any purchase decision. An online conjoint survey's primary aim is to set distinct values to the alternatives that thebuyers may consider when making a purchase decision. Equipped with this knowledge, marketers can target the features of products or services that are highly important and design messages more likely to strike a chord with target buyers.

4.8.1 Advantages of conjoint analysis

There are multiple advantages to using conjoint analysis in your surveys:

- It helps researchers estimate the trade-offs that consumers make on a psychological level when they evaluate numerous attributes simultaneously.
- Researchers can measure consumer preferences at an individual level.

- It gives researchers insights into real or hidden drivers that may not be too apparent.
- Conjoint analysis can study the consumers and attributes deeper and create a needs-based segmentation.

For example, assume a scenario where a product marketer needs to measure individual product features' impact on the estimated market share or sales revenue.

In this conjoint study example, we'll assume the product is a mobile phone. The competitors are Apple, Samsung, and Google. The organization needs to understand how different customers value attributes, such as brand, price, screen size, and screen resolution. Armed with this information, they can create their product range to match consumer preferences.

Conjoint analysis assigns values to these product attributes and levels by creating realistic choices and asking people to evaluate them.

It enables businesses to mathematically analyze consumer or client behaviour and make decisions based on real insights from customer data. This allows them to develop better business strategies that provide a competitive edge. To fulfil customer wishes profitably requires companies to fully understand which aspects of their product and service are most valued.

4.9 MULTIPLE – REGRESSION ANALYSIS

Multiple regression analysis is a statistical technique that analyzes the relationship between two or more variables and uses the information to estimate the value of the dependent variables. In multiple regression, the objective is to develop a model that describes a dependent variable y to more than one independent variable.

4.9.1 Multiple Regression Formula

In linear regression, there is only one independent and dependent variable involved. But, in the case of multiple regression, there will be a set of independent variables that helps us to explain better or predict the dependent variable y.

The multiple regression equation is given by

$$y = a + b_{1\times1} + b_{2\times2} + \ldots\ldots + b_{kxk}$$

where x_1, x_2, ….x_k are the k independent variables and y is the dependent variable.

Multiple regression analysis permits to control explicitly for many other circumstances that concurrently influence the dependent variable. The objective of regression analysis is to model the relationship between a dependent variable and one or more independent variables. Let k represent the number of variables and denoted by x_1, x_2, x_3, ……, x_k. Such an equation is useful for the prediction of valuefor y when the values of x are known.

Mostly, the statistical inference has been kept at the bivariate level. Inferential statistical tests have also been developed for multivariate analyses, which analyses the relation among more than two variables. Commonly used extension of correlation analysis for multivariate inferences is multiple regression analysis. Multiple regression analysis shows the correlation between each set of independent and dependent variables.

4.9.2 Multicollinearity

Multicollinearity is a term reserved to describe the case when the inter-correlation of predictor variables is high.

Signs of Multicollinearity

- The high correlation between pairs of predictor variables.
- The magnitude or signs of regression coefficients do not make good physical sense.
- Non-significant regression coefficients on significant predictors.
- The ultimate sensitivity of magnitude or sign of regression coefficients leads to the insertion or deletion of a predictor variable.

4.9.3 Steps of multiple regression

The five steps to follow in a multiple regression analysis are model building, model adequacy, model assumptions – residual tests and diagnostic plots, potential modeling problems and solution, and model validation.

4.9.4 APPLICATIONS OF MULTIPLE – REGRESSION ANALYSIS IN BUSINESS DECISION MAKING

Multiple regression analysis is a useful tool in a wide range of applications. From business, marketing and sales analytics to environmental, medical and technological applications, multiple regression analysis helps professionals evaluate diverse data that supports goals, processes and outcomes in many industries.

4.8.4.1 Gives insight into predictive factors

Conducting a multiple regression analysis is useful for determining what factors are affecting different aspects of a business' processes. For instance, revenue can be one type of Y-value, where different independent variables like the number of sales and cost of goods sold affect business revenue. With multiple regression analysis, analysts can identify the individual activities that affect specific metrics they want to measure, giving them better insight into how to improve efficiency and productivity.

4.9.4.2 Predicts factors affecting outcomes

When companies can analyze the factors that affect certain business operations, management can better predict which independent variables influence the dependent functions of the business. For example, a business analyst can predict which factors are likely to affect an organization's future profitability, based on the results of a multiple regression analysis.

In this case, the analyst may calculate the regression using the formula where profit is the predictive variable and factors like overhead, liabilities and total sales revenue represent the (b) and (X) values in the formula. When the analyst understands how much these factors affect profits, they can better predict the variables that may affect profits in the future.

4.9.4.3 Creates Models for Cause-and-Effect Analysis

Understanding the mathematical data that multiple regression analysis can provide allows professionals to model the information in a graph or chart. Displaying multiple regression–how external variables cause changes in a dependent variable–in this way can help you model the cause-and-effect relationship to better see the changes taking place in real time. This can be especially beneficial for financial activities like investing in stocks and securities, where traders can see the cause-and-effect relationship in a chart to understand how economic factors are influencing current market shares.

4.10 MULTI – DIMENSIONAL SCALING (MDS)

The term **scaling** comes from psychometrics, where abstract concepts ("objects") are assigned numbers according to a rule. For example, you may want to quantify a person's attitude to global warming. You could assign a "1" to "doesn't believe in global warming", a 10 to "firmly believes in global warming" and a scale of 2 to 9 for attitudes in between. You can also think of "scaling" as the fact that you're essentially *scaling down the data* (i.e. making it simpler by creating lower-dimensional data). Data that is scaled down in dimension keeps similar properties. For example, two data points that are close together in high-dimensional space will also be close together in low-dimensional space. The "**multidimensional**" part is due to the fact that you aren't limited to two dimensional graphs or data. Three-dimensional, four-dimensional and higher plots are possible.

Multidimensional scaling is a **visual representation of distances or dissimilarities between sets of objects**. "Objects" can be colours, faces, map coordinates, political persuasion, or any kind of real or conceptual stimuli. Objects that are more similar (or have shorter distances) are closer together on the graph than objects that are less similar (or have longer distances). As well as interpreting dissimilarities as distances on a graph, MDS can also serve as a dimension reduction technique for high-dimensional data.

MDS is now used over a wide variety of disciplines. It's use isn't limited to a specific matrix or set of data; In fact, just about any matrix can be analyzed with the technique as long as the matrix contains some type of relational data. Examples of relational data include correlations, distances, multiple rating scales or similarities.

4.10.1 Applications in Marketing:

- Brand image measurement
- Market segmentation
- New product development – by looking at the spatial map the empty spaces represent the unexplored by competitors market segments. The development of a product or a service to fit in the unused space on the spatial map might have a commercial sense.
- Assessing advertising activities and its effectiveness – some marketing campaigns are designed to re-place a brand from one part of the market to another (possibly more profitable or relevant to a company's competitive advantage).
- Distribution channel decisions – managers by looking at the multidimensional map of relevant brands can judge whether creating an additional retail outlet near other brand's outlets is compatible with their branding strategy.
- It can help identify how customers perceive products and brands in terms of attributes, benefits, or features, as well as associations, values, or personality.
- Additionally, MDS can be used to segment consumers based on their similarities or differences in perceptions or preferences for products or brands.
- Furthermore, it can be used to explore how consumers make decisions or choices among products or brands, and what factors or criteria influence them.
- MDS can be used to evaluate current positioning, assess brand equity, discover hidden segments, test hypotheses or theories, examine trade-offs or compromises, and predict or influence consumer behaviour.

