SCHEME : K

Name : _____

Roll No. : _____ Year : 20___ 20___

Exam Seat No. : _____

LABORATORY MANUAL FOR

STATISTICAL MODELLING FOR MACHINE LEARNING

(313307)

ARTIFICIAL INTELLIGENCE GROUP

MAHARASHTRA STATE BOARD OF TECHNICAL EDUCATION, MUMBAI (Autonomous) (ISO 9001: 2015) (ISO/IEC 27001:2013)

VISION

To ensure that the Diploma level Technical Education constantly matches the latest requirements of Technology and industry and includes the all-round personal development of students including social concerns and to become globally competitive, technology led organization.

MISSION

To provide high quality technical and managerial manpower, information and consultancy services to the industry and community to enable the industry and community to face the challenging technological & environmental challenges.

Quality Policy

We, at MSBTE are committed to offer the best-in-class academic services to the students and institutes to enhance the delight of industry and society. This will be achieved through continual improvement in management practices adopted in the process of curriculum design, development, implementation, evaluation and monitoring system along with adequate faculty development programmes.

Core Values

MSBTE believes in the following:

- Skill development in line with industry requirements.
- Industry readiness and improved employability of Diploma holders.
- Synergistic relationship with industry.
- Collective and Cooperative development of all stakeholders.
- Technological interventions in societal development.
- Access to uniform quality technical education.

A Practical Manual

for

STATISTICAL MODELLING FOR MACHINE LEARNING

(313307)

Semester– (III)

"K-SCHEME"

(AI/AN/DS)

Maharashtra State

Board of Technical Education, Mumbai

(Autonomous) (ISO-9001-2008) (ISO/IEC 27001:2013)

Maharashtra State
Board of Technical Education, Mumbai

(Autonomous) (ISO-9001-2008) (ISO/IEC 27001:2013)

4th Floor, Government Polytechnic Building, 49, Kherwadi,
Bandra (East), Mumbai -400051.

(Printed on: July 2024)

Maharashtra State
Board of Technical Education

Certificate

This is to certify that Mr. / Ms. ……. Roll No…….of Second Semester of Diploma in ……of Institute ……..
(Code……..) has completed the term work satisfactorily in course Statistical Modelling For Machine Learning **(313307)** for the academic year 20…….to 20…..... as prescribed in the curriculum.

Place ……. Enrollment No……
Date:…..... Exam Seat No. …….....

Course Teacher **Head of the Department** **Principal**

Seal of the Institute

Preface

The objective of all engineering laboratories or field experience in the technical education system is to help students acquire the critical competencies and skills that businesses demand. In light of this, MSBTE developed the cutting-edge "K" Scheme curricula for engineering diploma programs, emphasizing outcome-based learning and the National Education Policy 2020 (NEP2020). As a result, a sizable portion of the program is dedicated to practical work. This demonstrates how crucial laboratory work is in helping teachers, instructors, and students understand that every minute of lab time must be used efficiently to create these outcomes rather than being spent on pointless tasks. Consequently, each practical has been created to operate as a "vehicle" to advance this industry in order to ensure the successful implementation of this outcome-based curriculum. It is challenging to teach practical skills using only the "chalk and duster" activity. Because of this, the "K" scheme laboratory manual creation team focused on the outcomes when designing the practical rather than following the long-standing custom of doing the practical to "verify the theory" (which may turn out to be a by-product along the way).

This lab manual is intended to support all parties involved, particularly the students, instructors, and teachers, in helping the students achieve the pre-established objectives. It is required of every student to read through the relevant practical process in its entirety and comprehend the bare minimum of theoretical background related to the practical at least one day in advance of the practical. As a crucial starting point for carrying out the practical, each exercise in this manual starts with establishing the competency, industry-relevant skills, course outcomes, and practical outcomes. The skills, that students will acquire from the process outlined there, together with the necessary safety measures to be followed, will subsequently be made clear to them. These will enable them to apply the knowledge and abilities to solve real-world problems in their professional lives.

Statistic, Probability, Interpolation and sampling methods are the core components of AI/ML. Statistics is used in data analytics to identify patterns, trends, and relationships in the data. Integrating Statistics in AI can improve the accuracy and precision of the predictions made by AI systems.After going through these learning experiences students will be able to implement mathematical concepts using R-Programming which will enhance the knowledge and skills to use the methodology for solving AI/ML based problems of various domains.

The team responsible for developing the Practical manual would like to express its gratitude toMSBTE for taking the lead in developing and implementing the curriculum. Additionally, the team recognizes the valuable contributions made by individual course experts who have been directly or indirectly involved in the development of the "K" scheme curriculum and the laboratory manual. It is impossible to claim perfection in this laboratory manual, even though every effort has been made toverify it for errors, especially because this is the first edition. Any such mistakes and recommendations for enhancements are quite appreciated and can be brought to our attention.

Lab Manual Development Team

Programme Outcomes (POs) to be achieved through Practical of this Course

Following POs are expected to be achieved through the practical's of the Statistical Modelling for Machine Learning course.

PO1 : **Basic and Discipline specific knowledge:** Apply knowledge of basic mathematics, science and engineering fundamentals and engineering specialization to solve the engineering problems.

PO2 : **Problem analysis:** Identify and analyze well-defined engineering problems using codified standard methods.

PO3 : **Design/ development of solutions:** Design solutions for well-defined technical problems and assist with the design of systems components or processes to meet specified needs.

PO4 : **Engineering Tools, Experimentation and Testing:** Apply modern engineering tools and appropriate technique to conduct standard tests and measurements.

PO5 : **Engineering practices for society, sustainability and environment:** Apply appropriate technology in context of society, sustainability, environment and ethical practices.

PO6 : **Project Management:** Use engineering management principles individually, as a team member or a leader to manage projects and effectively communicate about well-defined engineering activities in diverse and multidisciplinary fields.

PO7 : **Life-long learning:** Ability to analyze individual needs and engage in updating in the context of technological changes.

List of Industry Relevant Skills-

The following industry relevant skills of the competency 'Apply Mathematics to solve real-world problems uses AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency are expected to be developed in you by undertaking the practical of this laboratory manual.

- Develop critical thinking and decision making skills.
- Develop skills related to Data Analysis and Prediction.
- Develop skills to make inferences about populations.

Practical- Course Outcome matrix

Course Outcomes (COs)

CO1 - Solve the given problem based on Statistic Techniques using R-Programming.
CO2 - Implement Statistic methods using R-Programming.
CO3 - Use Principles of Probability to solve the given Problem.
CO4 - Implement appropriate method based on the Interpolation.
CO5 - Apply Sampling Methods to solve given problem using R-Programming.

S. No.	Laboratory Practical Titles	CO1	CO2	CO3	CO4	CO5
1	* Download R Studio. Install and configure R studio and R Packages.	√	-	-	-	-
2	Write Program to- a. Demonstrate the use of R-Numbers. (numeric, integer, complex). b. Convert number from one type to other using functions. c. Perform following operations. i. Addition and Subtraction on numbers. ii. Find Square root using of number using built-in function.	√	-	-	-	-
3	Write Program to- a. Print any built-in data set of R. b. Get information about the data set. c. Find the dimensions of the data set and view the names of the variables. Hint: Use dim( ) and names( ) function. d. Find the name of each row in the first column. Hint: Use the rownames( ) function. e. Print all values that belong to a variable. f. Sort the values of variable. g. Get the statistical summary of the data.	√	-	-	-	-
4	Write a program to- a. Find the lowest or highest value in a data set. Hint: Use min( ) and max( ) functions. b. Find the index position of the max and min value in the table. Hint: use which.max( ) and which.min( ) functions.	√	-	-	-	-
5	* Write programs to calculate Measures of Central tendency. a. Import data into R. b. Calculate the Mean (Average value) of a variable from the given data set. c. Find the Median (Mid-Point value) of the variable from the given data set. d. Calculate the mode for the variable from the given	√	-	-	-	-

S. No.	Laboratory Practical Titles	CO1	CO2	CO3	CO4	CO5
	data set. (by sorting the column of the dataframe and by using the 'modest' package). e. Calculate the Percentile of the variable from the given data set.					
6	* Write programs to- a. Print Original Data Frame, Modified Frequency Table, Cumulative Frequency Table, Relative Frequency Table. b. Create the Frequency Table by using multiple arguments. c. Plot the frequency table using ggplot function.	√	-	-	-	-
7	* Write programs to calculate-Variance, Standard Deviation, Range, Mean Deviation for the given data.	√	-	-	-	-
8	* Write Programs to graphically represent mode and median of the given data. a. Draw Histogram for the given data. b. Draw Ogive Curve for the given data.	√	-	-	-	-
9	* Write a Program to calculate Skewness for the given data.	√	-	-	-	-
10	Write a Program to draw a scatter plot for two variables for the given dataset.	√	-	-	-	-
11	Write Program to perform the correlation test to evaluate the association between two or more variables. a. Install and load required R packages. b. Compute correlation in R. c. Visualize your data using scatter plots. d. Perform Preliminary test to check the test assumptions.	-	√	-	-	-
12	* Write Program to perform the correlation test to evaluate the association between two or more variables. a. Pearson correlation test. b. Interpretation of the result. c. Use Spearman rank correlation coefficient to estimate a rank-based measure of association.	-	√	-	-	-
13	Write a Program based on Line of Regression using two variables.	-	√	-	-	-
14	* Write Programs to- a. Calculate the probability of getting heads when flipping a fair coin. b. Calculate the probability of drawing a spade from a standard deck of 52 cards.	-	-	√	-	-

S. No.	Laboratory Practical Titles	CO1	CO2	CO3	CO4	CO5
15	Write a program to calculate the probability of rolling a 3 on a fair six-sided die.	-	-	√	-	-
16	Write a program to compute probability using prob() function.	-	-	√	-	-
17	Write a program to calculate the conditional probability.	-	-	√	-	-
18	* Write a program to use Bayes' Theorem in R-Programming.	-	-	√	-	-
19	* Write a Program to interpolate using Newton forward interpolation.	-	-	-	√	-
20	* Write a Program to interpolate using Newton backward interpolation.	-	-	-	√	-
21	* Write a program for the implementation of extrapolation.	-	-	-	√	-
22	Write a program to generate Samples using the Sampling Functions.	-	-	-	-	√
23	Write programs to perform following types of sampling- a. Simple Random Sampling. b. Stratified Sampling. c. Systematic Sampling. d. Biased Sampling.	-	-	-	-	√
24	* Write a program to generate a Sampling Distribution proportion.	-	-	-	-	√
25	* Write a program based on t-Distribution using dt, pt, qt&rt functions.	-	-	-	-	√
26	* Write a program based on Chi-Square Distribution using dchisq, pchisq, qchisq&rchisq functions.	-	-	-	-	√

Guidelines to Teachers

1. There will be two sheets of blank pages after every practical for the student to report other matters (if any), which is not mentioned in the printed practical.
2. For difficult practical if required, teacher could provide the demonstration of the practical emphasizing of the skills which the student should achieve.
3. Teachers should give opportunity to students for hands-on after the demonstration.
4. Assess the skill achievement of the students and COs of each unit.
5. One or two questions ought to be added in each practical for different batches. For this teacher can maintain various practical related question banks for each course.
6. For effective implementation and attainment of practical outcomes, teacher ought to ensure that in the beginning itself of each practical, students must read through the complete write-up of that practical sheet.
7. During practical, ensure that each student gets chance and takes active part in taking observations/ readings and performing practical.
8. Teacher ought to assess the performance of students continuously according to the MSBTE guidelines

Instructions for Students

1. For incidental writing on the day of each practical session every student should maintain a ***dated log book*** for the whole semester, apart from this laboratory manual which s/he has to ***submit for assessment to the teacher*** in the next practical session.
2. For effective implementation and attainment of practical outcomes, in the beginning itself of each practical, students need to read through the complete write-up including the practical related questions and assessment scheme of that practical sheet.
3. Student ought to refer the reference books, lab manuals and e-learning material etc.
4. Student should not hesitate to ask any difficulties they face during the conduct of practical.

Content Page

List of Practical and Progressive Assessment Sheet

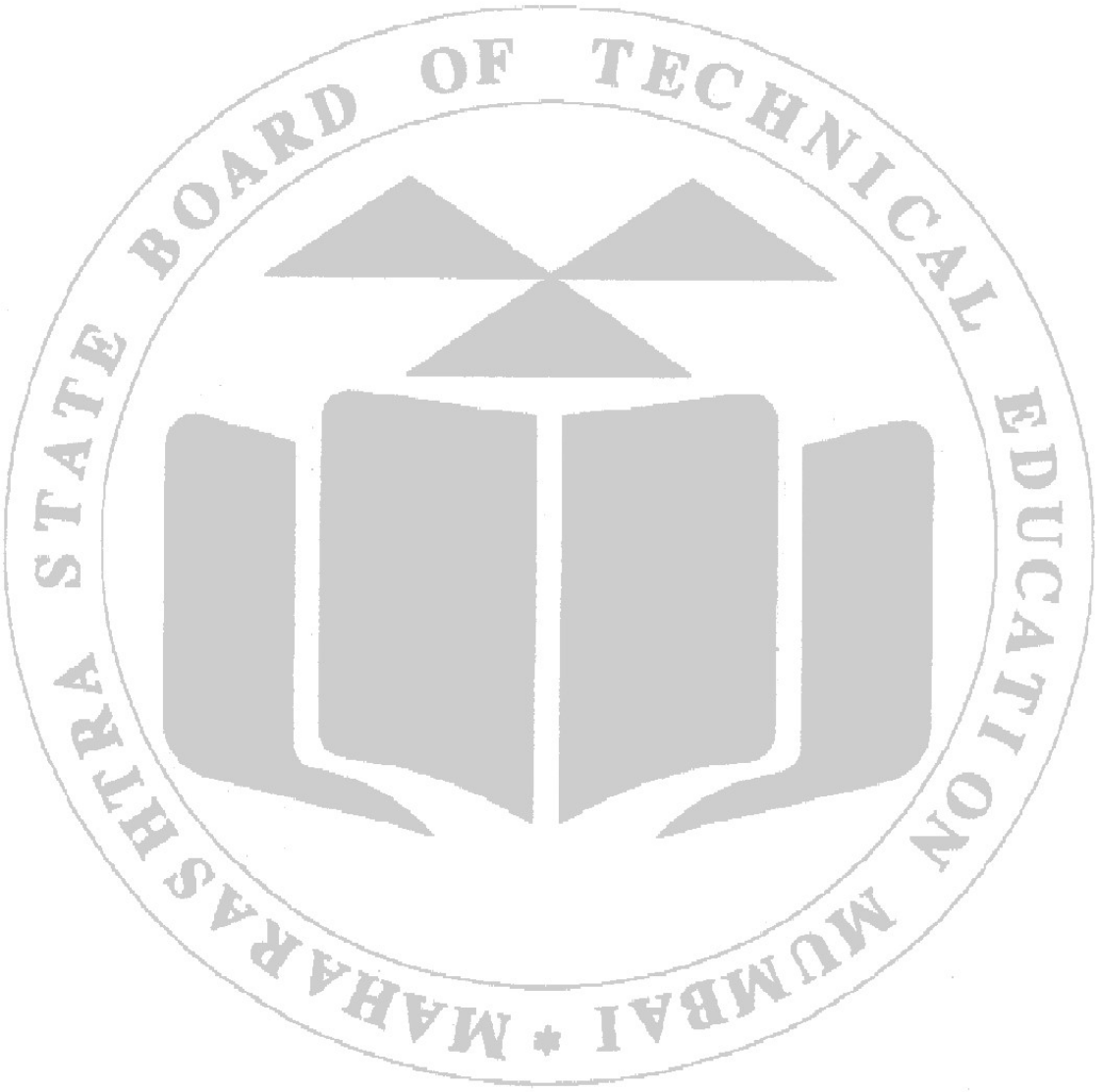
S. N.	Laboratory Practical Titles	Page No.	Date of performance	Date of submission	FA PR marks (25)	Dated sign. of teacher	Remarks (if any)
	column of the dataframe and by using the 'modest' package). e. Calculate the Percentile of the variable from the given data set.						
6	* Write programs to- a. Print Original Data Frame, Modified Frequency Table, Cumulative Frequency Table, Relative Frequency Table. b. Create the Frequency Table by using multiple arguments. c. Plot the frequency table using ggplot function.						
7	* Write programs to calculate-Variance, Standard Deviation, Range, Mean Deviation for the given data.						
8	* Write Programs to graphically represent mode and median of the given data. a. Draw Histogram for the given data. b. Draw Ogive Curve for the given data.						
9	* Write a Program to calculate Skewness for the given data.						
10	Write a Program to draw a scatter plot for two variables for the given dataset.						
11	Write Program to perform the correlation test to evaluate the association between two or more variables. a. Install and load required R packages. b. Compute correlation in R. c. Visualize your data using scatter plots. d. Perform Preliminary test to check the test assumptions.						
12	* Write Program to perform the correlation test to evaluate the association between two or more variables. a. Pearson correlation test. b. Interpretation of the result. c. Use Spearman rank correlation coefficient to estimate a rank-based measure of association.						
13	Write a Program based on Line of Regression using two variables.						
14	* Write Programs to- a. Calculate the probability of getting						

S. N.	Laboratory Practical Titles	Page No.	Date of performance	Date of submission	FA PR marks (25)	Dated sign. of teacher	Remarks (if any)
	heads when flipping a fair coin. b. Calculate the probability of drawing a spade from a standard deck of 52 cards.						
15	Write a program to calculate the probability of rolling a 3 on a fair six-sided die.						
16	Write a program to compute probability using prob() function.						
17	Write a program to calculate the conditional probability.						
18	* Write a program to use Bayes' Theorem in R-Programming.						
19	* Write a Program to interpolate using Newton forward interpolation.						
20	* Write a Program to interpolate using Newton backward interpolation.						
21	* Write a program for the implementation of extrapolation.						
22	Write a program to generate Samples using the Sampling Functions.						
23	Write programs to perform following types of sampling- a. Simple Random Sampling. b. Stratified Sampling. c. Systematic Sampling. d. Biased Sampling.						
24	* Write a program to generate a Sampling Distribution proportion.						
25	* Write a program based on t-Distribution using dt, pt, qt&rt functions.						
26	* Write a program based on Chi-Square Distribution using dchisq, pchisq, qchisq & rchisq functions.						
Total							

Note: To be transferred to Proforma of CIAAN-2023.