4.11 APPLICATIONS OF STATISTICAL SOFTWARE FOR DATA ANALYSIS

There are several software packages applied in data analysis ib business decision making.

4.11.1 MS Excel

There are plenty of statistical functions and data analysis tools available in MS Excel. Descriptive statistics, fitting of distributions and some important parametric tests (t test, F test etc.) can be done with statistical functions. However, by using 'Analysis Tool Pack' of MS Excel, ANOVA for single as well as multiple factors, regression, data summary can be computed directly for the given set of data.

4.11.2 SPSS

SPSS (Statistical Package for the Social Sciences), developed by IBM, offers advanced statistical analysis, a vast library of machine-learning algorithms, text analysis, open-source extensibility, integration with big data and seamless deployment into applications [3].It is one of the widely used statistical packages in the field of social science and human behaviour research. Descriptive statistics, various dimension reduction and classification techniques, parametric as well as non-parametric tests, regression and correlation analysis tests etc. can be performed with ease by dint of this user-friendly software.

4.11.3 R

R is a free software environment specifically meant for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.R along with its vast library help widely to implement time-series analysis, classification techniques, clustering, classical statistical tests, linear and nonlinear modelling etc. Another notable thing about R is its active community, members of which are always engaged in building packages and improving R and the associated plug-ins incessantly. Many of R's standard functions are written in R itself, making it more user-friendly to follow the algorithmic choices made.

4.11.4 SAS

SAS (previously, "Statistical Analysis System") is a software suite developed by SAS Institute for advanced analytics, predictive analytics, multivariate analyses, data management and business intelligence. Using the Output Delivery System, it is possible to publish SAS data in Excel,HTML, PDF and other formats. The SAS Enterprise Guide is SAS's point-and-click interface, generating code with a view to manipulate data or to perform analysis automatically.

4.11.5 Tableau

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It substantially helps to simplify raw data into the easily understandable formats. Visualizations created by Tableau are in the form of dashboards and worksheets. Some best features of Tableau are: data blending, real time analysis and collaboration of data. It is even suitable for non-technical users to create customized dashboards. Tableau has garnered interest among the people from all sectors such as business, research, industries etc.

4.11.6 Minitab

Minitab is a statistical package developed by the researchers of Pennsylvania State University in the year of 1972. It began as a light version of OMNITAB 80. Commands can be executed through either GUI orscripted commands, making it friendly to both novices as well as experts. Few other software, produced by Minitab, Inc., can also be used in conjunction with Minitab.

4.11.7 AMOS

AMOS is an easy-to-use software package intended for structural equation modelling. AMOS stands for **A**nalysis of **Mo**ment **S**tructures. AMOS provides with powerful and easy-to-use structural equation modelling (SEM) software. We can create more realistic models than if you used standard multivariate statistics or multiple regression models alone. Using AMOS, we can specify, estimate, assess, and present our model in an intuitive path diagram to show hypothesized relationships among variables.

4.12 STRUCTURAL EQUATION MODELLING (SEM)

Structural equation modelling may also be defined as a multivariate statistical analysis technique that is used for analyzing structural relationships. This technique may better be explained as a combination of factor analysis and multiple regression analysis.

Structural Equation Modelling is used to analyze the structural relationship between measured variables and latent constructs. Largely preferred by the

researchers Structural Equation Modelling estimates the multiple and interrelated dependence in a single analysis.

To explain in simpler words, two types of variables are used: endogenous variables and exogenous variables. Endogenous variables are equivalent to dependent variables and are equal to the independent variable.

Structural equation models are inclusive of both confirmatory and exploratory modelling. Confirmatory modelling usually starts out with a hypothesis that gets represented in a causal model. The concepts used in the model must then be operationalized to allow testing of the relationships between the concepts in the model.

Each Structural equation model is associated with a graph that represents the causal structure of the model and the form of the linear equations. There is a directed edge from X to Y ($X \rightarrow Y$) if the coefficient of X in the structural equation for Y is nonzero (i.e., X is a direct cause of Y). In addition, there is a bi-directed edge between the error terms εX and εY if and only if the covariance between the error terms is nonzero.

4.13 STRUCTURAL EQUATION MODEL TYPES

Structural Modelling falls into four broad categories. These structural equation models are Path Analysis, Latent Variable Structural Model, Growth Curve Model, and Latent Growth Model.

4.13.1 Path Analysis

Path Analysis, one of the major structural equation models in use is the application of structural equation modelling without latent variables.Path Analysis, another Structural Equation Model type, is an extension of the regression model. In a path analysis model from the correlation matrix, two or more casual models are compared. The path of the model is shown by a square and an arrow, which shows the causation. Regression weight is predicted by the model.

4.13.2 Confirmatory Factor Analysis

Confirmatory Factor Analysis, also known as CFA, is a way ahead of data reduction. CFA is also known within SEM as the measurement model because is the step taken to determine how the factors (ε1 and ε1) are measured by the indicators (x1 to x8).

The Confirmatory Factor Model in SEM treats intelligence as a latent variable which can be measured on the basis of test scores. These are spread out in four areas: reading, writing, math, and analysis.

4.13.3 Latent Variable Structural Model

The next structural equation model for analysis is the Latent Variable Structural Model.Very next step is to fit the structural model, which is what you probably think of when you hear about SEM. It is mainly using the measured latent variables within the path analysis framework. Once you have declared the latent variables you can hypothesize and test their relationships.

4.13.4 Growth Curve Models

Another popular use of Structural Equation Modelling is longitudinal models, commonly referred to as Growth Curve Models. Let's say for instance you have multiple observations of the same variable over time, you may declare an intercept. A slope for the subjects 'papers over time as latent variables by constraining the path coefficients in a specific way.

4.14 DECISION TREE ANALYSIS

A decision tree is a flowchart that starts with one main idea and then branches out based on the consequences of your decisions. It's called a "decision tree" because the model typically looks like a tree with branches.

These trees are used for decision tree analysis, which involves visually outlining the potential outcomes, costs, and consequences of a complex decision. You can use

a decision tree to calculate the expected value of each outcome based on thedecisions and consequences that led to it. Then, by comparing the outcomes to one another, you can quickly assess the best course of action. You can also use a decisiontree to solve problems, manage costs, and reveal opportunities.

4.14.1 Application of Decision Tree

We can use decision tree analysis to make decisions in many areas including operations, budget planning, and project management. Where possible, include quantitative data and numbers to create an effective tree. The more data you have, the easier it will be for you to determine expected values and analyze solutions basedon numbers.

For example, if you're trying to determine which project is most cost-effective, you can use a decision tree to analyze the potential outcomes of each project and choose the project that will most likely result in highest earnings.

4.14.2 Advantages and Disadvantages of Decision Tree Analysis

There are risks and rewards associated with the process of decision tree analysis. The advantages of decision tree analysis include: simple and easy tointerpret decision trees; valuable without requiring large amounts of hard data; helps decision makers ascertain best, worst, and expected results for various scenarios; andcan be combined with various decision techniques.