Note : Out of above suggestive LLOs -

- **'*' Marked Practical's (LLOs) Are mandatory.**
- **Minimum 80% of above list of lab experiment are to be performed.**
- **Judicial mix of LLOs are to be performed to achieve desired outcomes.**


MAHARASHTRA STATE BOARD OF TECHNICAL EDUCATION MUMBAI

Practical No.1

Download R Studio. Install and configure R studio and R Packages

I. Practical Significance

RStudio is an important tool for everyone who works with the R programming language. It is used in data analysis to import, access, transform, explore, plot, and model data, and for machine learning to make predictions on data. This practical is useful for students to understand procedure of Installation of RStudio and R Packages.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Install RStudio and RPackages'.

III. Course Level Learning Outcome (CO)

CO1 - Solve the given problem based on Statistic Techniques using R-Programming.
CO2 - Implement Statistic methods using R-Programming.
CO3 - Use Principles of Probability to solve the given Problem.
CO4 - Implement appropriate method based on the Interpolation.
CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 1.1 Download and Install R Studio and R Packages.

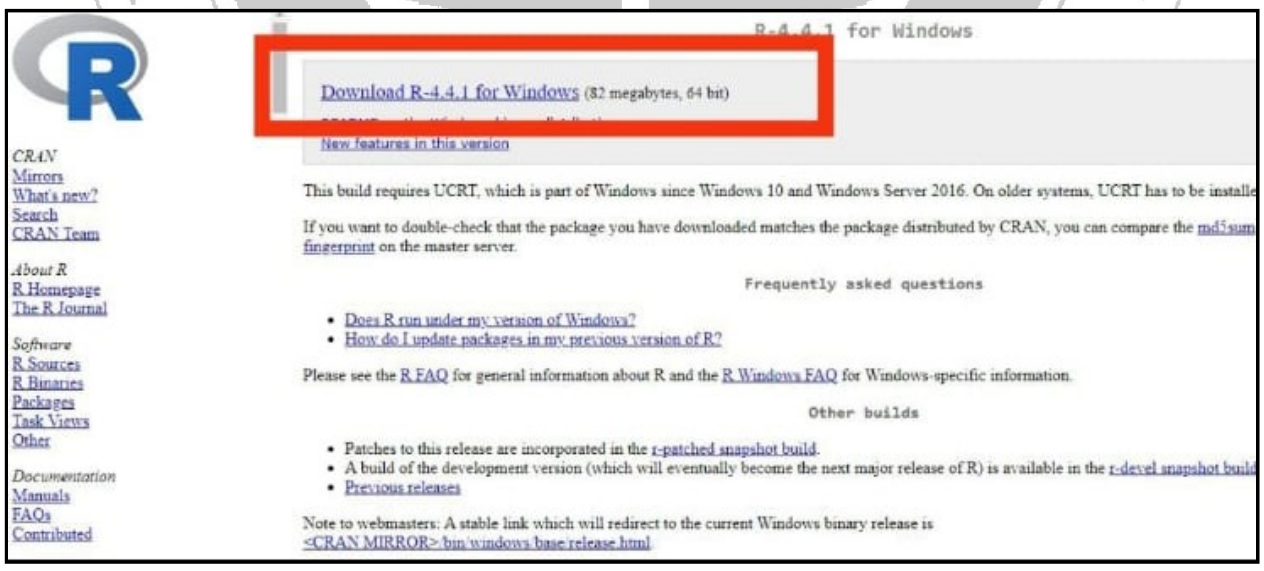
V. Relative Affective Domain related Outcome(s)-

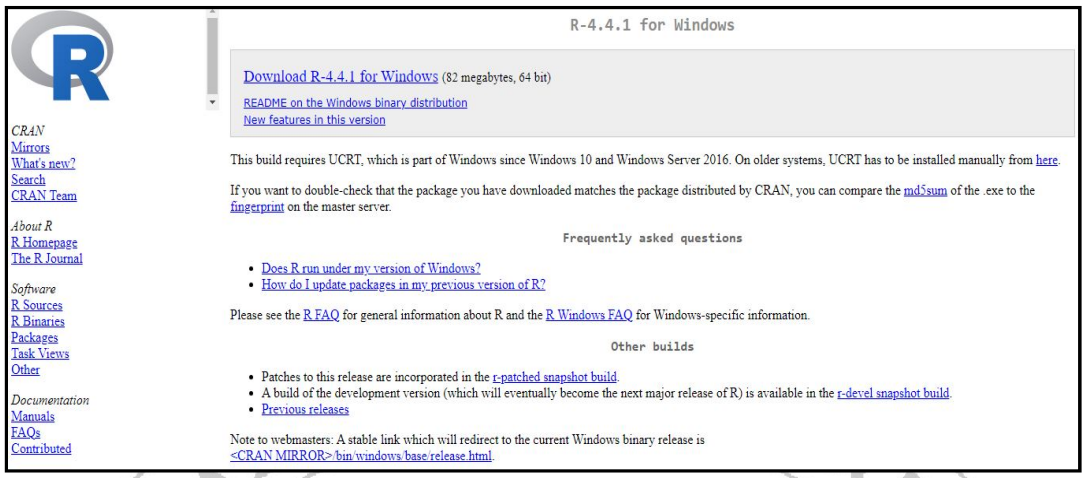
- Follow safety practices.
- Manage R studio and R packages.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram (if required)

Steps to Download and Install R and R Studio

Step 1: Download R for Windows platform.





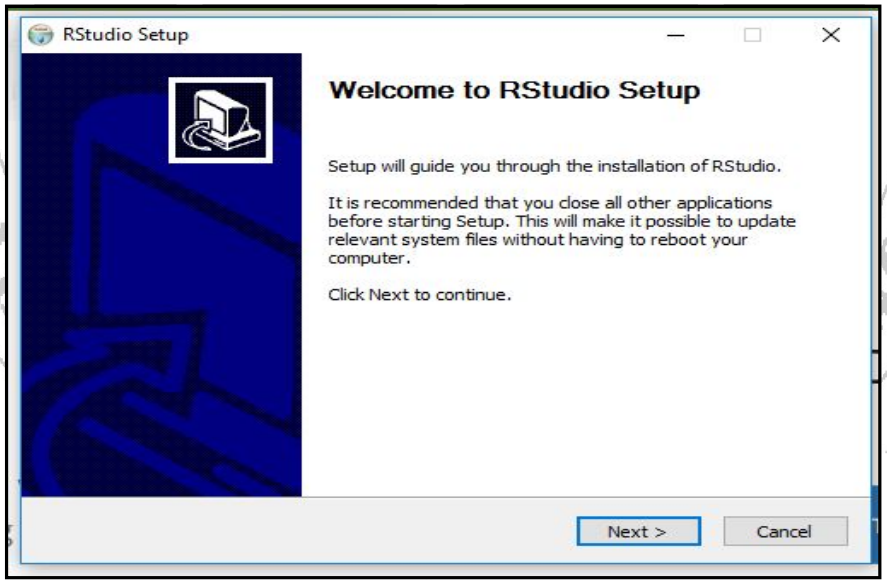
Step 2: After downloading R for the Windows platform, install it by double-clicking it.

Step 3: Download R Studio from their official page.

Step 4: After downloading, you will get a file named "RStudio-1.x.xxxx.exe" in your Downloads folder.

Step 5: Double-click the installer, and install the software.

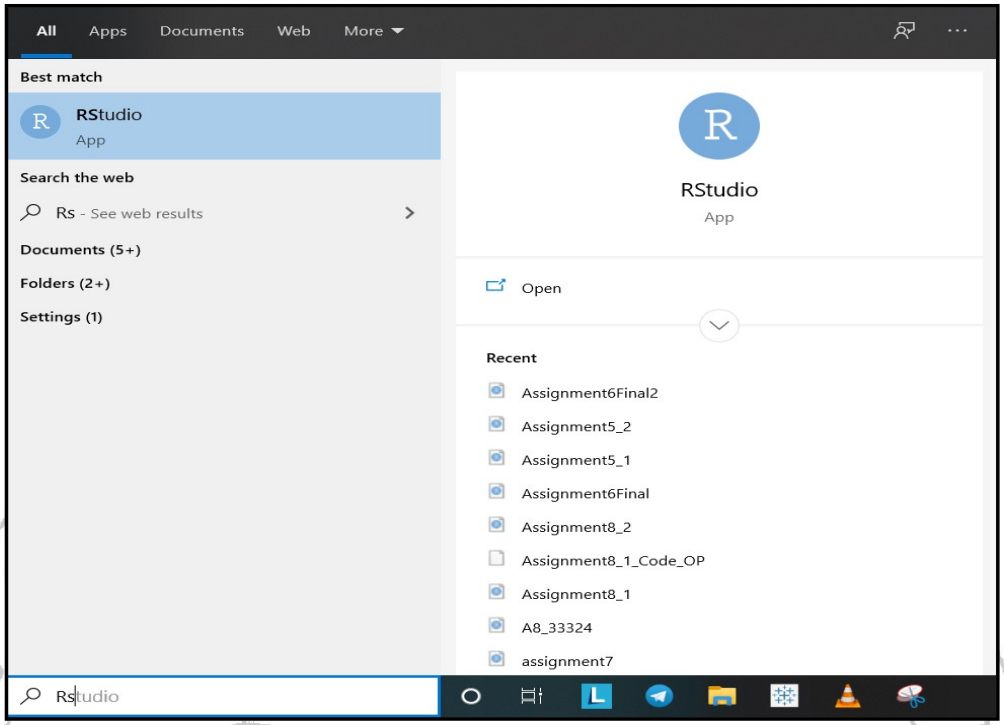
Step 6: Download and Install RStudio for Windows.



Step 7 : Once the packet has downloaded, the Welcome to RStudio Setup Wizard will open. Click Next and go through the installation steps.

Step 8: Test the R Studio installation.

- Search for RStudio in the Window search bar on Taskbar.



VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

- Write R program to import data using following functions:
  a. read.table()
  b. read.csv()
  c. read.csv2()
  d. read.delim()
  e. read.delim2()
  f. read_excel()
- Write steps to configure R Studio to work with multiple R versions.

(Spaceforanswers)

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write the system requirements for installing R Studio.
2. Write command/s to install a package from CRAN in R.
3. Write command/s to check if a package is installed and load in R.
4. Write command/s to install packages in a specific library directory.
5. Write command/s to install a package from GitHub in R.
6. Write command/s to remove a package from R.
7. Write command/s to view which packages are installed and find their versions.

(Spaceforanswers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/how-to-install-r-studio-on-windows-and-linux/
2. https://www.geeksforgeeks.org/packages-in-r-programming/
3. https://www.youtube.com/watch?v=9SzKJH93t5o
4. https://www.datacamp.com/tutorial/r-data-import-tutorial
5. https://www.youtube.com/watch?v=2_tW7e4e_dM

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10 Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25 Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.2

Write Programs to demonstrate the use of R-Numbers and perform operations on it

I. Practical Significance

In statistics, R-numbers (or R-values) are used to denote correlations between variables. This will help students understand the strength and direction of relationships between data sets.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Interpret data and make informed decisions.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 2.1 Implement basic mathematical operations in R programming.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

In R programming, numbers can be of various types, including numeric, integer, and complex.

1. Numeric

The default type for numbers in R is numeric, which usually means double precision floating-point numbers.

Exmple:

```
# Numeric (default)
num_default<- 3.14

# Precision and scientific notation
num_large<- 1e+10
```

2. Integer

In R programming, integers are explicitly defined by appending an L to the number. This indicates that the number is an integer, not a numeric (double).

Example:

```
# Integer
num_integer<- 42L
```

3. Complex

Complex numbers in R have both real and imaginary parts. They are defined using 1i to represent the imaginary unit.

Example:

```
# Complex number
num_complex<- 3 + 4i
```

Type Conversion:

In R programming, numbers can be converted between different types using various functions.

Conversion Functions

as.numeric()	Converts to numeric (double) type.
as.integer( )	Converts to integer type.
as.complex( )	Converts to complex type.
as.character()	Converts to character type, which can be useful for conversion back to other types.

Following example shows the conversion from Integer to numeric type. Students can use other conversion functions similarly.

```
# From Integer to Numeric
int_value<- 42L
num_value<- as.numeric(int_value)
print(num_value)                    # Output: 42
print(typeof(num_value))            # Output: "double"
```

Basic Arithmetic operations:

Basic arithmetic operations can be performed on R Numbers like addition, subtraction, multiplication, and division. Example for addition is given below. Students can perform all other arithmetic operations for each type of R Numbers.

```
# Define numeric values

num1 <- 7.5

num2 <- 2.3

# Addition

num_add<- num1 + num2

print(num_add)          # Output: 9.8
```

sqrt(x): Computes the square root of x. If x is numeric, it returns the principal square root. If x is complex or negative, it returns a complex number.

Example:

```
# Numeric value
num<- 16
sqrt_num<- sqrt(num)
print(sqrt_num) # Output: 4
```

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write program to demonstrate the use of R-Numbers. (numeric, integer, complex).
2. Write program to convert number from one type to other using functions.
3. Write program to perform following operations.
   i. Addition and Subtraction on numbers.
   ii. Find Square root using of number using built-in function.

(Spaceforanswers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design moresuch questions to ensure the achievement of identified CO.

1. Write R program to print the class name and the type of variable.
2. Write R program to extract real and imaginary parts of a complex number.
3. Write R program to find square root of a Complex Number, Vector of Numeric Values and Negative Number.
4. Write R program to handle floating-point precision errors using the all.equal() function.
5. Write functions to Compute logarithm and exponential of a number.

(Space for answers)

X. References/Suggestions for further Reading

1. https://www.programiz.com/r/numbers
2. https://www.w3schools.com/r/r_numbers.asp
3. https://www.youtube.com/watch?v=yjcSts7CW64
4. https://www.youtube.com/watch?v=2zZxsiOfhxY

XI. Assessment Scheme

	Performance Indicators	Weightage
	Process related (15Marks)	**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
	Product related (10Marks)	**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.3

Write Programs to perform operations on Dataset

I. Practical Significance

Working with datasets in R is fundamental for various practical applications, from data analysis and visualization to machine learning and statistical modeling. This practical will help in transforming and filtering data to make it suitable for analysis and to perform exploratory data analysis to understand data patterns.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Extract useful information for making data-driven decisions.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 3.1 Perform the given operation on Datasets using R programming.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Cultivate a sense of responsibility in performing data operations.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

R programming has some pre-built datasets for its users. These datasets cover a wide range of fields. These datasets can be used to learn R programming. One can perform various operations, practice model building, visualizations and other data analytic operations on these built-in datasets.

Note: One of the commonly used built-in datasets is mtcars, which contains data about various car attributes.

Following functions can be used while working with datasets.

data( )	To check the list of built-in datasets in R.
str(mtcars)	To view the structure of the dataset, including the type of each column and the first few values.
summary(mtcars)	To get a statistical summary of each column.
names(mtcars)	To get the names of the columns.
head(mtcars)	To view the first few rows of the dataset.
sapply(mtcars, class)	To get the type of each column.
sort(mpg_values)	To Sort the values of mpg_values columns in ascending order
sort(mpg_values,decreasing=TRUE)	To Sort the values of mpg_values columns in descending order

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write Program to print any built-in data set of R.
2. Write Program to get information about the data set.
3. Write Program to find the dimensions of the data set and view the names of the variables.
4. Write Program to find the name of each row in the first column.
5. Write Program to print all values that belong to a variable.
6. Write Program to sort the values of variable.
7. Write Program to get the statistical summary of the data.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more. such questions to ensure the achievement of identified CO.

1. Write R program to sort and Get Indices of a column.
2. Write R program to get the data type of each column.
3. Write R program to view the first few rows of the dataset.
4. Write R program to view the last few rows of the dataset.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

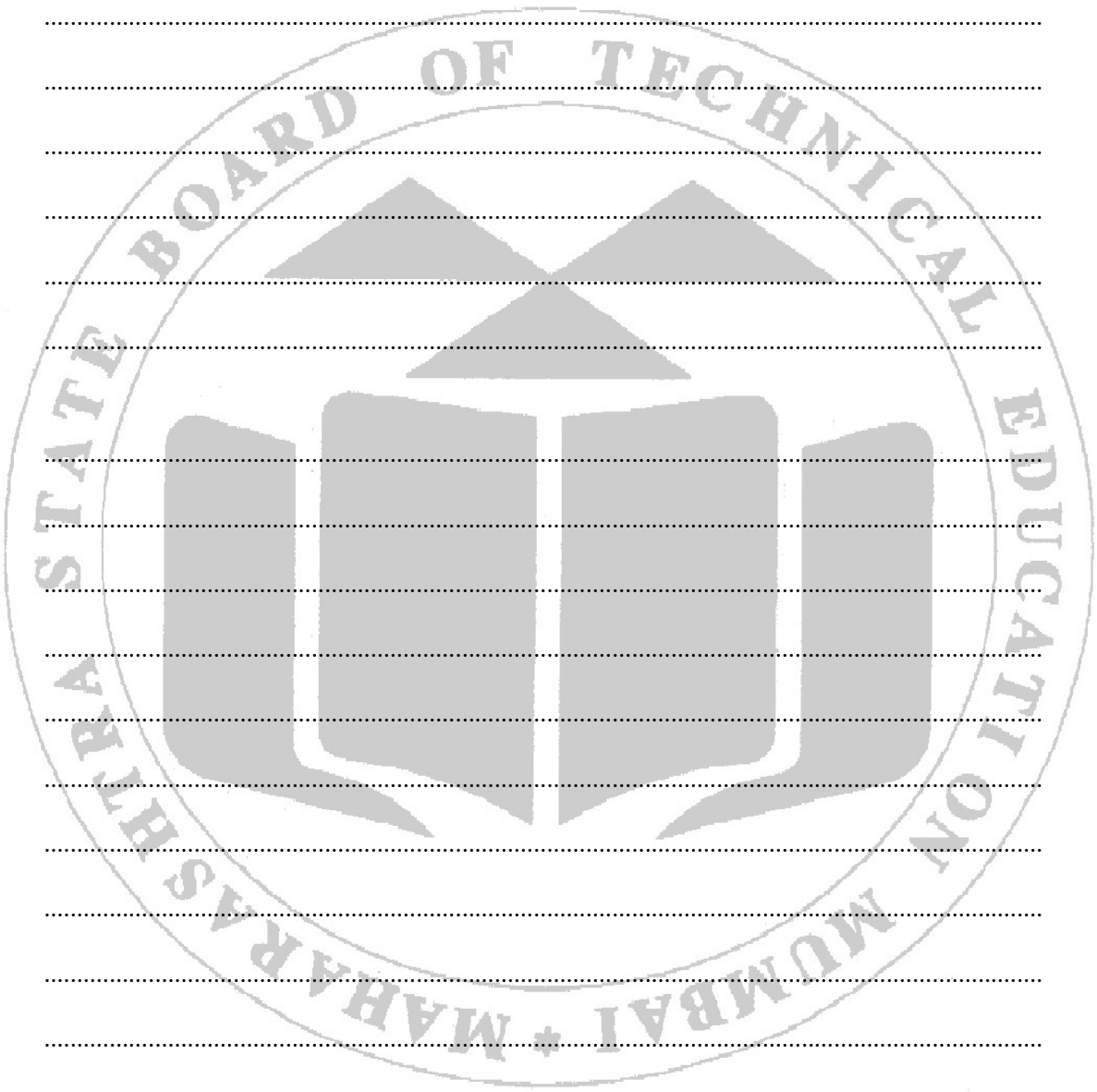
.....

.....

.....

.....

.....



.....

.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/a-complete-guide-to-the-built-in-datasets-in-r/
2. https://www.w3schools.com/r/r_stat_data_set.asp
3. https://www.youtube.com/watch?v=hb9hxCmK8wM
4. https://www.youtube.com/watch?v=LeN9KKvWnDA

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.4

Write a program to-

a. Find the lowest or highest value in a data set.

b. Find the index position of the max and min value in the table.

I. Practical Significance

Finding the lowest or highest value in a data set can have significant practical implications in data analysis, reporting, and decision-making. By understanding the extremes in your data, you can make more informed decisions, ensure data quality, and effectively communicate insights.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Optimize processes and improve operational efficiency.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 4.1 Find lowest and highest values and index position of max and min value from dataset using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Cultivate a sense of responsibility in performing data operations.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

In R, the min( ) and max( ) functions are used to find the smallest and largest values in a dataset, respectively. These functions are fundamental for basic data analysis and are useful in a variety of contexts such as descriptive statistics, data validation and preprocessing.

1. min( ) function:

Returns the smallest value in a numeric vector or dataset.

Syntax: min(x, na.rm = FALSE)

Where,

x: A numeric vector or object for which the minimum value is to be found.

na.rm: A logical value indicating whether to remove NA (missing values) before the computation. Default is FALSE.

Example:

```
data<- c(4, 2, 7, 1, 9)
smallest_value<- min(data)
print(smallest_value)                # Output: 1
```

2. max( ) function:

Returns the largest value in a numeric vector or dataset.

Syntax: max(x, na.rm = FALSE)

Where,

x: A numeric vector or object for which the maximum value is to be found.

na.rm: A logical value indicating whether to remove NA (missing values) before the computation. Default is FALSE.

Example:

```
data<- c(4, 2, 7, 1, 9)
largest_value<- max(data)
print(largest_value)                # Output: 9
```

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a program to find the lowest and highest value in a data set.
2. Write a program to find the index position of the max and min value in the table.
   Hint: use which.max( ) and which.min( ) functions.

(Space for answers)

.....

.....

.....

.....

.....

.....

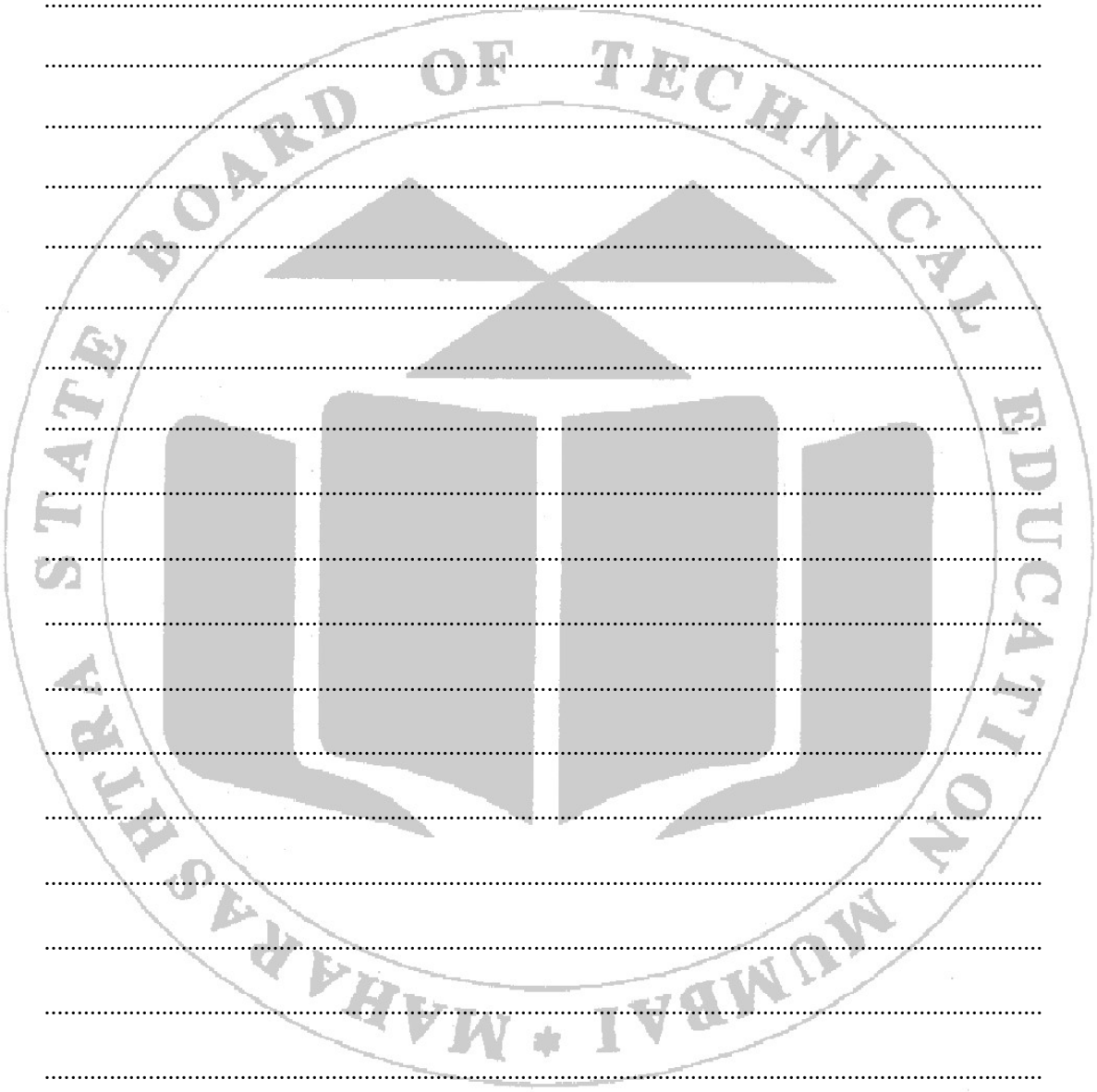
.....

.....

.....

.....

.....



.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more. such questions to ensure the achievement of identified CO.

1. Write R program to find minimum and maximum value from a dataset containing Missing Values.
2. Write R program to find minimum and maximum value from a dataframe.
3. Write R program to find minimum and maximum value from a Vector.
4. Write R program to find minimum and maximum value by combining Multiple Data Sources.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

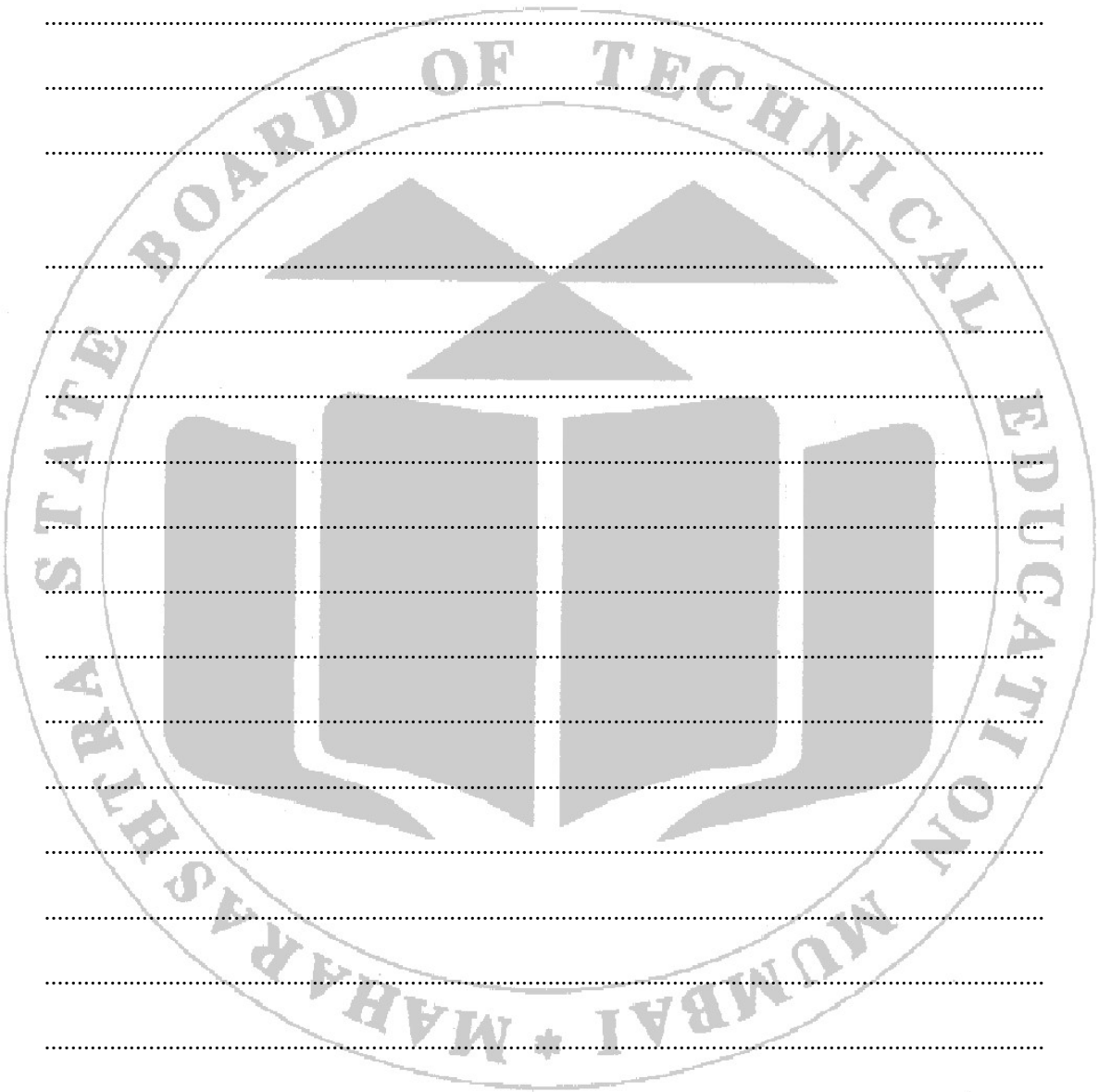
.....

.....

.....

.....

.....



.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/how-to-use-min-and-max-functions-in-r/
2. https://www.digitalocean.com/community/tutorials/min-max-in-r
3. https://www.youtube.com/watch?v=XF9b6XuAHgE
   https://www.youtube.com/watch?v=lFlqZ1shiog

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.5

Write programs to calculate Measures of Central tendency

I. Practical Significance

By understanding measures of central tendency, one can summarize data, make comparisons, support decision-making, and ensure data quality. Proficiency in these calculations is essential for effective data analysis and interpretation across various fields and industries.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Make data-driven decisions by providing clear metrics on central tendencies.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 5.1 Calculate Measures of Central tendency including Quartiles using R Program.

LLO 5.2 Calculate Deciles and Percentiles using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Understand the importance of measures of central tendency in summarizing and interpreting data.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

Central Tendency is one of the features of descriptive statistics. Central tendency tells about how the group of data is clustered around the center value of the distribution. In R, calculating measures of central tendency involves using built-in functions and a few simple commands. Below are R scripts to compute the mean, median, and mode of a data set.

1. Mean:

The mean is the average of a set of numbers. In R, it is calculate using the mean() function.

Syntax:

mean(x, trim = 0, na.rm = FALSE, ...)

Arguments:

- **x:** A numeric vector or data frame column containing the numbers you want to calculate the mean for.
- **trim:** A numeric value between 0 and 0.5 that specifies the fraction of observations to be trimmed from each end of the sorted data before the mean is computed. Default is 0 (no trimming).
- **na.rm:** A logical value indicating whether NA (missing values) should be removed before the computation. Default is FALSE.
- **...:** Additional arguments (not commonly used).

Example:

```
numbers<- c(1, 2, 3, 4, 5)
mean_value<- mean(numbers)
print(mean_value)# Output: 3
```

2. Median:

The median is the middle value when the data is ordered. You can calculate it using the median() function.

Syntax:

median(x, na.rm = FALSE, ...)

Arguments:

- **x:** A numeric vector containing the values for which you want to calculate the median.
- **na.rm:** A logical value indicating whether NA (missing values) should be removed before the computation. Default is FALSE.
- **. . .:** Additional arguments (not commonly used).

Example:

```
numbers<- c(1, 2, 3, 4, 5)
median_value<- median(numbers)
print(median_value)# Output: 3
```

3. Mode:

The mode is the value that appears most frequently. R does not have a built-in function for mode, so you can calculate the mode using custom functions or by leveraging existing R functions and packages.

1. Create a custom function for calculating mode
2. Use DescTools package that have built-in mode functions.

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write program to import data into R.
2. Write program to calculate the Mean (Average value) of a variable from the given data set.
3. Write program to find the Median (Mid-Point value) of the variable from the given data set.
4. Write program to calculate the mode for the variable from the given data set.(by sorting the column of the dataframe and by using the 'modest' package).
5. Write program to calculate the Percentile of the variable from the given data set.

(Space for answers)

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write R program to find mean and median of dataset containing Missing Values.
2. Write R program to find trimmed mean of a dataset.
3. Write R program to find median of a dataset containing Even Number of Values.

(Space for answers)

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/central-tendency-in-r-programming/

2. https://makemeanalyst.com/statistics-with-r/central-tendency-and-spread-in-r-programming/
3. https://www.scaler.com/topics/mean-median-mode-in-r/
4. https://www.youtube.com/watch?v=Hmj5e2dxLAA
5. https://www.youtube.com/watch?v=5Q6Uo73cnT0

XI. Assessment Scheme

	Performance Indicators	Weightage
	Process related (15Marks)	**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
	Product related (10Marks)	**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.6

Write programs to work on Frequency Tables

I. Practical Significance

Frequency tables are essential tools in data analysis and statistical research. They provide a way to organize and summarize data, making it easier to understand and interpret patterns and distributions.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Enhance Data Analysis Capabilities and Improve Decision-Making skills.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 6.1 Create Frequency Distribution table using R Programming.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Recognize the importance of accurate data representation in deriving meaningful conclusions.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

Frequency tables are a fundamental tool in data analysis, used to summarize the distribution of data values. A frequency table is a list of objects with the frequency of each item shown in the table. In R, frequency tables can be created and analyzed using various functions and techniques.

1. Creating Frequency Tables:

Basic Frequency Table:

To create a basic frequency table, you use the table() function. This function computes the frequency of each unique value in a vector or factor.