When using decision tree analysis, there may also be some disadvantages. Disadvantages include: uncertain values can lead to complex calculations and uncertain outcomes; decision trees are unstable, and minor data changes can lead to major structure changes; information gain in decision trees can be biased; anddecision trees can often be relatively inaccurate. A popular alternative to decision treesis the influence diagram, which is a more compact, mathematical graphical representation of a decision situation.

Let us Sum Up:

In this unit we have learnt about the theoretical aspects of multi-variate analysis of data. They are applied for the data sourced from more than two variables. In that way, this study content has provided the various multi-variate techniques generally used in research and data analysis. The meaning and features of the multi-variate techniques like exploratory analysis, confirmatory factor analysis, discriminant function analysis, cluster analysis, conjoint analysis and multi-regression analysis have been explained in this unit. The features of multi-dimensional scaling, an advanced tool is also explained here. The application of multi-dimensional scaling is given. Various statistical software packages used in data analysis have been briefly introduced. The Structural equation Modelling analysis process and types are given for the factor analysis.

Check Your Progress -Quiz

1. **Which of the following is not the part of the exploratory factor analysis process?**
   A. Extracting factors
   B. **D**etermining the number of factors before the analysis
   C. Rotating the factors
   D. Refining and interpreting the factors
2. **CFA differs from EFA in that CFA involves (a) initial hypotheses regarding the number and nature of factors and (b) tests of_____.**
   A. **m**odel-fit
   B. data originality
   C. local control
   D. logical deduction
3. **An examination of difference across groups lies at the heart of the basic concept of_____.**
   A. Regression analysis
   B. Conjoint analysis
   C. **D**iscriminant analysis
   D. Factor analysis
4. **Which of the following is cluster analysis?**

A. Simple segmentation

B. Labeled Classification

C. Query Results Grouping

D. **G**rouping similar objects

5. **Conjoint analysis may be employed to _____**

A. identify important attributes that influence consumer choice

B. estimate market share potential

C. analyze product image

D. **A**ll the above

6. **_____analysis is used to estimate the utility of the levels of various attributes or features of an object, as well as the relative importance of the attributes themselves.**

A. ANOVA

B. Factor

C. **C**onjoint

D. Attribute

7. **A multiple regression model has _____**

A. only one independent variable

B. more than one dependent variable

C. more than one independent variable

D. none of the above

8. **Multidimensional scaling involves at least_____dimensions**

A. **T**hree

B. Two

C. One

D. Zero

9. **SPSS widely used for _____**

A. Image and video editing

B. Mobile app development

C. Web development and programming

D. **D**ata visualization and analysis

10. **Which of the following is not a typical structural equation model?**

A. Confirmatory factor analysis

B. Latent path analysis

C. Latent mean analysis

D. **E**xploratory factor analysis

UNIT SUMMARY

Multivariate Statistical Analysis focuses on advanced statistical techniques used to analyze multiple variables simultaneously, including Exploratory and Confirmatory Factor Analysis (for identifying latent structures), Discriminant Analysis (for classifying data into groups), Cluster Analysis (for grouping similar data points), and Conjoint Analysis (for understanding consumer preferences). It also covers Multiple Regression (for predicting a dependent variable based on several independent variables) and Multidimensional Scaling (for visualizing similarities between data points). Applications in marketing and the use of statistical software for data analysis, including SEM (Structural Equation Modeling), are also emphasized.

GLOSSARY

Keywords	Meaning
Exploratory Factor Analysis (EFA)	A technique used to uncover the underlying structure of a large set of variables by identifying common factors.
Confirmatory Factor Analysis (CFA)	A statistical technique used to test if a set of observed variables fit a hypothesized measurement model.
Discriminant Analysis	A method to predict group membership based on one or more predictor variables.
Cluster Analysis	A technique used to group similar objects or cases into clusters based on their characteristics.
Conjoint Analysis	A statistical technique used in marketing to determine how consumers value different attributes of a product or service.

Self-Assessment Questions:

1. List and explain the various techniques used for multi-variate data analysis
2. Bring out the features and applications of multiple regression analysis
3. What is multidimensional scaling? What are the applications of MDS in

marketing?

4. Enumerate various software packages available for data analysis with their features
5. Explain the meaning and types of SEM analysis
6. Write the types of cluster analysis with suitable diagrams

Activities:

1. Choose any product or service of your choice. Prepare the data relevant to various marketing elements of the product. Use the multidimensional scaling for the data collection and analysis appropriately.
2. There are various software packages available for multi-variate data analysis. Go through original websites of the software packages and learn the tutorials of the same.

ANSWERS FOR CHECK YOUR PROGRESS

1. B) Determining the number of factors before the analysis
2. A) Model-fit
3. C) Discriminant analysis
4. D)Grouping similar objects
5. D) All the above.
6. C) Conjoint
7. C) More than one independent variable
8. B) Two dimensions
9. D) Data visualization and analysis
10. D) Exploratory factor analysis

Suggested Readings:

1. Linda S. Fidell Barbara G. Tabachnick, “Using Multivariate Statistics”, Pearson, 2020
2. Johnson and Wichern, “Applied Multivariate Statistical Analysis’, Pearson, 6th Edition, 2015
3. Joseph F Hair, Barry J. Babin, Rolph E. Anderson and William C. Black, “Multivariate Data Analysis”, Cengage Publishers, 201

UNIT V

REPORT WRITING AND ETHICS IN BUSINESS RESEARCH

REPORT WRITING AND ETHICS IN BUSINESS RESEARCH
Research Reports- Different Types -Report Writing Format- Content of Report- Need For Executive Summary- Chapterisation -Framing the Title of the Report- Different Styles Of Referencing -Academic Vs Business Research Reports - Ethics In Research.

REPORT WRITING AND ETHICS IN BUSINESS RESEARCH

UNIT OBJECTIVES:

- To provide the need, importance and presentation of research work in the form of research report
- To enhance the knowledge on types of research reports, format and content of the research reports
- To exhibit the need and framing of executive summary, chapterisation and title of the research report
- To teach different styles of referencing
- To enumerate the concept of ethics in research with its importance and issues

5.1 INTRODUCTION ABOUT RESEARCH REPORTS

One of the reasons for carrying out research is to add to the existing body of knowledge. Therefore, when conducting research, you need to document your processes and findings in a research report. With a research report, it is easy to outline the findings of your systematic investigation and any gaps needing further inquiry. Knowing how to create a detailed research report will prove useful when you need to conduct research. In many ways, a research report can be considered as a summary of the research process that clearly highlights findings, recommendations, and other important details. Reading a well-written research report should provide you with all the information you need about the core areas of the research process.

Mostly, research work is presented in a written form. The practical utility of research study depends heavily on the way it is presented to those who are expected to act on the basis of research findings. Research report is a written document containing key aspects of research project.

5.2 RESEARCH REPORT

A research report is a document prepared by an analyst or strategist who is a part of the investment research team in a stock brokerage or investment bank. A research report may focus on a specific stock or industry sector, a currency, commodity or fixed-income instrument, or on ageographic region or country. Research reports generally, but not always, have actionable recommendations such as investment ideas that investors can act upon.

Definition

Research report is the systematic, articulate, and orderly presentation of research work in awritten form.