Syntax:**table(x)**

Example:

```
# Sample data
data <- c("Red", "Blue", "Red", "Green", "Blue", "Blue", "Red")

# Create frequency table
freq_table<- table(data)
print(freq_table)

# Output:
```

```
# data
# Blue Green  Red
#   3     1     4
```

2. Plot the frequency table using ggplot function:

To visualize a frequency table using the ggplot2 package in R, you can transform the frequency table into a data frame suitable for plotting and then use ggplot() to create various types of plots such as bar charts, stacked bar charts, or faceted plots. Below is a step-by-step example demonstrating how to do this.

i) Install and Load Required Packages : install.packages("ggplot2")
library(ggplot2)
ii) Create a Sample Data Frame.
iii) Create a Frequency Table.
iv) Convert Frequency Table to Data Frame.
v) Plot the Frequency Table Using ggplot2.Here is one example of Bar Plot. Similarly students can write code for Faceted Plot.
Example code for Bar Plot of Frequencies:

```
barplot(frequency_table, main="Frequency of Colors", xlab="Colors", ylab="Frequency", col="lightblue")
```

VII. Resources Required

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write program to print Original Data Frame, Modified Frequency Table, Cumulative Frequency Table and Relative Frequency Table.
2. Write program to create the Frequency Table by using multiple arguments.
3. Write program to plot the frequency table using ggplot function.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write R program to create a Pie Chart for frequency table.
2. Write R program to Customize Plot by adding Labels.
3. Write R program to print frequency table with Missing Values.

(Space for answers)

.....

.....

.....

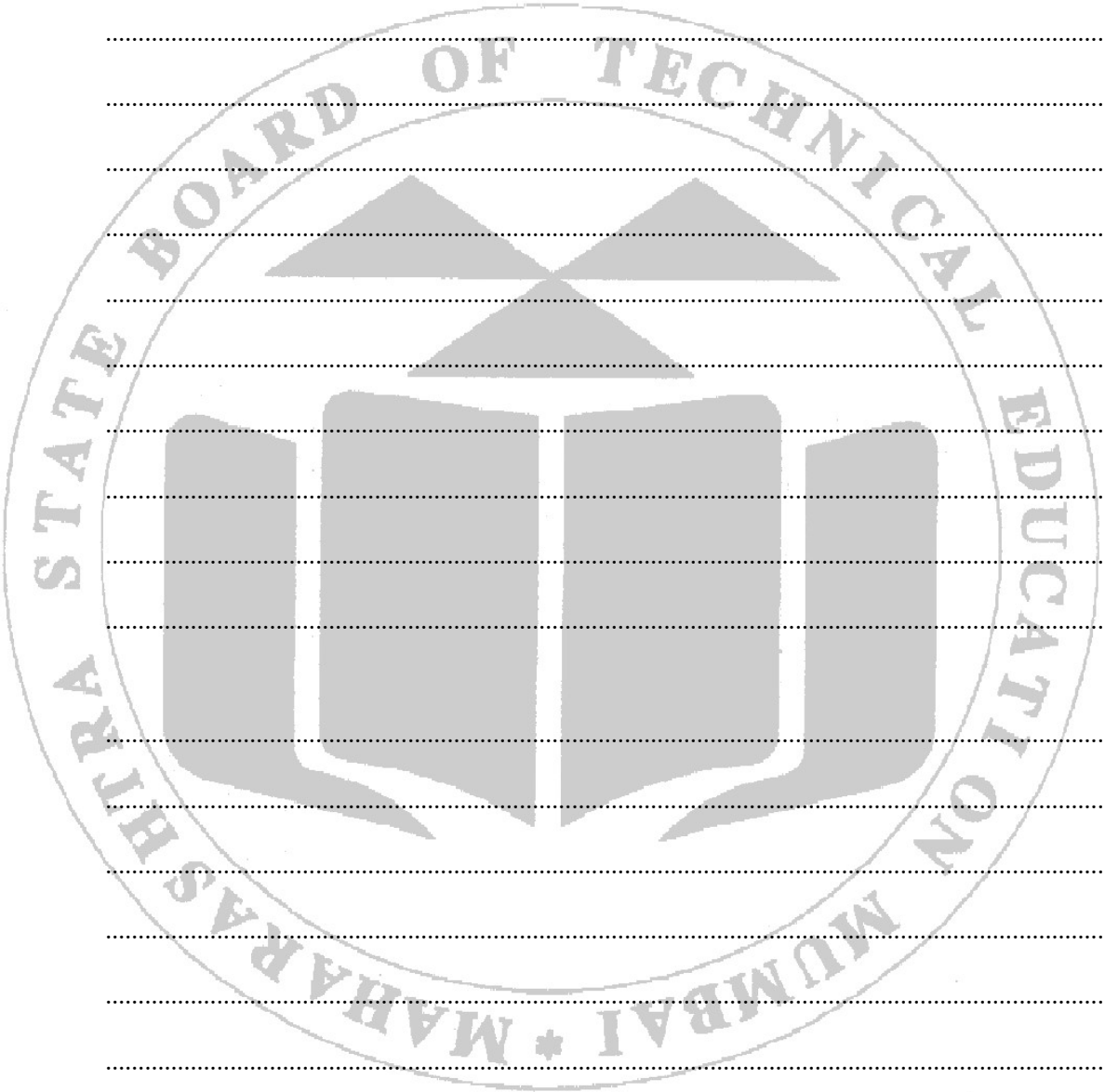
.....

.....

.....

.....

.....



.....

.....

.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/frequency-table-in-r/
2. https://www.tutorialspoint.com/how-to-create-a-frequency-table-in-data-frame-format-in-r
3. https://www.youtube.com/watch?v=-rrPpXz4uFM
4. https://www.youtube.com/watch?v=BX60iXae6k4

XI. Assessment Scheme

	Performance Indicators	Weightage
	Process related (15Marks)	**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
	Product related (10Marks)	**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.7

Write programs to calculate-Variance, Standard Deviation, Range, Mean Deviation for the given data

I. Practical Significance

Calculating measures like variance, standard deviation, range, and mean deviation is essential for understanding the spread and dispersion of the data. It provides insight into data variability and is critical for risk assessment, quality control, and performance evaluation.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Identify variability and risk in processes, assess and identify outliers or extreme values and evaluate consistency.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 7.1 Calculate measures of Dispersion using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Cultivate a positive and proactive attitude towards statistical analysis.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

1. Variance:

Variance measures the average squared deviation of each data point from the mean.

Syntax: **var(x, na.rm = FALSE)**

x: A numeric vector or a data frame column for which you want to calculate the variance.

na.rm: Logical value indicating whether NA (missing) values should be removed before computation. Default is FALSE.

2. Standard Deviation:

Standard deviation is the square root of the variance and measures the average distance of each data point from the mean.

Syntax: **sd(x, na.rm = FALSE)**

x: A numeric vector for which you want to calculate the standard deviation.

na.rm: Logical value indicating whether NA (missing) values should be removed before computation. Default is FALSE.

3. Range:

Range is the difference between the maximum and minimum values in the dataset.

Syntax: **range(x)**

x:A numeric vector for which you want to calculate the range.

4. Mean Deviation (Absolute Deviation):

Mean deviation measures the average absolute distance of each data point from the mean. Mean deviation is not a built-in function in R, so you need to compute it manually using basic functions.

Example:

```
# Sample data
data<- c(5, 7, 8, 9, 10)

# Calculate mean
mean_value<- mean(data)

# Calculate mean deviation
mean_deviation<- mean(abs(data - mean_value))

# Print the result
print(paste("Mean Deviation:", mean_deviation))
```

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

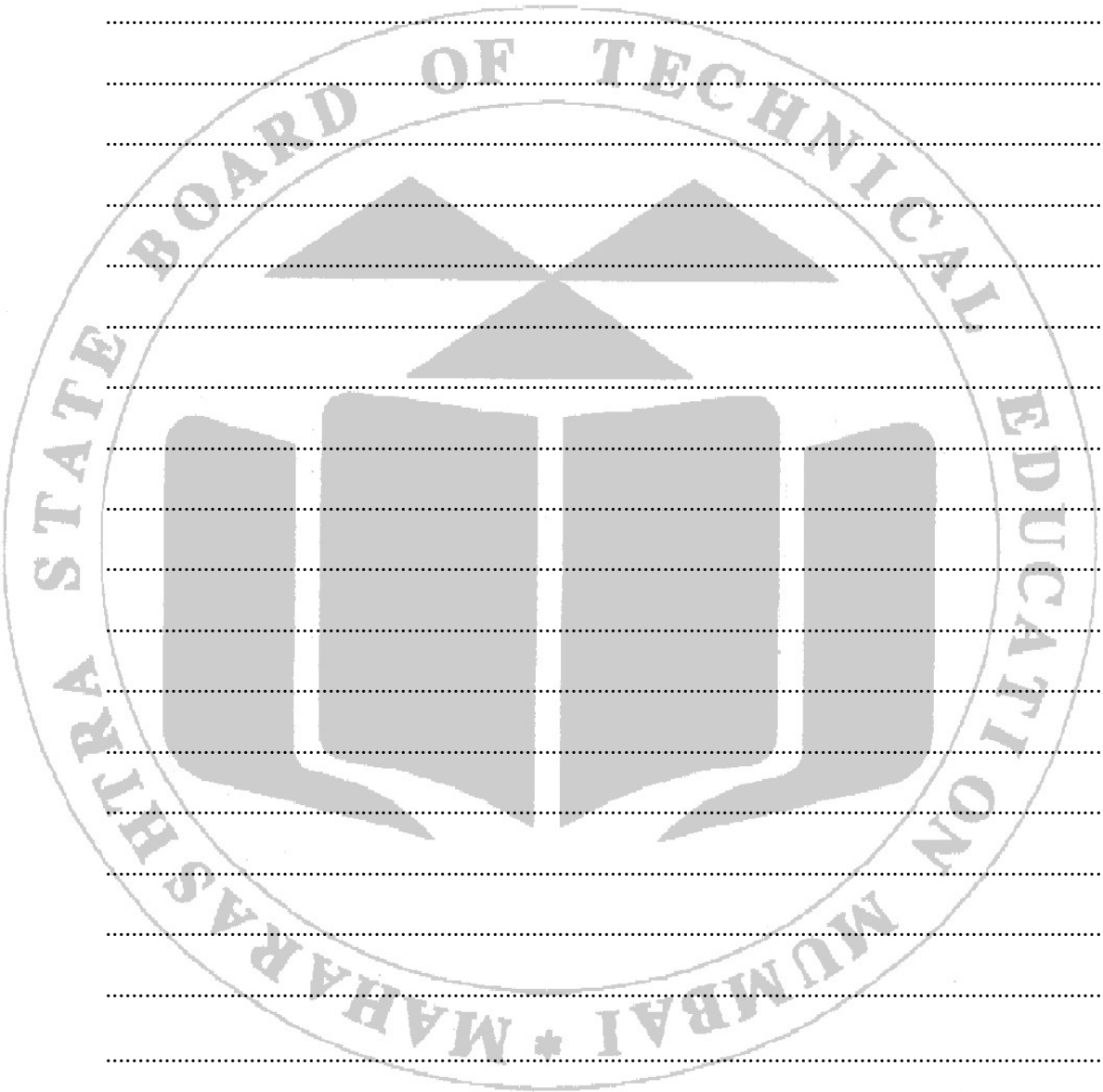
(Use blank space for answers or attach more pages if needed)

1. Write program to calculate Variance for the dataset and print the result.
2. Write program to calculate Standard Deviation for the dataset and print the result.
3. Write program to calculate Range for the dataset and print the result.
4. Write program to calculate Mean Deviation for the dataset and print the result.

(Space for answers)

.....

.....



.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write R program to calculate the Interquartile Range (IQR) for the given dataset.
2. Write R program to find the Standard deviation of the values stored in a CSV file.
3. Write R program to find Standard Deviation for a List of Even Numbers.

(Space for answers)

.....

.....

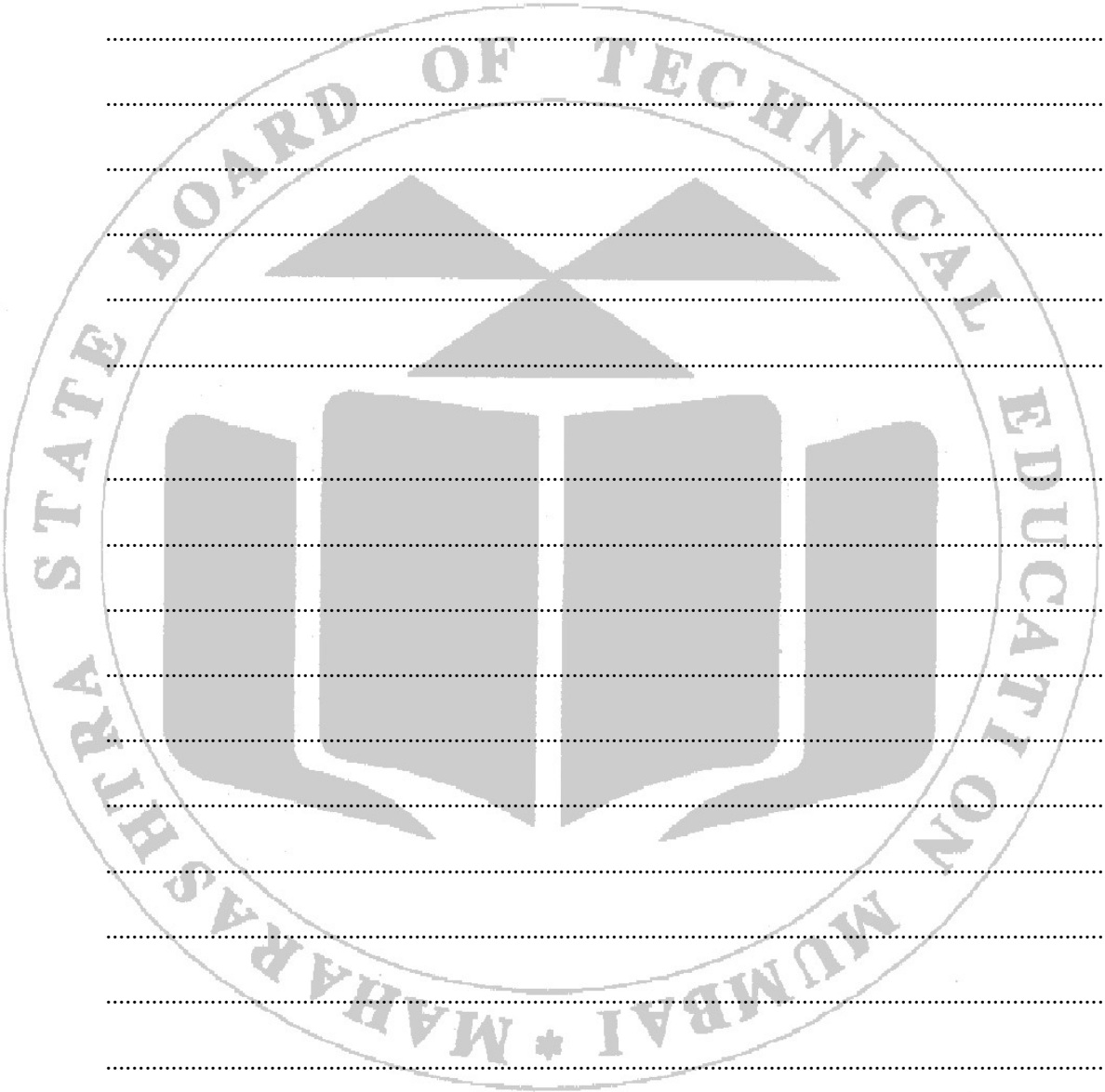
.....

.....

.....

.....

.....



.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/calculate-the-average-variance-and-standard-deviation-in-r-programming/
2. https://www.digitalocean.com/community/tutorials/find-standard-deviation-in-r
3. https://www.youtube.com/watch?v=HeFwGmG40lw
4. https://www.youtube.com/watch?v=6SqE6BNPLhE

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.8

Write Programs to graphically represent mode and median of the given data

I. Practical Significance

Graphically representing the mode and median of a dataset in R is useful for visualizing the central tendencies and distribution of the data, providing insights that are not always apparent from numerical summaries alone. This helps to understand how data points are distributed around these central values.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Create visually appealing and informative plots which can be useful in the early stages of data analysis.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 8.1 Draw Histogram and Ogive Curve to graphically represent Mode and Median for the given data.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Understand the value of visualizing statistical measures in enhancing data comprehension.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

To graphically represent the mode and median of a given dataset in R, one can create a histogram and an ogive (cumulative frequency curve).

1. Draw a Histogram

A histogram displays the distribution of data and is useful for visualizing the frequency of data points within certain ranges.`ggplot2` is a powerful package for creating more customizable and aesthetically pleasing plots. The `geom_histogram()` function is used to create histograms.

geom_histogram() (ggplot2):

Arguments

- bins: Number of bins for the histogram.
- fill: Fill color of the bars.
- color: Border color of the bars.
- labs(): Function to set titles and axis labels.
- theme_minimal(): Applies a minimal theme to the plot for a cleaner look.

2. Draw an Ogive Curve

An ogive is a cumulative frequency curve that shows the cumulative frequency of data points up to a certain value. It's useful for understanding the distribution of data across

different ranges.ggplot2 provides a more flexible and visually appealing way to create an ogive curve. You can use geom_step() to create a step plot representing the cumulative frequency.

geom_step(): Creates a step plot for cumulative frequencies.
labs(): Sets titles and axis labels.
theme_minimal(): Applies a minimal theme for a cleaner look.

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write program to draw Histogram to graphically represent mode and median of the given data.
2. Write program to draw Ogive Curve to graphically represent mode and median of the given data.

(Space for answers)

.....

.....

.....

.....