5.3 IMPORTANCE OF RESEARCH REPORT

- **Investigation:** Whenever there is any problem, a committee or commission or study group investigates the problem to find out the reason behind the problem and present the findings with or without the recommendation in the form of a report. It is anotherimportance of report.
- **Evaluation:** Large scale organizations are engaged in multidimensional activities. It is not possible for a single top executive to keep personal watch on what others are doing. So, the executive depends on reports to evaluate the performance of various departments or units.
- **Decision-Making Tool:** Today's complex business organizations require thousands of information. Â reports provide the required information a large number of important decisions in business or any other area are taken on the basis of information presented in the reports. This is one of the great importance of report.
- **Neutral presentation of facts:** Facts are required to be presented in a neutral way; such presentation is ensured through a report as it investigates, explains and evaluates any fact independently.

- **Quick Location:** There is no denying the fact that business executives need information for quick decision-making. As top executives are found to be busy for various purposes), they need vital sources of information. Such sources can be business reports.
- **Development of skill:** Report writing skill develops the power of designing, organization coordination, judgment and communication.
- **Professional Advancement:** Report also plays a major role in professional achievement. For promotion to the rank and file position, satisfactory job performance is enough tohelp a person. But for promotion to high level position, intellectual ability is highly required. Such ability can be expressed through the reportsubmitted to higher authority.
- **Proper Control:** Whether activities are happening according to plan or not is expressed through a report. So, controlling activities are implemented based on the information of a report.
- **A managerial Tool:** Various reports make activities easy for the managers. For planning,organizing, coordinating, motivating and controlling, manager needs help from a report which acts as a source of information.
- **Encountering Advance and Complex Situation: I**n a large business organization, thereis always some sort of labour problems which may bring complex situations. To tackle thatsituation, managers take the help of a report.

5.4 CHARACTERISTICS OF A GOOD RESEARCH REPORT

The characteristics of a good research report are:

- **Simplicity:**

The language shall be as simple as possible so that a report is easily understandable. Jargons and technical words should be avoided. Even in a technical report there shall be restricted use of technical terms if it has to be presented to laymen.

- **Clarity:**

The language shall be lucid and straight, clearly expressing what is intended to be

expressed. Forthat the report has to be written in correct form and following correct steps.

- **Brevity:**

A report shall not be unnecessarily long so that the patience of the reader is not lost and there is no confusion of ideas. But, at the same time, a report must be complete. A report is not an essay.

- **Positivity:**

As far as possible positive statements should be made instead of negative ones. For example, it isbetter to say what should be done and not what should not be done.

- **Punctuation:**

Punctuations have to be carefully and correctly used otherwise the meaning of sentences may be misunderstood or misrepresented.

- **Approach:**

There are two types of approaches: (a) Person–When a report is written based on personal enquiry or observations, the approach shall be personal and the sentences shall be in the first person and in direct speech, (b) Impersonal–When a report is prepared as a source of information and when it is merely factual (e.g. a report on a meeting), the approach shall be impersonal and the sentences shall be in the third personand in indirect speech.

- **Readability:**

The keynote of a report is readability. The style of presentation and the diction (use of words) shall be such that the readers find it attractive and he is compelled to read the report from the beginning to the end.' Then only a report serves its purpose. A report on the same subject matter can be written differently for different classes of readers.

- **Accuracy:**

A report shall be accurate when facts are stated in it. It shall not be biased with personal feelings of the writer.

- **Logical Sequence:**

The points in a report shall be arranged with a logical sequence, step by step and not in a haphazard manner. A planning is necessary before a report is prepared.

- **Proper Form:**

A report must be in the proper form. Sometimes there are statutory forms to follow.

- **Presentation:**

A report needs an attractive presentation. It depends on the quality of typing or printing as wellas quality of paper used. Big companies make very attractive and colorful Annual Reports.

5.5 IMPORTANT CONSIDERATIONS OF PREPARING RESEARCH REPORT

- **Selectiveness:**

It is important to exclude the matter, which is known to all. Only necessary contents should be included to save time, costs, and energy. However, care should be taken that the vital points should not be missed.

- **Comprehensiveness:**

Report must be complete. It must include all the necessary contents. In short, it must contain enough detail to covey meaning.

- **Cost Consideration:**

It must be prepared within the budgeted amount. It should not result into excessive costs.

- **Accuracy:**

As far as possible, research report must be prepared carefully. It must be free from spelling mistakes and grammatical errors.

- **Objectivity:**

Report must be free from personal bias, i.e., it must be free from one's personal liking and disliking. The report must be prepared for impersonal needs. The facts must be stated boldly. It must reveal the bitter truth. It must suit the objectives and must meet expectations of the relevant audience/readers.

- **Clarity:**

Report must reveal the facts clearly. Contents and conclusions drawn must be free from ambiguities. In short, outcomes must convey clear-cut implications.

Preciseness: Research report must not be unnecessarily lengthy. It must contain only necessary parts with adequate description.

- **Simplicity:**

Report must be simple to understand. Unnecessary technical words or terminologies (jargons) should be avoided.

- **Proper Language:**

Researcher must use a suitable language. Language should be selected as per its target users.

- **Reliability:**

Research report must be reliable. Manager can trust on it. He can be convinced to decide on the basis of research reports.

- **Proper Format:**

An ideal report is one, which must be prepared as per commonly used format. One must comply with the contemporary practices; completely a new format should not be used.

- **Attractive:**

Report must be attractive in all the important regards like size, colour, paper quality, etc. Similarly, it should use liberally the charts, diagrams, figures, illustrations, pictures, and multiple colours.

5.6 TYPES OF RESEARCH REPORT

5.6.1 Long Report and Short Report:

These kinds of reports are quite clear, as the name suggests. A two-page report or sometimes referred to as a memorandum is short, and a thirty-page report is absolutely long. But what makes a clear division of short reports or long reports? Well, usually, notice that longer reports are generally written in a formal manner.

Internal and External Report:

As the name suggests, an internal report stays within a certain organization or group of people. Inthe case of office settings, internal reports are for within the organization. We

prepare external reports, such as a news report in the newspaper about an incident or the annual reports of companies for distribution outside the organization. We call these as public reports.

5.6.2 Vertical and Lateral Report:

This is about the hierarchy of the reports' ultimate target. If the report is for your management orfor your mentees, it's a vertical report. Wherever a direction of upwards or downwards comes into motion, we call it a vertical report.

Lateral reports, on the other hand, assist in coordination in the organization. A report traveling between units of the same organization level (for example, a report among the administration andfinance departments) is lateral.

5.6.3 Periodic Report:

Periodic reports are sent out on regularly pre-scheduled dates. In most cases, their direction is upward and serves as management control. Some, like annual reports, is not vertical but is a Government mandate to be periodic in nature.

5.6.4 Formal and Informal Reports:

Formal reports are meticulously structured. They focus on objectivity and organization, contain deeper detail, and the writer must write them in a style that eliminates factors like personal pronouns.

Informal reports are usually short messages with free-flowing, casual use of language. Wegenerally describe the internal report/memorandum as an informal report. For example, a report among your peers, or a report for your small group or team, etc.

5.6.5 Informational Report:

Informational reports (attendance reports, annual budget reports, monthly financial reports, and such) carry objective information from one area of an organization to maybe a larger system.

5.6.6 Proposal Report:

These kinds of reports are like an extension to the analytical/problem-solving reports. A proposal is a document one prepares to describe how one organization can provide a solution to a problemthey are facing. There's usually always a need to prepare a report in a business set-up. The end goal is usually very solution-oriented. We call such kinds of reports as proposal reports.