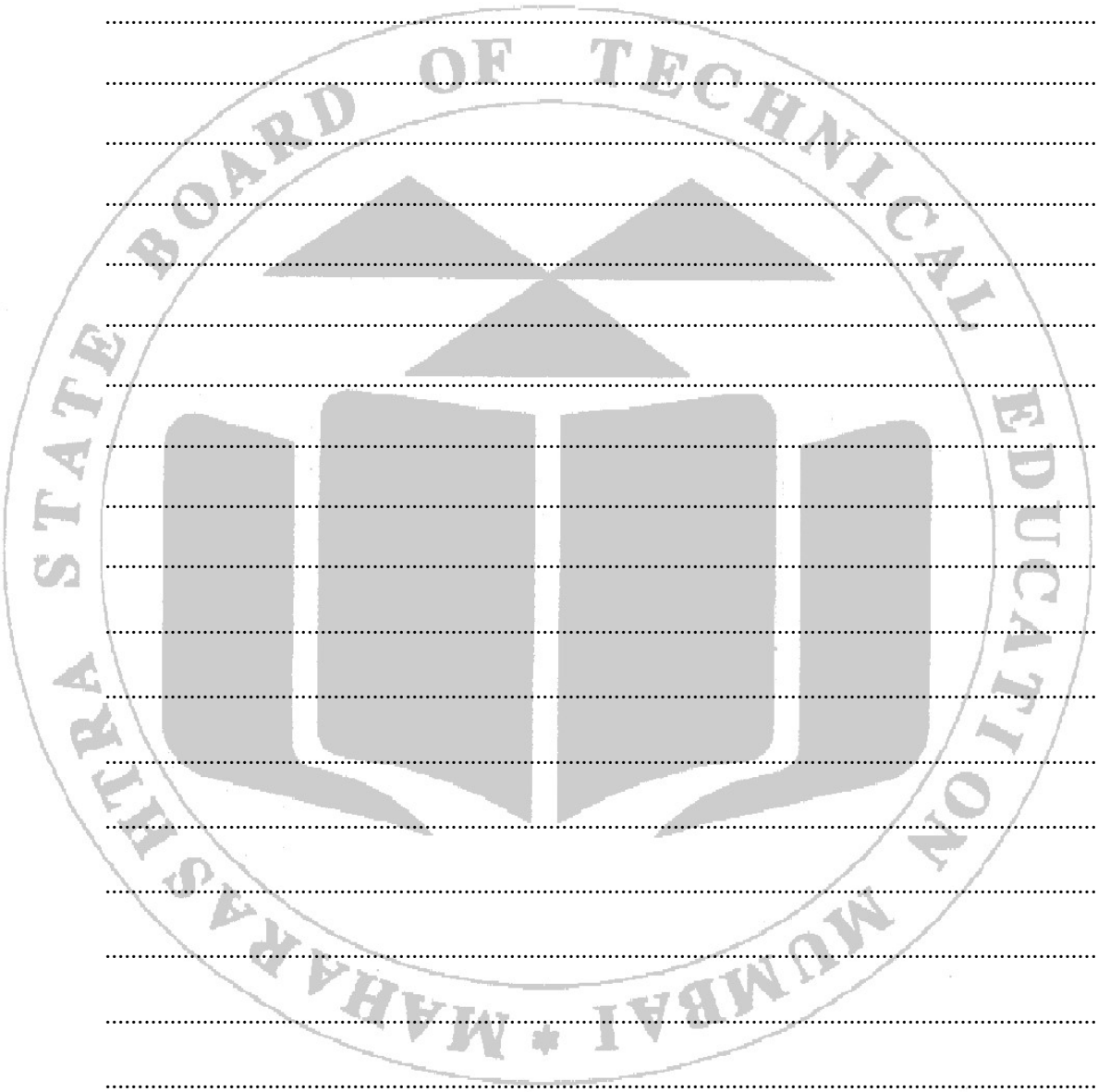
.....

.....

.....

.....

.....



.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write R program to create a Histogram return value chart.
2. Write R program to create Histogram using non-uniform width.
3. Write R program to Extract Frequency Counts from Histogram.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/histograms-in-r-language/
2. https://www.statology.org/ogive-graph-in-r/
3. https://www.youtube.com/watch?v=_GN81FhJ7qM
4. https://www.youtube.com/watch?v=lXvka6db1tw

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.9

Write a Program to calculate Skewness for the given data

I. Practical Significance

Calculating skewness for a given dataset is important because it provides insight into the asymmetry of the data distribution. It enhances the understanding of the distribution characteristics of the data and guides to make better decision and more accurate modeling.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Calculate skewness and integrate these insights into their data analysis processes to achieve the aforementioned outcomes.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 9.1 Calculate Skewness for the given data using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Troubleshoot issues related with distribution of data.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

Skewness is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side is not mirror images. To calculate the skewness of a dataset using R programming, follow these steps:

1. Install and Load the Required Package:

First, ensure you have the e1071 package installed, as it provides the function for calculating skewness.

```
install.packages("e1071")  # Run this line if you haven't installed the package
library(e1071)             # Load the package
```

2. Create or Load Your Dataset:

Define your dataset. You can either create a dataset manually or load it from an external source.

3. Calculate Skewness:

Use the skewness() function from the e1071 package to compute the skewness.

```
# Calculateskewness
skewness_value<- skewness(data)
print(skewness_value)
```

VII. Resources Required

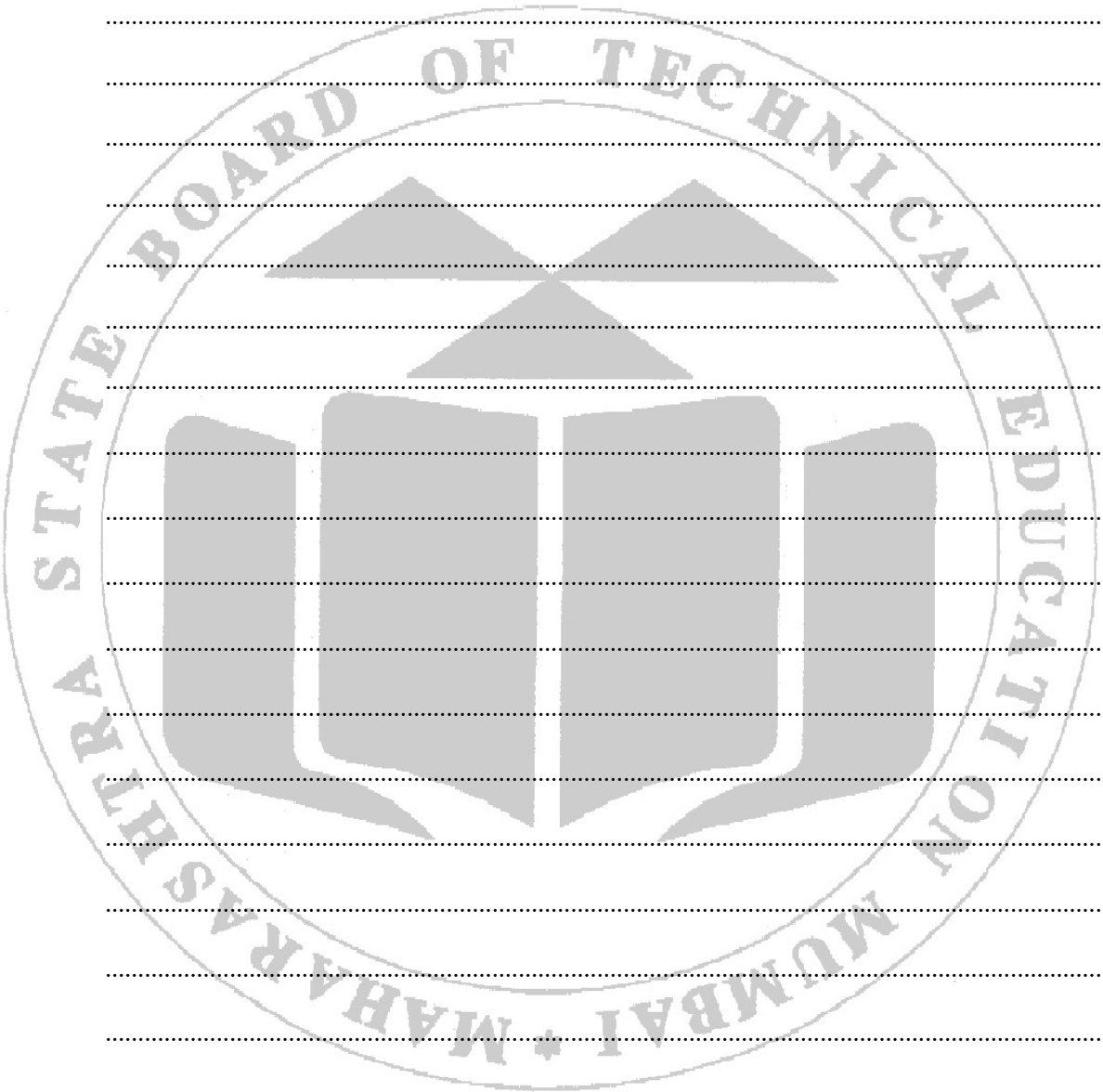
Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a Program to calculate Skewness for the given data by following below steps.
   a) Define normally distributed data vector in R.
   b) Present output as PNG file.
   c) Print skewness of distribution.
   d) Draw Histogram of distribution.
   e) Save the file.

(Space for answers)



.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write R program to calculate Positive Skewness for the given dataset and draw Histogram.
2. Write R program to calculate Negative Skewness for the given dataset and draw Histogram.
3. Write R program to calculate Zero Skewness for the given dataset and draw Histogram.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

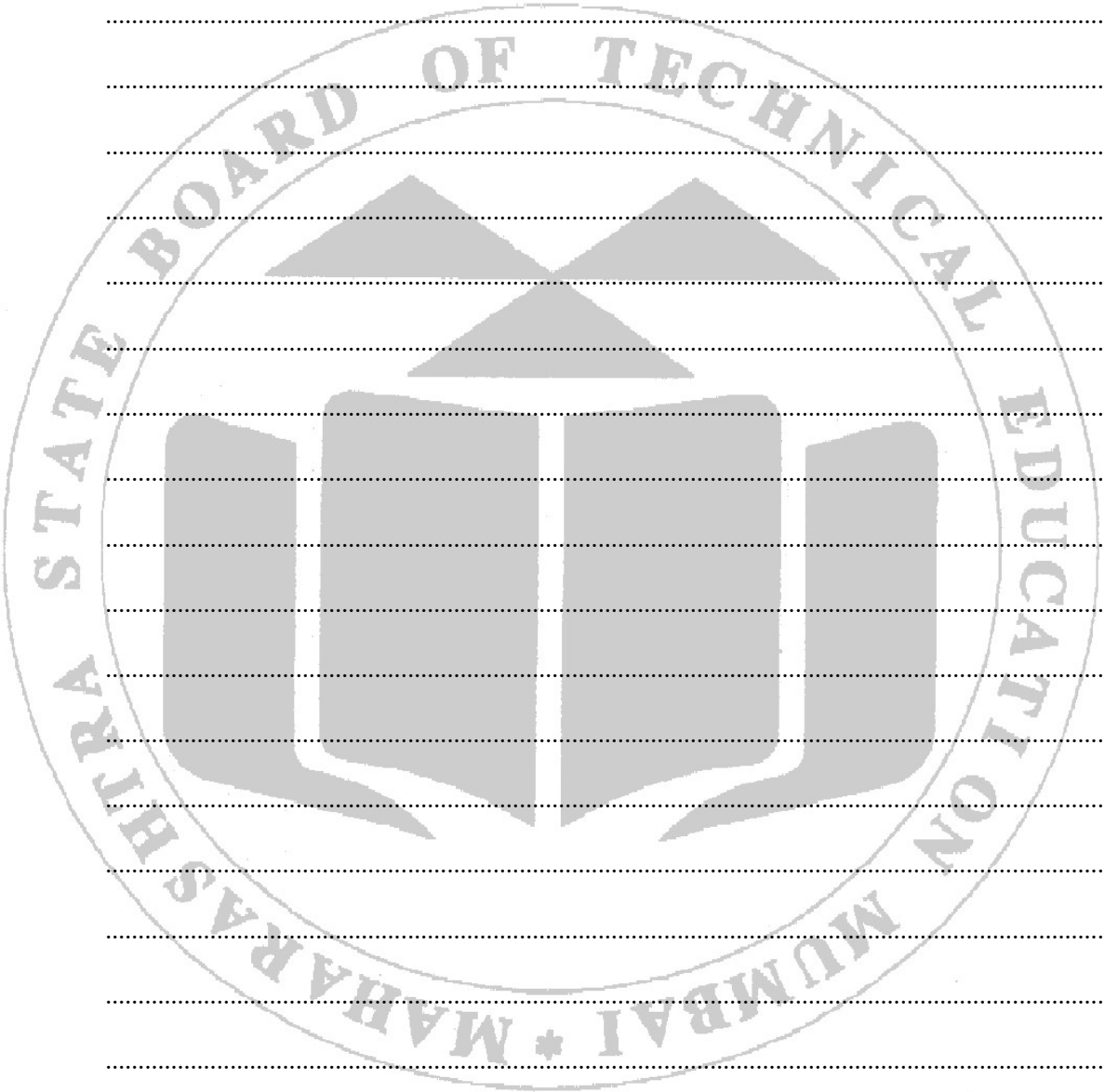
.....

.....

.....

.....

.....



X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/skewness-and-kurtosis-in-r-programming/
2. https://methodenlehre.github.io/SGSCLM-R-course/functions.html
3. https://www.programmingr.com/statistics/skewness/
4. https://www.youtube.com/watch?v=j4aC22l-D5k
5. https://www.youtube.com/watch?v=nIVsZ0l-DBc

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.10

Write a Program to draw a scatter plot for two variables for the given dataset

I. Practical Significance

Drawing a scatter plot for two variables is a powerful way to visually analyze the relationship between them. Using R to create scatter plots helps in making data-driven decisions and facilitates a deeper understanding of the dataset.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Visualize data, identify correlations and outliers, enhance decision-making and reporting.'

III. Course Level Learning Outcome (CO)

CO1: Solve the given problem based on Statistic Techniques using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 10.1 Draw a scatter plot for the given data using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Develop a positive, engaged, and collaborative approach to data visualization.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

A scatter plot is a set of dotted points representing individual data pieces on the horizontal and vertical axis. In a graph in which the values of two variables are plotted along the X-axis and Y-axis, the pattern of the resulting points reveals a correlation between them. Students can use the base R plotting system or various packages like ggplot2 for more advanced and customizable plots.

Using Base R:

To create a simple scatter plot using base R, you can use the plot() function.

Example:

```
# Sample data
x <- c(1, 2, 3, 4, 5)
y <- c(2, 4, 6, 8, 10)

# Create scatterplot
plot(x, y,
main = "Scatterplot of x vs. y",
xlab = "X-axis label",
ylab = "Y-axis label",
pch = 19,          # Type of point (19 is solid circle)
col = "blue")      # Color of points
```

Using ggplot2:

ggplot2 is a powerful and flexible package for creating plots in R. Here's how to create a scatterplot with ggplot2.

1. Install ggplot2 if you haven't already.
2. Use ggplot2 to create the scatterplot.

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a Program to draw a scatterplot for two variables for the given dataset using Base R.
2. Write a Program to draw a scatterplot for two variables for the given dataset using ggplot2 package.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write a Program to draw a scatterplot for two variables for the given dataset using Base R. Add more parameters to the plot() function like pch, cex, col etc for customization.
2. Write a Program to draw a scatterplot for two variables for the given dataset using ggplot2 package. Add layers and themes to customize the plot.
3. Write a Program to draw a scatterplot with fitted values.
4. Write a Program to draw 3D Scatterplots.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

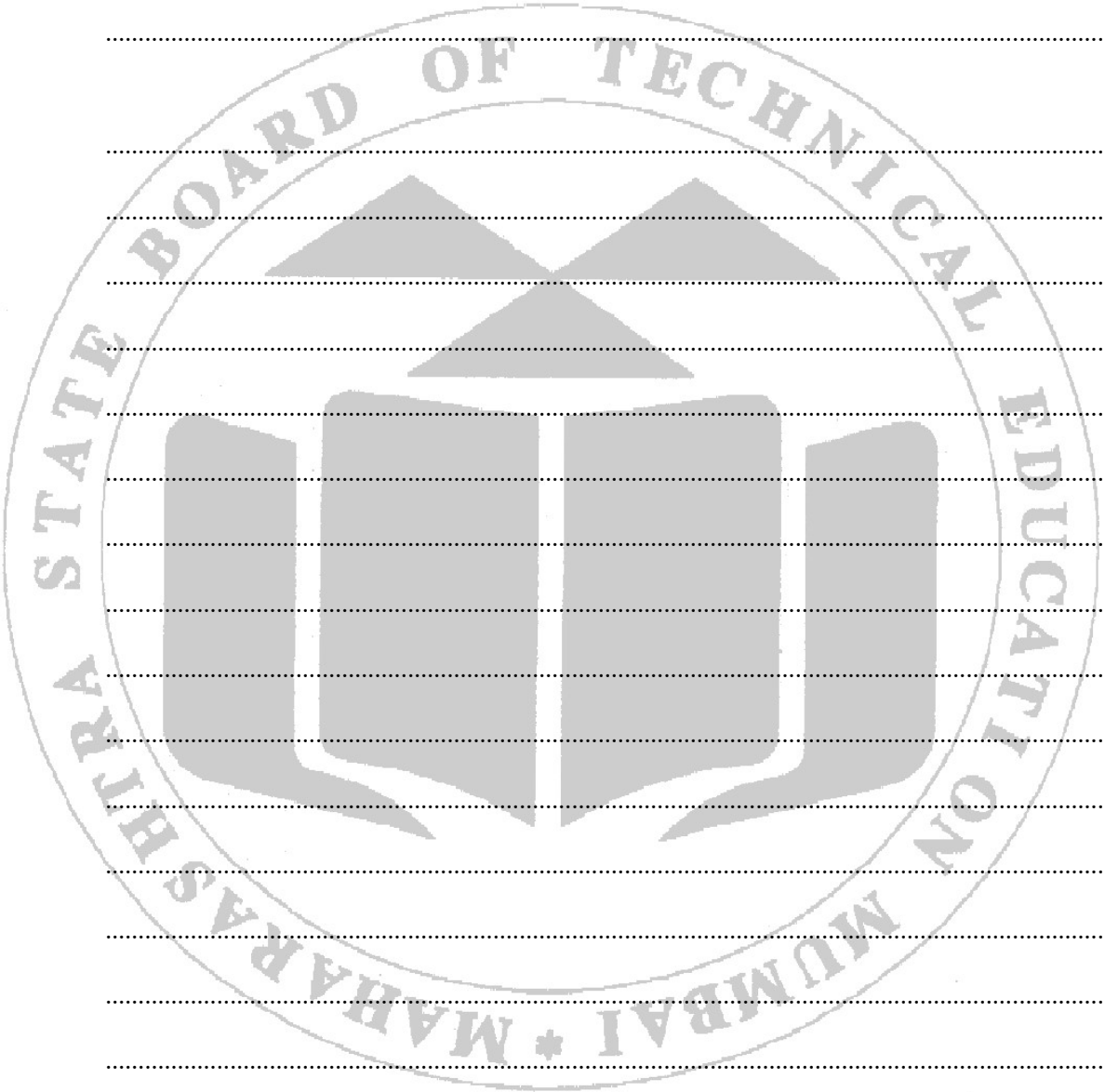
.....

.....

.....

.....

.....



.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/scatter-plots-in-r-language/
2. https://www.w3schools.com/r/r_graph_scatterplot.asp
3. https://www.youtube.com/watch?v=FEAS3akVxD8
4. https://www.youtube.com/watch?v=lPOSwfxMd3c

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.11

Write Program to perform the correlation test to evaluate the association between two or more variables

I. Practical Significance

Applying correlation analysis effectively allows one to make informed decisions based on the relationships between variables, leading to more accurate insights and better outcomes in various fields. Whether it is in business, research, or everyday life, understanding the correlations between different factors can provide valuable insights and help in making informed decisions.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Enhance decision-making; improve predictive modeling, operational efficiency, data-driven insights, quality control, and effective cost-benefit analysis.'

III. Course Level Learning Outcome (CO)

CO2 - Implement Statistic methods using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 11.1 Perform the correlation test to evaluate the association between two or more variables using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Build learners' confidence in their ability to perform correlation tests accurately and interpret the results meaningfully.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

Correlations between variables play an important role in a descriptive analysis. A correlation measures the relationship between two variables, that is, how they are linked to each other. In this sense, a correlation allows one to know which variables evolve in the same direction, which ones evolve in the opposite direction, and which ones are independent.

Following are the steps to perform the correlation test to evaluate the association between two or more variables.

(Keep a dataset in CSV format and you want to evaluate the correlations between two or more variables.)

1. Install and Load Required R Packages(if needed)

```
install.packages("tidyverse")

library(tidyverse)
```

2. Load Your Dataset.

Load your dataset into R. Replace 'path/to/your/dataset.csv' with the path to your CSV file.

3. Compute Correlations.

Compute the correlation coefficients between the variables by using a correlation matrix.

4. Perform Correlation Tests.

Use the cor.test() function to test the significance of the correlation.

This test will provide you with the correlation coefficient, confidence intervals and p-values which indicate the significance of the correlation.

5. Visualize the Data.

Create scatter plots to visually assess the relationship between variables.

For correlation matrix visualization:

```
# Visualize the correlation matrix
corrplot(cor_matrix, method = "circle")
```

6. Perform Preliminary Tests.

Check for assumptions of the correlation test:

- Linearity: Ensure the relationship between variables is linear (visualized by scatter plots).
- Normality of Residuals: Check if residuals of a linear model are normally distributed.
- Homogeneity of Variances: Check for constant variance of residuals.

VII. Resources Required

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write Program to perform the correlation test to evaluate the association between two or more variables.
   a. Install and load required R packages.
   b. Compute correlation in R.
   c. Visualize your data using scatter plots.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write a program to Perform Preliminary test to check the test assumptions of the correlation test for the following:
    a) Linearity
    b) Normality of Residuals
    c) Homogeneity of Variances

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/pearson-correlation-testing-in-r-programming/
2. https://statsandr.com/blog/correlation-coefficient-and-correlation-test-in-r/
3. https://www.scribbr.com/statistics/correlation-coefficient/
4. https://www.youtube.com/watch?v=qo1FVrlvW1Y
5. https://www.youtube.com/watch?v=oqD_wr9_eDg

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.12

Write Program to perform the correlation test to evaluate the association between two or more variables

I. Practical Significance

Applying correlation analysis effectively allows one to make informed decisions based on the relationships between variables, leading to more accurate insights and better outcomes in various fields. Pearson correlation is useful when one expects a linear relationship and the data meets the assumptions of normality and homoscedasticity. Spearman correlation is useful when the relationship is expected to be monotonic but not necessarily linear, or when your data is ordinal or does not meet the assumptions required for Pearson's correlation.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Compute relationships between variables which can help to make strategies, improve processes and guide decision-making.'

III. Course Level Learning Outcome (CO)

CO2 - Implement Statistic methods using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 12.1 Perform the correlation test to evaluate the association between two or more variables using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Build learners' confidence in their ability to perform correlation tests accurately and interpret the results meaningfully.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram

Pearson Correlation Test:

Data Preparation: data contains two continuous variables.

Correlation Calculation:cor() computes the Pearson correlation coefficient.

Interpretation: The script interprets the sign of the correlation coefficient to determine the type of relationship.

Significance Test:cor.test() provides additional statistics, including p-value, which helps determine if the correlation is statistically significant.

Spearman Rank Correlation Coefficient:

Data Preparation:ata_ranks contains ordinal or ranked data.

Correlation Calculation:cor() computes the Spearman rank correlation coefficient.

Interpretation: The script interprets the sign of the coefficient to determine the type of monotonic relationship.

Significance Test:cor.test() again provides statistics and a p-value for significance testing.

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write Program to perform the correlation test to evaluate the association between two or more variables.
   a. Pearson correlation test.
   b. Interpretation of the result.
   c. Use Spearman rank correlation coefficient to estimate a rank-based measure of association.

(Space for answers)

.....

.....

.....

.....

.....

.....

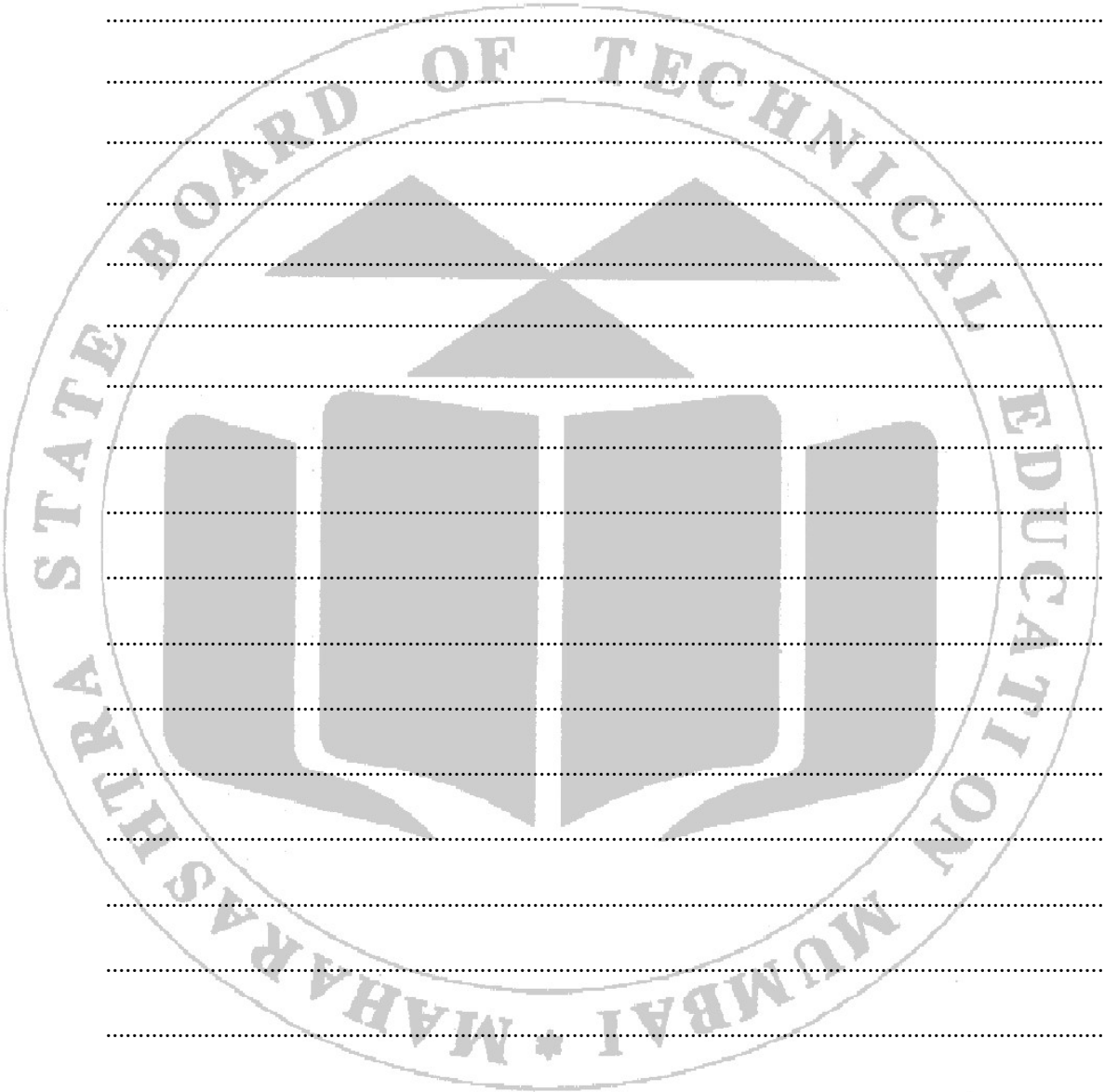
.....

.....

.....

.....

.....



.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write an R program to compute the Pearson correlation coefficient between height and weight. Interpret whether taller individuals tend to weigh more. (**Data**: Collect data on height (in cm) and weight (in kg) from a group of individuals.)
2. Write an R program to compute the Pearson correlation coefficient. Interpret the relationship between temperature and ice cream sales. (**Data**: Collect data on daily temperature and ice cream sales for a month.)
3. Write an R program to compute the Spearman rank correlation coefficient and interpret the relationship between job satisfaction and performance. (**Data**: Gather ordinal rankings for job satisfaction and performance for a set of employees.)
4. Write an R program to calculate the Spearman rank correlation coefficient. Interpret whether higher customer ratings are associated with better perceived quality. (Data: Collect ordinal ratings for customer satisfaction and perceived product quality.)

(Space for answers)

.....

.....

.....

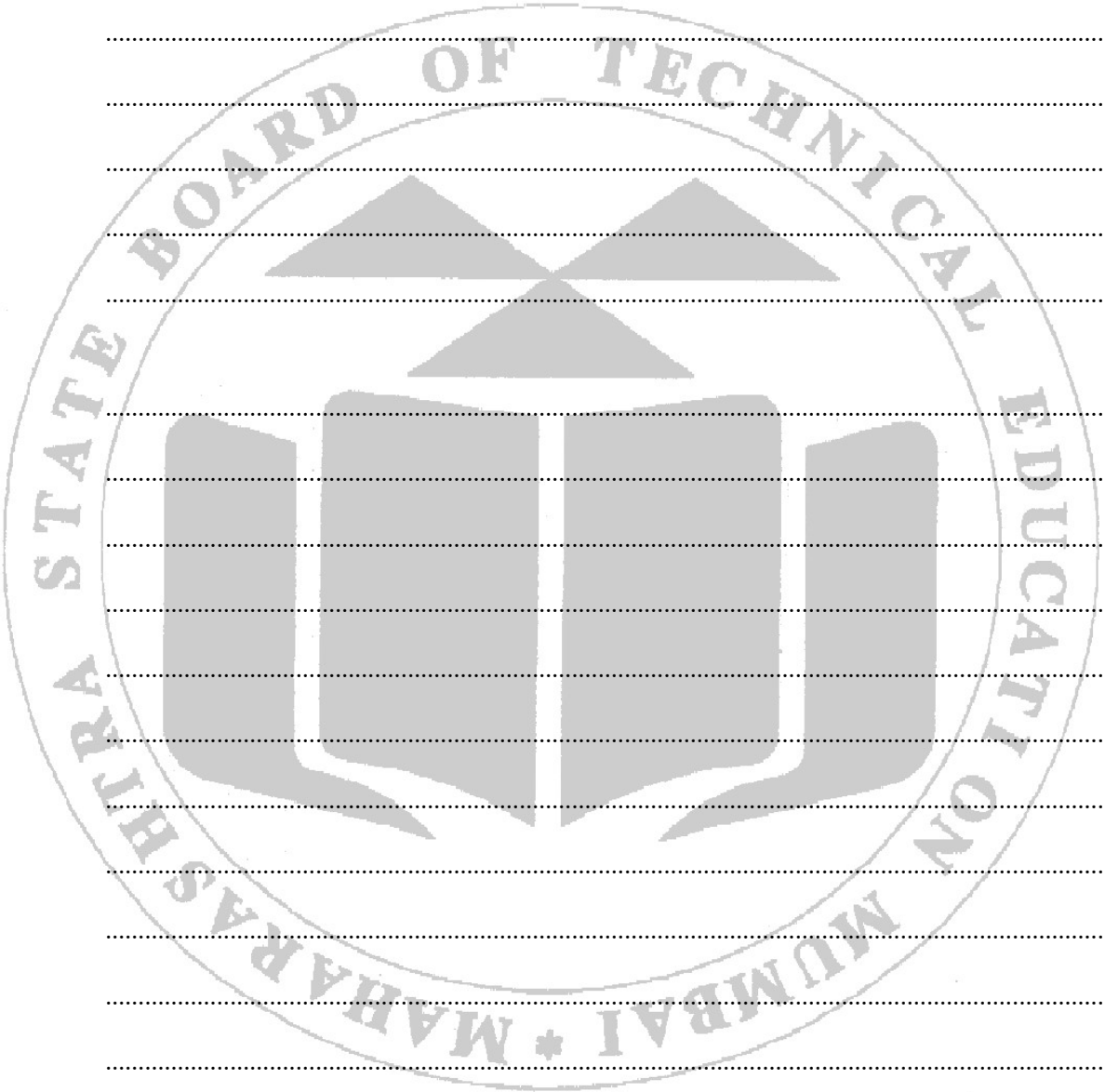
.....

.....

.....

.....

.....



.....

.....

X. References/Suggestions for further Reading

1. https://www.geeksforgeeks.org/pearson-correlation-testing-in-r-programming/
2. https://www.onlinespss.com/pearson-correlation-in-r/
3. https://www.geeksforgeeks.org/spearman-correlation-testing-in-r-programming/
4. https://www.onlinespss.com/spearmen-correlation-in-r/
5. https://www.youtube.com/watch?v=2J_ZlxLeuQU
6. https://www.youtube.com/watch?v=81GifK9vKxc

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No.13

Write a Program based on Line of Regression using two variables

I. Practical Significance

The line of regression provides valuable insights into the relationship between two variables. Applying linear regression helps in making predictions, identifying relationships and guiding decision-making processes. The R program can demonstrate how to fit a linear regression model, interpret the results and visualize the relationship which can be applied to various real-world scenarios.

II. Industry/Employer Expected Outcome (s)

This practical is expected to develop the following skills for the industry/Employer 'Generate predictions based on the linear relationship between two variables and use regression results to optimize performance.'

III. Course Level Learning Outcome (CO)

CO2 - Implement Statistic methods using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO 13.1 Find the Line of Regression for two variables using R Program.

V. Relative Affective Domain related Outcome(s)-

- Follow safety practices.
- Develop a positive, proactive and collaborative approach to regression analysis.
- Follow ethical Practices.

VI. Minimum Theoretical Background with diagram (if required)

Linear Regression is a statistical approach for modelling the relationship between a dependent variable and a given set of independent variables. It is predicted that a straight line can be used to approximate the relationship. The goal of linear regression is to identify the line that minimizes the discrepancies between the observed data points and the line's anticipated values.

Steps to Perform Linear Regression in RStudio:

1. Set Up Your Environment.

- **Open RStudio**: Launch RStudio on your computer.
- **Create a New R Script**: Go to File > New File > R Script to open a new script editor where you can write and save your R code.

2. Load Your Data.

You can either input data manually or import it from a file (e.g., CSV).

3. Fit the Linear Regression Model.

Use the lm() function to fit a linear regression model.

4. Summarize the Model.

Use the summary() function to get detailed information about the model.

5. Make Predictions.
Use the predict() function to make predictions based on the model.

6. Visualize the Data and Regression Line.
Visualize the data points and the fitted regression line.

7. Save and Export Your Results.
- **Save Your Script**: Click File > Save to save your R script.
- **Export Plots**: Click on the "Plots" pane, then use the export options to save your plot as an image or PDF.