5.6.7 Functional Report:

These kinds of reports include marketing reports, financial reports, accounting reports, and a spectrum of other reports that provide a function specifically. By and large, wecan include almost all reports in most of these categories. Furthermore, we can include a single report in several kinds of reports.

5.6.8 Descriptive Report:

In descriptive report, researcher describes the facts, trends or opinions experienced or gathered during the research work. In such reports, data presentation and analysis are more importantly presented. Such reports are more suitable in case of describing current situations, etc. It is more popular method of report writing.

5.6.9 Analytical Report:

As name given analytical, such reports are prepared with analyzing and interpretation of the factsor trends or situations. This means analytical report is one step ahead than descriptive reports. Such reports follow the scientific investigation and reporting. Analytical reports also recommend some measures to improve the situation with stating different problems on the situation. Policy research and managerial research which are normally funded by any agencies seeking solution ofprevailing problems demand analytical report.

5.6.10 Technical Report:

In the technical report the main emphasis is on

i. the methods employed,
ii. assumptions made in the course of the study,
iii. the detailed presentation of the findings including their limitations and supporting data.

A general outline of a technical report can be as follows:

- *Summary of results:* A brief review of the main findings just in two or three pages.
- *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.
- *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.
- *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.
- *Analysis of data and presentation of findings:* The analysis of data and presentation of thefindings 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.
- *Conclusions:* A detailed summary of the findings and the policy implications drawn fromthe results be explained.
- *Bibliography:* Bibliography of various sources consulted be prepared and attached.
- *Technical appendices:* Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones.
- *Index:* Index must be prepared and be given invariably in the report at the end.

5.6.11 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.

The General Outline of a Popular Report.

- *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.
- *Recommendations for action:* Recommendations for action on the basis of the findings ofthe study is made in this section of the report.
- *Objective of the study:* A general review of how the problem arise is presented along withthe specific objectives of the project under study.
- *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 partof the report.
- *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.
- *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.

5.7 REPORT WRITING FORMAT:

Research Report Format depends on several relevant variables. One must employ a suitable format to create desirable impression with clarity. Report must beattractive. It should be written systematically and bound carefully. A report must use the format (often called

structure) that best fit the needs and wants of its readers. Normally, following format is suggested as a basic outline, which has sufficient flexibly to meet the most situations.

Research report is divided into three parts as:

I. First Part (Front Pages):

i. Cover page
ii. Title page
iii. Certificate or statement
iv. Index (brief contents)
v. Table of contents (detailed index)
vi. Acknowledgement
vii. List of tables and figures used
viii. Preface/forwarding/introduction
ix. Summary report

II. Main Report (Central Part of Report):

1. Statement of objectives
2. Methodology and research design
3. Types of data and its sources
4. Sampling decisions
5. Data collection methods
6. Data collection tools
7. Fieldwork
8. Analysis and interpretation (including tables, charts, figures, etc.)
9. Findings
10. Limitations
11. Conclusions and recommendations
12. Any other relevant detail

III. Appendix:

i. Copies of forms used
ii. Tables not included in findings
iii. A copy of questionnaire
iv. Detail of sampling and rate of response
v. Statement of expenses
vi. Bibliography – list of books, magazines, journals, and other reports
vii. Any other relevant information

5.8 CONTENTS OF A RESEARCH REPORT

5.8.1 Purpose of study:

Research is one direction-oriented study. He should discuss the problem of his study. He must give background of the problem. He must lay down his hypothesis of the study. Hypothesis is the statement indicating the nature of the problem. He should be able to collect data, analyze it and prove the hypothesis. The importance of the problem for the advancement of knowledge or removed of some evil may also be explained. He must use review of literature or the data from secondary source for explaining the statement of the problems.

5.8.2 Significance of study:

Research is re-search and hence the researcher may highlight the earlier research in new manner or establish new theory. He must refer earlier research work and distinguishhis own research from earlier work. He must explain how his research is different and how his research topic is different and how his research topic is important. In a statement of his problem, he must be able to explain in brief the historical account of the topic and way in which he can make and attempt. In his study to conduct the research on his topic.

5.8.3 Review of Literature:

Research is a continuous process. He cannot avoid earlier research work. He must start with earlier work. He should note down all such research work, published in books, journals or unpublished thesis. He will get guidelines for his research from taking a review of

literature. He should collect information in respect of earlier research work.

5.8.4 Methodology:

It is related to collection of data. There are two sources for collecting data; primary and secondary. Primary data is original and collected in field work, either through questionnaire interviews. The secondary data relied on library work. Such primary data are collected by sampling method. The procedure for selecting the sample must be mentioned. The methodology must give various aspects of the problem that are studied for valid generalization about the phenomena. The scales of measurement must be explained along with different concepts used in the study.

While conducting a research based on field work, the procedural things like definition of universe, preparation of source list must be given. We use case study method, historical research etc. He must make it clear as to which method is used in his research work. When questionnaire is prepared, a copy of it must be given in appendix.

5.8.5 Interpretation of data:

Mainly the data collected from primary source need to be interpreted in systematic manner. The tabulation must be completed to draw conclusions. All the questions are not useful for report writing. One has to select them or club them accordingto hypothesis or objectives of study.

5.8.6 Conclusions/suggestions:

Data analysis forms the main part of the research problem. The information collected in field work is useful to draw conclusions of study. In relation with the objectives of study the analysis of data may lead the researcher to pin point his suggestions. This is the most important part of study. The conclusions must be based on logical and statistical reasoning. The report should contain not only the generalization of inference but also the basis on which the inferences are drawn. All sorts of proofs, numerical and logical, must be given in support of any theory that has been advanced. He should point out the limitations of his study.

5.8.7 Bibliography:

The list of references must be arranged in alphabetical order and be presented in appendix. The books should be given in first section and articles are in second section and research projects in the third. The pattern of bibliography is considered convenient and satisfactory from the point of view of reader.

5.8.8 Appendices:

The general information in tabular form which is not directly used in the analysis of data but which is useful to understand the background of study can be given in appendix.

5.9 EXECUTIVE SUMMARY/ABSTRACT

An executive summary provides an overview of a larger document or research and is usually the first thing your reader will see. Oftentimes, executive summaries are the only place decision makers will go to determine if action is warranted on a particular action or idea.

Executive summaries will analyze a problem, draw conclusions, and recommend a course of action in a complete but brief synopsis. Remember, the people who are reading the summary often do not have much time, so your executive summary must grab their attention and entice them to read through the larger, more in-depth documentation.

5.9.1 IMPORTANCE OF EXECUTIVE SUMMARY

- Although an executive summary is similar to an abstract in that they both summarize the contents of a research study, there are several key differences. With research abstracts, the author's recommendations are rarely included, or if they are, they are implicit rather than explicit.
- Recommendations are generally not stated in academic abstracts because scholars operate in a discursive environment, where debates, discussions, and dialogs are meant to precede the implementation of any new research findings.
- The conceptual nature of much academic writing also means that recommendations arising from the findings are distributed widely and not easily or usefully encapsulated.

- Executive summaries are used mainly when a research study has been developed for an organizational partner, funding entity, or other external group that participated in the research.
- In such cases, the research report and executive summary are often written for policy makers outside of academe, while abstracts are written for the academic community.
- Professors, therefore, assign the writing of executive summaries so students can practice synthesizing and writing about the contents of comprehensive research studies for external stakeholder groups.

5.10 CHAPTERISATION

One of the basic steps in planning a report writing process is to decide how it should be structured. A report should have at least 5 chapters depending on its scope. This is known as the report chapterisation scheme. Deciding on the report chapterisation scheme is important because it helps organise data and critical findings in a legible manner.

5.10.1 FRAMING REPORT CHAPTERISATION

The report chapterisation should be presented while introducing the report's objectives and aims the introduction. It should be placed at the end of the introduction to the report chapter just before the chapter summary. It should range from 200 to 400 words. It can also be presented as a diagram.

However, before framing the chapterisation, decide on all the chapters of your report. The key point to remember while preparing the report chapterisation is to keep it short and informative.

The chapterisation scheme for technical subjects like information technology, information systems, and computer sciences is different. This is because of a more complex analysis like the development and verification of a new algorithm or framework. Such researches involve multiple steps in analysis like reviewing, coding and developing, testing, comparing and finalising a model.

5.11 FRAMING RESEARCH REPORT TITLE

A research report title summarizes the aim and purpose of your research study. Making a title for your research is one of the most important decisions. The research title is the first thing that examiners and reviewers see when they look at your report and the only piece of information that fellow researchers will see in a database or search engine query.

Therefore, when you title research work, make sure it captures all of the relevant aspects of your study, including the specific topic and problem being investigated. It also should present these elements in a way that is accessible and will captivate readers.

In a nutshell, your research title should accurately capture what you have done, it should sound interesting to the people who work on the same or a similar topic, and it should contain the important title keywords that other researchers use when looking for literature in databases.

5.11.1 STEPS IN FRAMING SUITABLE RESEARCH TITLE

- Convey the key research findings: Before writing a research report title, list down what your study is about, what you have achieved or discovered, and the methodology used. Try and identify the one or two key elements that make your study novel or significant in your subject area. Combine these elements to create the best research title that showcases your report accurately and effectively.
- Choose a declarative research paper title: Declarative titles are more informative and help readers to quickly grasp what the body of the article may contain. Therefore, it is considered to be more impactful and more likely to attract the reader's attention.
- A good research title must pique reader interest: Researchers browsing through online platforms during their literature search often spend only a few seconds to read the title and evaluate a report's relevance. This makes it important to create a catchy title for your research report that will spark curiosity in the minds of your audience, which may prompt them pause, read, share, and discuss your research report.
- Avoid making any unsubstantiated claims: This is an important aspect to keep in mind when creating research report titles. While it may be tempting to write titles with claims

that will immediately attract reader attention and get you more citations, your research should be able to back-up these claims with substantive, studied evidence. Failing to do so can create mistrust about the research and even hurt your reputation.

- Keep it simple and avoid jargon: It's tempting to use technical words in a research report's title when you know that your primary audience is most likely to be other researchers working in the same field. However, this can prove counter-productive as readers who are not familiar with these complicated words may end up skipping your article.

5.12 REFERENCES

A references page is the last page of an essay or research paper that's been written in APA style. It lists all the sources you've used in your project so readers can easily find what you've cited.

References can be described as giving credit, with citation, to the source of information used in one's work. Research is a buildup on what other people have previously done thus referencing helps to relate your own work to previous work.

References are a way to provide evidence to support the assertions and claims in your own assignments. References are also a way to give credit to the writers from whom you have borrowed words andideas.

References should always be accurate, allowing your readers to trace the sources of information you have used. The best way to make sure you reference accurately is to keep a record of all the sources you used when reading and researching for an assignment.

5.12.1 IMPORTANCE OF REFERENCES

Referencing is important for a number of reasons, some of which include:

- It allows for acknowledgement of the use of other people's opinions, ideas, theories andinventions.

- Helps readers understand what influenced the writer's thinking and how their ideas wereformulated.
- Helps the readers evaluate the extent of the writer's reading.
- Enables readers to visit source materials for themselves and verify the information.

5.12.2 REFERENCE STYLES

Reference styles are standardized rules for presenting information about the sources used in a text. Typically, a style will describe how to organize information about author(s), publication year, title and page numbers.

There are many different ways to organise the references of a text. Some reference styles follow the author-year format, while others are based on footnotes and/or numerical references.

Some ofthe most commonly used reference styles in academic writing are:

- **American Psychological Association (APA6th) – author-year, commonly used in psychology, economics, educational sciences and health sciences**
- **The Chicago styles (see the free guide at Purdue University or the licensed *Quick guide*):**
  - **Chicago 16 A – footnote system**
  - **Chicago 16 B – author-year, commonly used in the humanities**
- **Harvard – author-year, a general reference system used in many disciplines. or example *Quote, Unquote*.**
- **Modern Language Association (MLA) – author -page number, widely used in linguisticsand**
- **Vancouver – numbered system, commonly used in medicine, health sciences and naturalsciences.**
- **IEEE (Institute of Electrical and Electronics Engineers) – used in, e.g., engineering andcomputer science (cf. the Citation Compass).**

Different academic journals use different reference styles.

5.12.3 KINDS OF INFORMATION NEED TO REFERENCE

Printed books are not the only sources that require acknowledgement. Any words, ideas orinformation taken from Any source requires a reference.

Reference when you are using words or ideas from:

- books and journal articles
- newspapers and magazines
- pamphlets or brochures
- films, documentaries, television programs or advertisements
- websites or electronic resources
- letters, emails, online discussion forums
- personal interviews
- lecturers or tutors. (Not always necessary but check with your lecturer or tutor about theirpreferences before you draw on their ideas.)

You also need to reference when you reprint any diagrams, illustrations, charts or pictures.

5.13 ACADEMIC RESEARCH AND BUSINESS RESEARCH REPORTS

5.13.1 Academic Research

Academic Research is defined as a "*Systematic investigation into a problem or situation, where the intention is to identify facts and/or opinions that will assist in solving the problem or dealing with the situation"*. This academic or scholarly research focuses on research goals/questions that arise from independent researchers. It uses formal, scientific

and systematic procedures to discover answers. The scholarly research is guided by an already existing theory in order to reject or support the theory.

The term 'research' is applied in so many ways in our daily life, from our quest of customer knowledge to writing PhD level research, to exploring a problem at work. Research is a systematic process of collecting, analysing and interpreting information (data) in order to better understand a phenomenon about which we are interested or concerned. It is a lengthy process, focused, specific, intensive, accumulative and educational, and is not mere information gathering, transportation of facts from one location to another and rummaging for information.

5.13.2 Business Research

On the other hand, **business research** is defined as work performed to advance an individual's profession. It is a form of communication produced in a professional manner in order to facilitate work. Business research focuses on research goals/questions that emerge from business requirements. It may or may not use the formal, scientific and systematic procedures to discover answers. It is not grounded in theories and may not require a representative sample.

5.13.3 Differences Between Academic Research and Business Research

The common differences between academic research and professional research are listed below.

Academic Research	Business Research
Also called as Scholarly Research Seeks to add to a larger "body of knowledge"	Also called as Applied Research Seeks to find solutions to instant problems and issues
Questions tend to be more conceptual	Problems tend to be more practical
Theoretically focused	Organizationally focused
Findings are generally made public	Findings are generally kept private

Results generally spur ideas and questions for future research	Results are generally used internally to make decisions and set up strategy
Assessed through peer review by means of academic discipline standards	Assessed by client-organisation and/or industry standards
Shared primarily through academic writings (doctoral dissertation, thesis, dissertation research, scholarly journals, academic conferences & presentations, academic articles and other publications (e.g., books)	Shared mainly through internal reports to reveal results; may also be shared more widely through business conferences and industry/trade publications (e.g., articles, case studies, etc.)