VII. Resources Required

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a Program based on Line of Regression using two variables.
   a) Load the Data.
   b) Fit the Linear Regression Model.
   c) Summarize the Model.
   d) Make Predictions.
   e) Visualize the Data and Regression Line.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related Questions

Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO.

1. Write an R program to investigate how study hours influence exam scores.
2. Write an R program to analyze the Relationship between Advertising Spend and Sales.
3. Write an R program to analyze the Effect of Experience on Salary.

(Space for answers)

.....

.....

.....

.....

.....

.....

.....

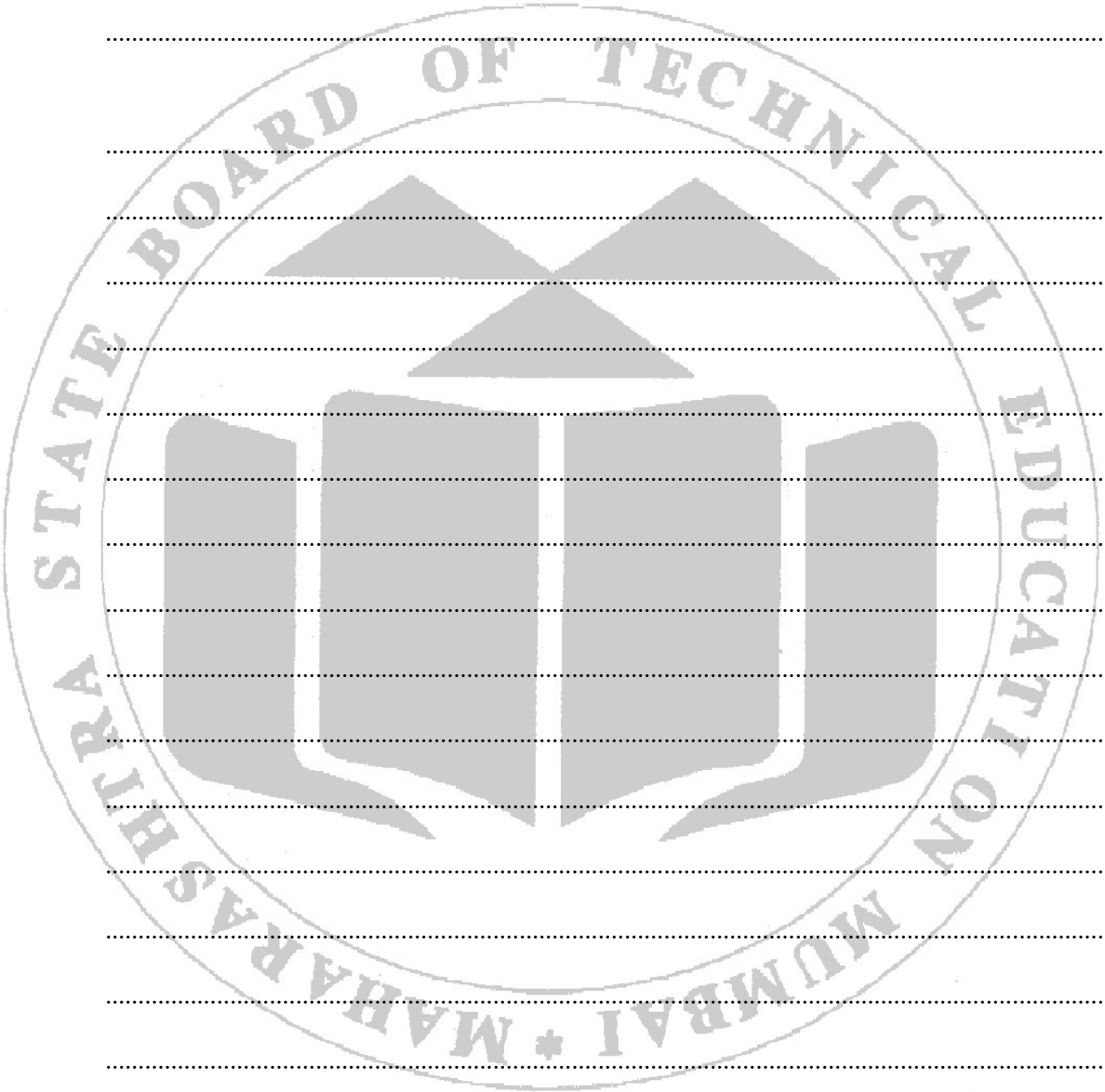
.....

.....

.....

.....

.....



.....

.....

.....

.....

.....

X. References/Suggestions for further Reading

1. https://www.datacamp.com/tutorial/multiple-linear-regression-r-tutorial
2. https://www.geeksforgeeks.org/simple-linear-regression-using-r/
3. https://www.scribbr.com/statistics/linear-regression-in-r/
4. https://www.youtube.com/watch?v=xlJyKDzru1o
5. https://www.youtube.com/watch?v=-mGXnm0fHtI

XI. Assessment Scheme

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 14

Write Programs to calculate Probability.

I. Practical Significance

Classical probability theory provides a solid foundation for understanding random events and their likelihood in various scenarios. Students can explore mathematical foundations, properties, and practical codes of classical probability, offering insights suitable for both newcomers and seasoned data analysts seeking to grasp the core principles of probability in the R environment. This practical will make learner to use fundamental concepts of classical probability within the context of the R programming language.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO3 - Use Principles of Probability to solve the given Problem.

IV. Laboratory Learning Outcome(s)

LLO.14.1 Implement Classical Probability using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

Classical Probability:

Classical probability, often referred to as "a priori" probability, is a branch of probability theory that deals with situations where all possible outcomes are equally likely. It provides a foundational understanding of how probability works and forms the basis for more advanced probability concepts.

Mathematical Foundations:

Sample Space: The sample space represents the set of all possible outcomes in a given experiment. It serves as the foundation for calculating probabilities. For instance, when rolling a fair six-sided die, the sample space is {1, 2, 3, 4, 5, 6}.

Events: An event is a subset of the sample space, representing a specific outcome or set of outcomes. Events can range from simple, such as rolling an even number, to complex, like drawing a red card from a deck.

Probability Distribution: A probability distribution assigns probabilities to each event in the sample space. For classical probability, all outcomes are equally likely, so each event has the same probability.

Calculating Classical Probability:

Classical probability is based on the principle of equally likely outcomes. Consider an experiment with a finite sample space S, consisting of n equally likely outcomes. Let A be an event of interest within S.

The classical probability of event A, denoted as P(A), is calculated as:

P(A) = Number of favorable outcomes for event A/Total number of equally likely outcomes in S

Mathematically, this can be expressed as:

P(A) = {n(A)}/{n(S)}

Where:

P(A) is the probability of event A.

n(A) is the number of favorable outcomes for event A.

n(S) is the total number of equally likely outcomes in the sample space S.

This formula allows us to calculate the probability of an event by counting the favourable outcomes and dividing by the total number of equally likely outcomes.

In R, you can use this formula to calculate classical probabilities for various events, making it a fundamental concept in probability theory for data analysis and statistics.

Implementing Classical Probability in R:

1. The probability of getting heads when flipping a fair coin can be calculated as:

```
outcomes<- c("Heads", "Tails")

total_outcomes<- length(outcomes)

favorable_outcomes<- length(outcomes[outcomes == "Heads"])

classical_prob<- favorable_outcomes / total_outcomes

classical_prob
```

Output:

[1] 0.5

2. Calculating the probability of drawing a spade from a standard deck of 52 cards.

```
deck<- rep(c("Spades", "Hearts", "Diamonds", "Clubs"), each = 13)

total_cards<- length(deck)

spades<- length(deck[deck == "Spades"])

classical_prob_spade<- spades / total_cards

classical_prob_spade
```

Output:

[1] 0.25

VII. Required Resources

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

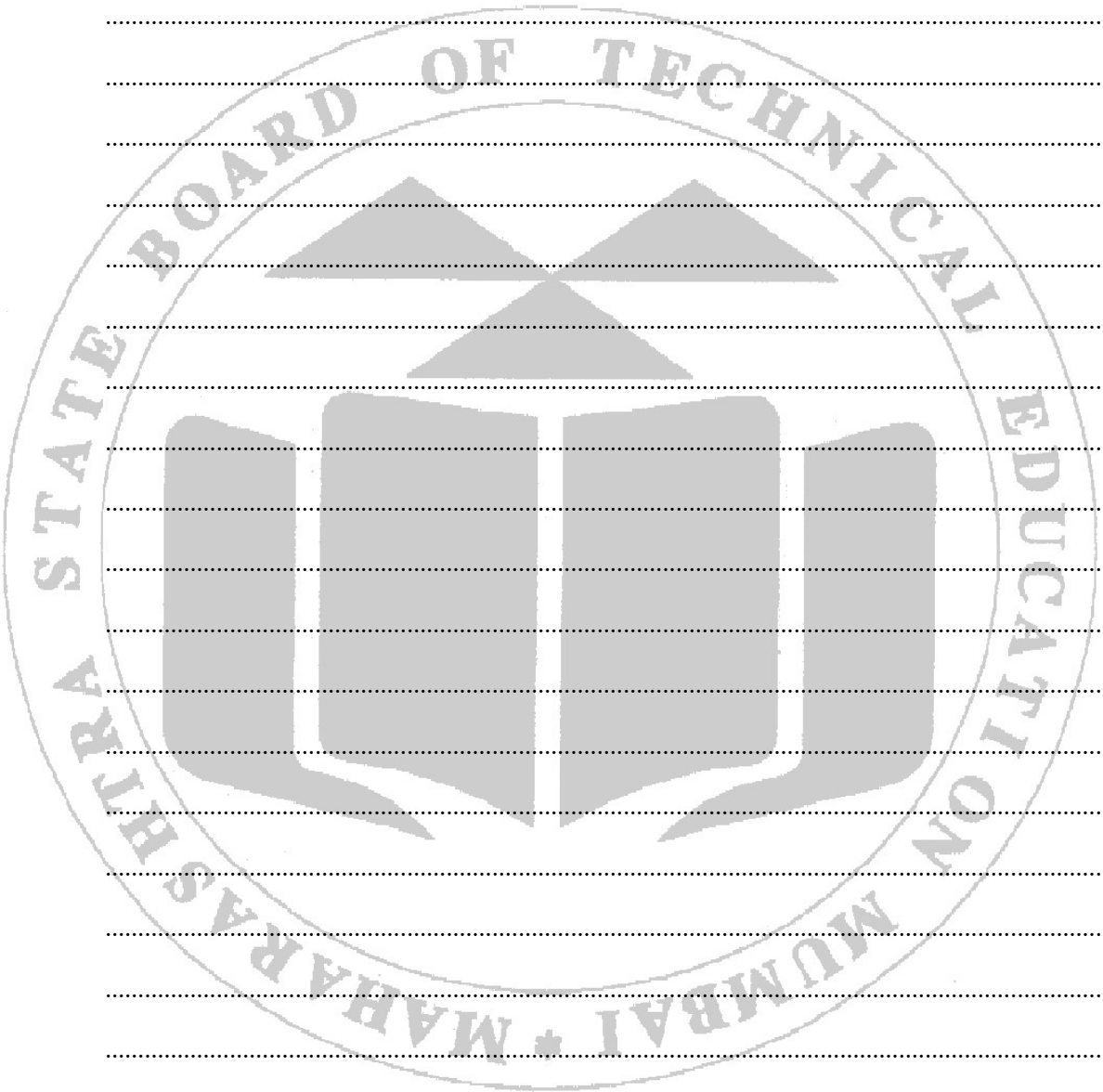
1. Write program to calculate the probability of getting heads when flipping a fair coin.
2. Write program to calculate the probability of drawing a spade from a standard deck of 52 cards.

.....

.....

.....

.....



IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write program to calculate the probability of getting tails when flipping a fair coin.
2. Write program to calculate the probability of getting all tails when flipping a fair coin 5 times.
3. Write program to calculate the Probability of rolling a specific sum with two dice.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
Product related (10 Marks)		**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 15

Write a program to calculate the probability of rolling a 3 on a fair six-sided die

I. Practical Significance

Classical probability theory provides a solid foundation for understanding random events and their likelihood in various scenarios. Students can explore mathematical foundations, properties, and practical codes of classical probability, offering insights suitable for both newcomers and seasoned data analysts seeking to grasp the core principles of probability in the R environment. This practical will make learner to use fundamental concepts of classical probability within the context of the R programming language.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO3 - Use Principles of Probability to solve the given Problem.

IV. Laboratory Learning Outcome(s)

LLO.15.1Implement Classical Probability using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

Classical Probability:

Classical probability, often referred to as “a priori” probability, is a branch of probability theory that deals with situations where all possible outcomes are equally likely. It provides a foundational understanding of how probability works and forms the basis for more advanced probability concepts.

Mathematical Foundations:

Sample Space: The sample space represents the set of all possible outcomes in a given experiment. It serves as the foundation for calculating probabilities. For instance, when rolling a fair six-sided die, the sample space is {1, 2, 3, 4, 5, 6}.

Events: An event is a subset of the sample space, representing a specific outcome or set of outcomes. Events can range from simple, such as rolling an even number, to complex, like drawing a red card from a deck.

Probability Distribution: A probability distribution assigns probabilities to each event in the sample space. For classical probability, all outcomes are equally likely, so each event has the same probability.

Calculating Classical Probability:

Classical probability is based on the principle of equally likely outcomes. Consider an experiment with a finite sample space S, consisting of n equally likely outcomes. Let A be an event of interest within S.

The classical probability of event A, denoted as P(A), is calculated as:

P(A) = Number of favorable outcomes for event A/Total number of equally likely outcomes in S

Mathematically, this can be expressed as:

P(A) = {n(A)}/{n(S)}

Where:

P(A) is the probability of event A.

n(A) is the number of favorable outcomes for event A.

n(S) is the total number of equally likely outcomes in the sample space S.

This formula allows us to calculate the probability of an event by counting the favourable outcomes and dividing by the total number of equally likely outcomes.

In R, you can use this formula to calculate classical probabilities for various events, making it a fundamental concept in probability theory for data analysis and statistics.

Example: Rolling a fair six-sided die

Probability of an event = (number of favorable event) / (total number of event).

P(A) =(Event A) / (total number of event).

Probability of getting any number = 1/6.

Rolling die is an equally likely event hence the outcome will have the same possibility.

Probability of getting 3 in a die = (number of favourable event) / (total number of event).

P(3) = (1)/(6).

Hence, the probability of getting 3 after tossing a rolling die is 1/6.

Sample Code:

```
calculate_probability<- function(favorable, total) {
```

```
if (favorable > total) {
 return("Invalid Input. Favorable outcomes must be less than or equal to the total number of outcomes")
} else if (favorable < 0 || total <= 0) {
 return("Invalid input. Favorable outcomes must be non-negative, and the total number of outcomes must be positive.")
} else {
 return(favorable / total)
}
}
```

 # Example usage:

```
result<- calculate_probability(3, 6)
cat("Probability:", result, "\n")

result<- calculate_probability(6, 3)
cat("Probability:", result, "\n")

result<- calculate_probability(-3, 6)
cat("Probability:", result, "\n")
```

Output:

Probability: 0.5

Probability: Invalid Input. Favorable outcomes must be less than or equal to the total number of outcomes

Probability: Invalid input. Favorable outcomes must be non-negative, and the total number of outcomes must be positive.

VII. Required Resources

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a program to calculate the probability of rolling a 3 on a fair six-sided die.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write program to calculate the probability of not rolling a 6 on a fair six-sided die.
2. Write a program to calculate the probability of rolling a 3 or 6 on a fair six-sided die.

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
Product related (10 Marks)		**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 16

Write a program to compute probability using prob( ) function

I. Practical Significance

The Probability theory is a fundamental concept in mathematics and statistics that plays a crucial role in various fields such as finance, engineering, medicine, and more. Understanding probabilities allows us to make informed decisions in uncertain situations. This practical will make learner to compute probability using prob() function within the context of the R programming language.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO3 - Use Principles of Probability to solve the given Problem.

IV. Laboratory Learning Outcome(s)

LLO.16.1 Compute probability using RProgram.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

Probability is the measure of the likelihood that an event will occur. The probability of an event A, denoted as P(A), lies between 0 and 1, where 0 indicates impossibility and 1 indicates certainty. Some key concepts include:
Sample Space (S): The set of all possible outcomes of a random experiment.
Event: Any subset of the sample space.
Probability of an Event: The likelihood of occurrence of an event, calculated as the ratio of favorable outcomes to the total number of outcomes.

Prob():Calculates probability and conditional probability of events.

Usage:

Prob(x, ...)

Default S3 method:

Prob(x, event = NULL, given = NULL, ...)

S3 method for class 'ps'

Prob(x, event = NULL, given = NULL, ...)

Arguments:

x: A probability space or a subset of one.

event: Logical expression indicating elements or rows of space to keep: missing values are taken asfalse.

given: Either a subset of a probability space or a logical expression indicating elements or rows ofspace to keep: missing values are taken as false.

... further arguments to be passed to or from other methods.

This function calculates the probability of events or subsets of a given sample space. Conditional probability is also implemented. In essence, the Prob() function operates by summing the probs column of its argument. It will find subsets on the fly if desired.

The event argument is used to define a subset of x, that is, the only outcomes used in the probability calculation will be those that are elements of x and satisfy event simultaneously. In other words, Prob(x,event) calculates Prob(intersect(x, subset(x, event))). Consequently, x should be the entire probability space in the case that event is non-null.

There is some flexibility in the given argument in that it can be either a data frame or it can be a logical expression that defines the subset. However, that flexibility is limited. In particular, if given is a logical expression, then event must also be specified (also a logical expression). And in this case, the argument x should be the entire sample space, not a subset thereof.