Academic or scholarly research writing concerns more about methodology as it needs to be scientifically validated. It generally circulates within the academy, and has an objective stance, clearly states the importance of the topic, and is managed with sufficient detail so that other academic researchers/scholars may also try to replicate the results. Academic research writing focuses on dissertation research and dissertation writing.

Whereas, business research writing is more focused on the use of the information received and is less worried about methodology, and even validation seems to be natural. It uses particular language to communicate information that is easily understood by its target audience, and it may be managed to inform, instruct, persuade, debate, stimulate, or encourage action.

5.14 ETHICS IN RESEARCH

Research is the pillar of knowledge, and it constitutes an integral part of progress. Ethics are broadly the set of rules, written and unwritten, that govern our expectations of our own and others' behaviour.

Research ethics is a core aspect of the research work and the foundation of research design. Research ethics govern the standards of conduct for scientific researchers. It is important to adhere to ethical principles in order to protect the dignity, rights and welfare of research participants.

Research ethics are the set of ethics that govern how research is performed at research institutions such as universities and how it is disseminated. When most people think of research ethics, they think about issues that arise when research involves human or animal subjects.

While these issues are indeed a key part of research ethics, there are also wider issues about standards of conduct. These include the importance of publishing findings in a transparent way, not plagiarising others' work, and not falsifying work.

5.14.1 IMPORTANCE OF RESEARCH ETHICS

Research ethics are important for a number of reasons.

- They promote the aims of research, such as expanding knowledge.
- They support the values required for collaborative work, such as mutual respect and fairness. This is essential because scientific research depends on collaboration between researchers and groups.
- They mean that researchers can be held accountable for their actions. Many researchers are supported by public money, and regulations on conflicts of interest, misconduct, and research involving humans or animals are necessary to ensure that money is spent appropriately.
- They ensure that the public can trust research. For people to support and fund research, they have to be confident in it.
- They support important social and moral values, such as the principle of doing no harm to others.

5.14.2 ETHICAL ISSUES IN RESEARCH

5.14.2.1 Study design and ethics approval

Good research should be well adjusted, well-planned, appropriately designed, and ethically approved. To conduct research to a lower standard may constitute misconduct. This may appear to be a stringent criterion, but it highlights the basic requirement of a researcher to conduct a research responsibly. To achieve this, a research protocol should be developed and adhered to. It must be carefully agreed to by all contributors andcollaborators, and the precise roles of each team member should be spelled out early, including matters of authorship and publications.

Research should seek to answer specific questions, rather than just collect data. The research proposal should discuss potential ethical issues pertaining to the research. The researchers should pay special attention to vulnerable subjects to avoid breech of ethical codes (e.g. children, prisoners, pregnant women, mentally challenged, educationally and economically disadvantaged). Consent should be explained and obtained from the subjects or guardians, and steps should be taken to ensure confidentiality of information provided by the subjects.

5.14.2.2 Data analysis

It is the responsibility of the researcher to analyse the data appropriately. Although inappropriate analysis does not necessarily amount to misconduct, intentional omission of result may cause misinterpretation and mislead the readers. Fabrication and falsification of data do constitute misconduct. To ensure appropriate data analysis, all sources and methods used to obtain and analyse data should be fully disclosed. Failure to do so may lead the readers to misinterpret the results without considering possibility of the study being underpowered. The discussion section of a paper should mention any issues of bias, and explain how they have been dealt with in the design and interpretation of the study.

5.14.2.3 Authorship

There is no universally agreed definition of authorship. It is generally agreed that an author should have made substantial contribution to the intellectual content, including

conceptualising and designing the study; acquiring, analysing and interpreting the data. The author should also take responsibility to certify that the manuscript represents valid work and take public responsibility for the work. Finally, an author is usually involved in drafting or revising the manuscript, as well as approving the submitted manuscript.

Data collection, editing of grammar and language, and other routine works by itself, do not deserve an authorship. It is crucial to decide early on in the planning of a research who will be credited as authors, as contributors, and who will be acknowledged. It is also advisable to read carefully the "Advice to Authors" of the target journal which may serve as a guide to the issue of authorship.

5.14.2.4 Conflicts of interest

This happens when researchers have interests that are not fully apparent and that may influence their judgments on what is published. These conflicts include personal, commercial, political, academic or financial interest. Financial interests may include employment, research funding, stock or share ownership, payment for lecture or travel, consultancies and company support for staff. Such interests, where relevant, should be discussed in the early stage of research. The researchers need to take extra effort to ensure that their conflicts of interest do not influence the methodology and outcome of the research. It would be useful to consult an independent researcher, or Ethics Committee, on this issue if in doubt. When publishing, these conflicts of interest should be declared to editors, and readers will judge for themselves whether the research findings are trustworthy.

5.14.2.5 Redundant publication and plagiarism

Redundant publication occurs when two or more papers, without full cross reference, share the same hypothesis, data, discussion points, or conclusions. However, previous publication of an abstract during the proceedings of meetings does not preclude subsequent submission for publication, but full disclosure should be made at the time of submission. This is also known as self-plagiarism. In the increasing competitive environment where appointments, promotions and grant applications are strongly influenced by publication

record, researchers are under intense pressure to publish, and a growing minority is seeking to bump up their CV through dishonest means.

On the other hand, plagiarism ranges from unreferenced use of others published and unpublished ideas, including research grant applications to submission under “new” authorship of a complete paper, sometimes in different language. Therefore, it is important to disclose all sources of information, and if large amount of other people’s written or illustrative materials is to be used, permission must be sought.

5.14.2.6 Research Methods

We know there are numerous research methods. However, when it comes to ethical considerations, some key questions can help us find the right approach for our studies.

- Which methods most effectively fit the aims of your research?
- What are the strengths and restrictions of a particular method?
- Are there potential risks when using a particular research method?

5.14.2.7 Voluntary Participation and Consent

An individual should at no point feel any coercion to participate in a study. This includes any type of persuasion or deception in attempting to gain an individual’s trust. Informed consent states that an individual must give their explicit consent to participate in the study. You can think of consent form as an agreement of trust between the researcher and the participants.

5.14.2.8 Validity

The research design must address specific research questions. Hence, the conclusions of the study must correlate to the questions posed and the results. Also, research ethics demands that the methods used must relate specifically to the research questions.

5.14.2.9 Sampling

Sampling is the first step in research design. You will need to explain why you want a particular group of participants. You will have to explain why you left out certain people or groups. In addition, if your sample includes children or special needs individuals, you will have additional requirements to address like parental permission.

5.14.3 ETHICAL CODES TO BE FOLLOWED IN RESEARCH

5.14.3.1 Honesty and Integrity

This means that you need to report your research honestly, and that this applies to your methods (what you did), your data, your results, and whether you have previously published any of it. You should not make up any data, including extrapolating unreasonably from some of your results, or do anything which could be construed as trying to mislead anyone. It is better to undersell than over-exaggerate your findings.

When working with others, you should always keep to any agreements, and act sincerely.

5.14.3.2 Objectivity

You should aim to avoid bias in any aspect of your research, including design, data analysis, interpretation, and peer review. For example, you should never recommend as a peer reviewer someone you know, or who you have worked with, and you should try to ensure that no groups are inadvertently excluded from your research. This also means that you need to disclose any personal or financial interests that may affect your research.

5.14.3.3 Carefulness

Take care in carrying out your research to avoid careless mistakes. You should also review your work carefully and critically to ensure that your results are credible. It is also important to keep full records of your research. If you are asked to act as a peer reviewer, you should take the time to do the job effectively and fully.

5.14.3.4 Openness

You should always be prepared to share your data and results, along with any new tools that you have developed, when you publish your findings, as this helps to further knowledge and advance science. You should also be open to criticism and new ideas.

5.14.3.5 Recognition for Intellectual Property

You should never plagiarise, or copy, another people's work and try to pass it off as your own. You should always ask for permission before using other people's tools or methods, unpublished data or results. Not doing so is plagiarism. Obviously, you need to respect copyrights and patents, together with other forms of intellectual property, and always acknowledge contributions to your research. If in doubt, acknowledge, to avoid any risk of plagiarism.

5.14.3.6 Confidentiality

An important ethical principle of research is that the confidentiality of the information supplied by research subjects and the anonymity of respondents must be respected. However, sometimes confidentiality is limited. For example, if a participant is at risk of harm, we must protect them. This might require releasing confidential information. You should respect anything that has been provided in confidence. You should also follow guidelines on protection of sensitive information such as patient records.

5.14.3.7 Responsible Publication

You should publish to advance to state of research and knowledge, and not just to advance your career. This means, in essence, that you should not publish anything that is not new, or that duplicates someone else's work.

5.14.3.8 Legality

You should always be aware of laws and regulations that govern your work, and be sure that you conform to them.

5.14.3.9 Human Subjects Protection

If your research involves people, you should make sure that you reduce any possible harm to the minimum, and maximise the benefits both to participants and other people. This

means, for example, that you should not expose people to more tests than are strictly necessary to fulfil your research aims. You should always respect human rights, including the right to privacy and autonomy. You may need to take particular care with vulnerable groups, which include, but are not limited to, children, older people, and those with learning difficulties.

5.14.3.10 Applying for Ethical Approval

Applications for ethical approval will differ across institutions. Regardless, they focus on the benefits of your research and the risk to benefit ratio concerning participants. Therefore, you need to effectively address both in order to get ethical clearance. Participants

It is vital that you make it clear that individuals are provided with sufficient information in order to make an informed decision on their participation. In addition, you need to demonstrate that the ethical issues of consent, risk of harm, and confidentiality are clearly defined.

Benefits of the Study You need to prove to the panel that your work is essential and will yield results that contribute to the scientific community. For this, you should demonstrate the following:

- The conduct of research guarantees the quality and integrity of results.
- The research will be properly distributed.
- The aims of the research are clear and the methodology is appropriate.

5.14.3.11 Integrity

Integrity and transparency are vital in the research. Ethics committees expect you to share any actual or potential conflicts of interest that could affect your work. In addition, you have to be honest and transparent throughout the approval process and the research process.

Let's Sum Up:

In this unit we have discussed about the research report and its importance in presentation. We have covered the characteristics of a research report and considerations while preparing the research report. The unit has explained the different types of research reports

and content of a research report. The writing of executive summary, the framing of chapter scheme and the framing of title of a research are described in this units. The different reference styles are given with their format. The differences between academic report and business report are clearly explained. The ethics in research is described with the areas of ethics and codes to be followed in doing research.

Check your Progress – Quiz

1. The last stage of research process is _____

A. Review of literature

B. Report writing

C. Research design

D. Analysis of data

2. An executive summary contains _____

A. a brief summary of research problems

B. a brief summary of the findings of the report

C. a brief analysis of data

D. a brief interpretation of data

3. The first page of research report is _____

A. Appendix

B. Bibliography

C. Index

D. Title page

4. The content of the research report can be classified into_____of the following parts.

A. Front pages

B. Middle Part

C. appendix

D. All the above

5. Which section of research report presents the theoretical rationale for the research?

A. Introduction

B. Literature review

C. Methods

D. Findings

6. In presentation of Social Science report, particularly business studies reports, which style of reference is used?

A. APA

B. MLA

C. IEEE

D. Chicago

7. What does the 'Ethical Considerations' section in a research report typically discuss?

A. Ethical principles followed in the study and treatment of human or animal subjects

B. Theoretical framework of the study

C. Statistical analyses performed in the study

D. The limitations of the study

8. In research reporting, the "Literature Review" section typically focuses on _____

A. Previous research relevant to the current study

B. Presenting the raw data collected in the study

C. Acknowledging the limitations of the study

D. Describing the research methods used in the study

9. What should be included in the "Appendix" section of a research report?

A. Supplementary materials that provide more detailed information

B. Summary of the study's key findings

C. List of references and citations

D. Main research methods and techniques

10. A Set of principles to guide and assist researchers in deciding which goals are most important and in reconciling conflicting values when conducting research is called _____

A. Research ethics

B. Deontological approach

C. Utilitarianism

D. None of the above

UNIT SUMMARY

Report Writing and Ethics in Business Research covers the essentials of creating research reports, including understanding different types and formats, structuring the content, and the importance of an executive summary. It emphasizes chapterization for logical organization, framing a precise report title, and employing various referencing styles. The unit distinguishes between academic and business research reports and underscores the importance of ethical considerations in research practices.

GLOSSARY

Keywords	Meaning
Research Reports	Written documents presenting the methodology, findings, and conclusions of a research study.
Framing the Title of the Report:	Crafting a clear and concise title that reflects the main focus of the research.
Chapterisation:	Dividing the report into logical sections or chapters to structure the flow of information.
Content of Report:	The essential components of a report, including background, data, analysis, and recommendations.
Ethics in Research	Principles ensuring integrity, transparency, and responsibility in conducting and reporting research.

Self-Assessment Questions:

1. What are the different types of research report? Explain them with suitable illustration?
2. Discuss the contents and format of a research report
3. Describe the need and importance of executive summary, framing of chapter schemes and choosing suitable title for a research
4. Differentiate academic report and business report
5. Enumerate the ethical issues and codes to be followed in research

Activities for this unit:

1. Consider few areas of academic research like science, social science, professional degrees and health science. List out the format and content of the research reports for these fields of academic
2. Collect and exhibit different refence styles followed in different journals by visitingtheir websites

ANSWERS FOR CHECK YOUR PROGRESS

1. B) Report writing

2. B) A brief summary of the findings of the report

3. D) Title page

4. D) All the above

5. B) Literature review

6. A) APA

7. A) Ethical principles followed in the study and treatment of human or animal subjects

8. A) Previous research relevant to the current study

9. A) Supplementary materials that provide more detailed information

10. A) Research ethics

Suggested readings:

1. Cooper, D.R., Schindler, P. and Sharma, J.K., Business Research Methods,11th Edition, Tata-McGraw Hill, 12 th Edition, 2018.
2. Ranjith Kumar, Research Methodology: A Step-by-Step guide for Beginners, Sage, South Asia, 4th Edition, 2023.
3. Kothari C.R. & Gaurav Garg, Research Methodology: Methods and Techniques, New Age International Publishers, 2023