Example:

```
S <- rolldie(times = 3, makespace = TRUE )
Prob(S, X1+X2 >9 )
Prob(S, X1+X2 > 9, given = X1+X2+X3 >7 )
```

Output:

```
[1] 0.1666667
[1] 0.198895
```

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise
(Use blank space for answers or attach more pages if needed)

1. Write a program to compute probability using prob( ) function.

.....

.....

.....

IX. Practical related questions. (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write a program to calculate uniform distribution probability using prob.
2. Write a program to calculate the conditional probability using prob() function.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X.References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
Product related (10 Marks)		**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 17

Write a program to calculate the conditional probability

I. Practical Significance

The probability of occurrence of one event conditioned over the occurrence of another event i.e., an event occurs depending on the condition of another event is termed as conditional probability. This practical will make learner to compute conditional probability within the context of the R programming language.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO3 - Use Principles of Probability to solve the given Problem.

IV. Laboratory Learning Outcome(s)

LLO.17.1 Calculate the conditional probability using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

The probability of occurrence of one event conditioned over the occurrence of another event(i.e., an event occurs depending on the condition of another event) is termed as conditional probability. In simple terms, it means if A and B are two events, then the probability of occurrence of Event B conditioned over the occurrence of Event A is given by P(B|A). In another way, it is also the conditional probability of Event B given that event A has already occurred.

Similarly, the probability of occurrence of Event A conditioned over the occurrence of Event B is given by P(A|B), which also represents the conditional probability of Event A given that Event B has already occurred.

The formula for conditional probability can be represented as

$P(A|B) = P(A \cap B) / P(A)$

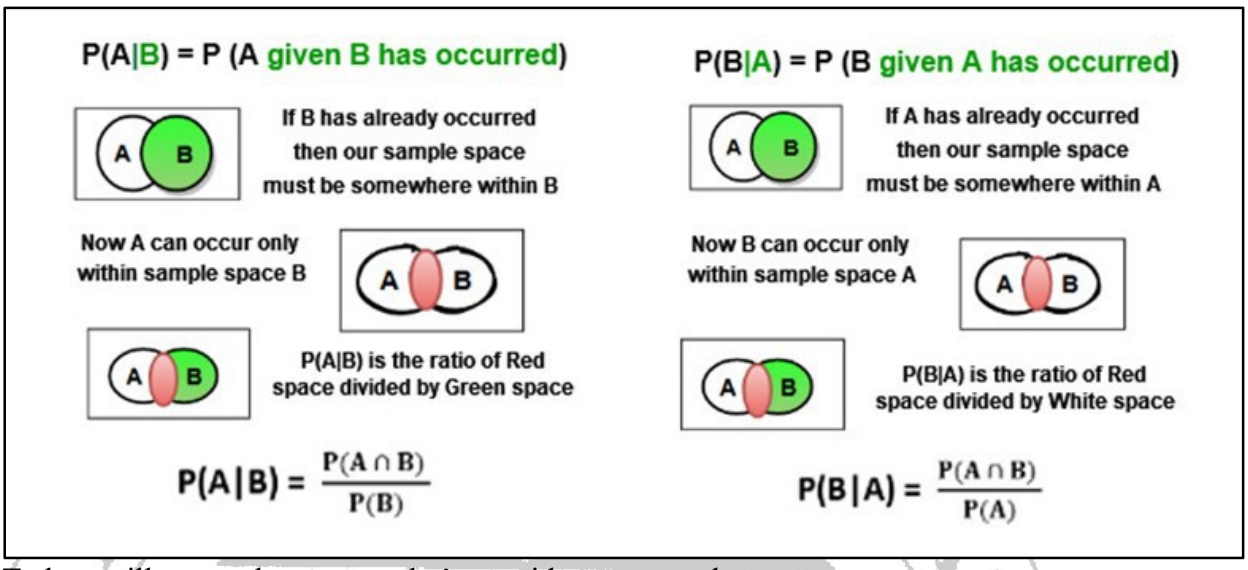
This is valid only when $P(A) \neq 0$ i.e. when event A is not an impossible event.

Similarly,

$P(B|A) = P(A \cap B) / P(B)$

This is valid only when $P(B) \neq 0$ i.e. when the event B is not an impossible event.

The below figure depicts the Venn diagram representation



To better illustrate the concept, let's consider an example:

Suppose we have a deck of 52 playing cards. We know that there are 13 hearts and 12 face cards in the deck. Let's calculate the conditional probability of drawing a face card, given that the card is a heart.

First, we need to determine the probability of drawing a heart (event A). There are 13 hearts in the deck, so: P(A) = 13 / 52 = 1/4
Next, we need to find the joint probability of drawing a face card and a heart (event A and B). There are 3 face cards that are also hearts (the King, Queen, and Jack of hearts), so: P(A and B) = 3 / 52
Finally, we can calculate the conditional probability of drawing a face card given that the card is a heart (P(B | A)): P(B | A) = P(A and B) / P(A) = (3 / 52) / (1/4) = 3/13
So, the probability of drawing a face card given that the card is a heart is 3/13 or approximately 0.2308.

Calculate Conditional Probability in R:
R is a powerful programming language for statistical computing and graphics. It offers various functions to calculate conditional probabilities. In this section, we will discuss a step-by-step process to calculate conditional probabilities in R using the prop.table() function.

Step 1: Create a Data Frame
First, create a data frame containing the variables A and B. Each row in the data frame represents an observation, while each column represents a variable.

Step 2: Create a Contingency Table
A contingency table, also known as a cross-tabulation or crosstab, is a tabular method to display the relationship between two or more categorical variables.
In R, you can create a contingency table using the table() function.

Step 3: Calculate the Conditional Probability Table
To calculate the conditional probability table P(B | A), use the prop.table() function in R. The prop.table() function converts a contingency table into a conditional probability table by dividing each cell by the row sums (i.e., the probabilities are conditioned on the first variable, A).

Step 4: Access Specific Conditional Probabilities
If you want to find a specific conditional probability, such as P(B=b1 | A=a1), you can access the corresponding cell in the conditional probability table using the appropriate row and column names.

Example: Calculating Conditional Probability for Student Information
In this example, we will calculate the conditional probability of passing an exam given high attendance.

Step 1: Create a Data Frame

```
student_data<- data.frame(
  Attendance = c("High", "High", "Low", "Low"),
  Pass = c("Yes", "No", "Yes", "No"),
  Frequency = c(80, 20, 30, 70)
)
```

Step 2: Calculate the Conditional Probability

```
total_high_attendance<-     sum(student_data$Frequency[student_data$Attendance     ==
"High"])
pass_and_high_attendance<-     student_data$Frequency[student_data$Attendance     ==
"High" &student_data$Pass == "Yes"]
P_pass_given_high_attendance<- pass_and_high_attendance / total_high_attendance
P_pass_given_high_attendance
```

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a program to calculate the conditional probability.
2. Write a program to calculate Conditional Probability Using a Contingency Table.
3. Write a program to calculate Conditional Probability Using dplyr.

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write a program to calculate the conditional probability of rain given the presence of clouds.
2. Write a program to calculate the conditional probability of drawing a face card given that the card is a heart.

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
Product related (10 Marks)		**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 18

Write a program to use Bayes' Theorem in R-Programming.

I. Practical Significance

Bayes' Theoremis used to determine the conditional probability of an event. Bayes Theorem is a very important theorem in mathematics that laid the foundation of a unique statistical inference approach called the Bayes' inference. It is used to find the probability of an event, based on prior knowledge of conditions that might be related to that event. This practical will make learner to use Bayes' Theorem in R-Programming.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO3 - Use Principles of Probability to solve the given Problem.

IV. Laboratory Learning Outcome(s)

LLO.18.1 Calculate the probability of an event based on given information using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

Bayes' Theorem states the following for any two events A and B:

$P(A|B) = P(A)*P(B|A) / P(B)$

where:

P(A|B): The probability of event A, given event B has occurred.

P(B|A): The probability of event B, given event A has occurred.

P(A): The probability of event A.

P(B): The probability of event B.

For example, suppose the probability of the weather being cloudy is 40%. Also suppose the probability of rain on a given day is 20% and that the probability of clouds on a rainy day is 85%.

If it's cloudy outside on a given day, what is the probability that it will rain that day?

Solution:

P(cloudy) = 0.40

P(rain) = 0.20

P(cloudy | rain) = 0.85

Thus, we can calculate:

P(rain | cloudy) = P(rain) * P(cloudy | rain) / P(cloudy)

P(rain | cloudy) = 0.20 * 0.85 / 0.40

P(rain | cloudy) = 0.425

If it's cloudy outside on a given day, the probability that it will rain that day is 42.5%.

We can create the following simple function to apply Bayes' Theorem in R:

```
#define function for Bayes' Theorem
bayesTheorem<- function(pA, pB, pBA) {
pAB<- pA * pBA / pB
return(pAB)
}
#define probabilities
pRain<- 0.2
pCloudy<- 0.4
pCloudyRain<- .85

#use function to calculate conditional probability
BayesTheorem(pRain, pCloudy, pCloudyRain)
```

Output:

[1] 0.425

VII. Required Resources

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)1.

1. Write a program to use Bayes' Theorem in R-Programming.

.....

.....

.....

.....

.....

.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write a program to create the following simple function to apply Bayes' Theorem in R. Assume that there is a 30% chance of rainfall in a day. Assume that the probability that I walk outside is 50% and that the probability that I walk on a rainy day is 10%. What are the chances that it will be a rainy day given I walk outside?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

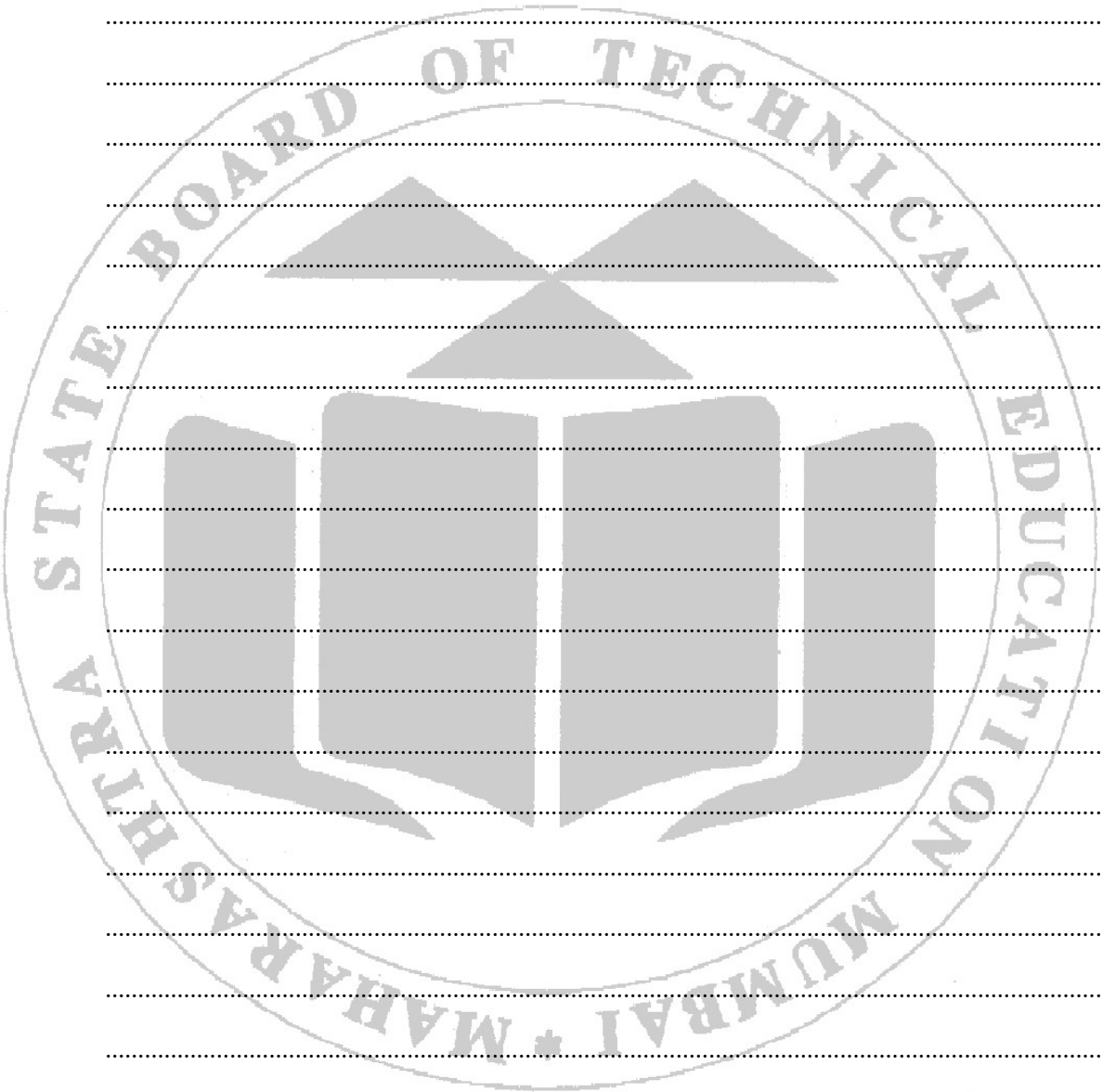
.....

.....

.....

.....

.....



.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

	Performance Indicators	Weightage
	Process related (15 Marks)	**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
	Product related (10 Marks)	**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 19

Write a Program to interpolate using Newton Forward Interpolation

I. Practical Significance

Interpolation is the technique of estimating the value of a function for any intermediate value of the independent variable, while the process of computing the value of the function outside the given range is called extrapolation. In this practical, students will be able to write a code to interpolate using newton forward interpolation.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO4 - Implement appropriate method based on the Interpolation.

IV. Laboratory Learning Outcome(s)

LLO.19.1 Find forward Interpolation using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

Newton forward difference formula is used to interpolate the values of y that is closer to the beginning of a set of tabular values. y0 may be taken as any point of the table, but the formula contains only those values of y which comes after the value chosen as y0 The Newton forward difference formula is given as:

$$f(x) = y_0 + r\Delta y_0 + \frac{r(r-1)}{2!}\Delta^2 y_0 + \frac{r(r-1)(r-2)}{3!}\Delta^3 y_0 + \frac{r(r-1)(r-2)(r-3)}{4!}\Delta^4 y_0 + \dots$$

Where

$$x = x_0 + rh$$

and

$$r = \frac{x - x_0}{h}$$

Sample Code:

```
#n = as.integer(readline(prompt = "Enter n: "))
##x = c(rep(0,n))
#y = c(rep(0,n))
#for(i in 1:n)
#{
#  x[i] = as.double(readline(prompt = "Enter x: "))
#  y[i] = as.double(readline(prompt = "Enter y: "))
#}
nfd = function(x,y,a){
n = length(x)
A = matrix(c(rep(0,(n-1)^2)), nrow = n-1, ncol = n-1, byrow = TRUE)
for (j in1:(n-1)){
for (iin1:(n-j))
  {
if(j==1)
A[i,j]=(y[i+1]-y[i])
else
A[i,j]=(A[(i+1),(j-1)]-A[i,(j-1)])
  }
}
#no = as.integer(readline(prompt = "Enter number to find: "))

u = (a - x[1])/(x[2]-x[1])
print(A)
sum = y[1]
for (iin1:(n-1))
{
prod = 1
for(j in0:(i-1))
prod = prod*(u-j)
prod = prod/factorial(i)
sum = sum + prod*A[1,i]
```

```
}
print(sum)
}
```

If the population of a town was given in the table below and we are to estimate the population for year 1895 using Newton forward difference formula,

Year	1891	1901	1911	1921	1931
Population	46	66	81	93	101

then we define the table by assigning Year to x and Population to y as a vector, as shown below.

```
x = c(1891,1901,1911,1921,1931)
y = c(46,66,81,93,101)
a = 1895
```

The function nfd(x, y, a) is then used to estimate the population, using the three parameters, where a is 1895.

```
nfd(x,y,a)
##      [,1] [,2] [,3] [,4]
## [1,]   20   -5    2   -3
## [2,]   15   -3   -1    0
## [3,]   12   -4    0    0
## [4,]    8    0    0    0
## [1] 54.8528
```

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a Program to interpolate using Newtonforward interpolation method.

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1) Write an R program to implement Newton's Forward Interpolation for Temperature Data.
2) Write an R program to implement Newton's Forward Interpolation for Stock Prices.
3) Write an R program to implement Newton's Forward Interpolation for Polynomial Data.

X.References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
Product related (10 Marks)		**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 20

Write a Program to interpolate using Newton Backward Interpolation

I. Practical Significance

Interpolation is the technique of estimating the value of a function for any intermediate value of the independent variable, while the process of computing the value of the function outside the given range is called extrapolation. In this practical, students will be able to write a code to interpolate using Newton backward interpolation..

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO4 - Implement appropriate method based on the Interpolation.

IV. Laboratory Learning Outcome(s)

LLO.20.1 Find backward Interpolation using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

Newton's forward difference formula cannot be used for interpolating a value of y near the end of table of values. For this purpose, we use another formula known as Newton backward difference formula and it can be written as:

$$f(x) = y_n + r\nabla y_n + \frac{r(r+1)}{2!}\nabla^2 y_n + \frac{r(r+1)(r+2)}{3!}\nabla^3 y_n + \frac{r(r+1)(r+2)(r+3)}{4!}\nabla^4 y_n + \ldots$$

Where

$$x = x_0 + rh$$

and

$$r = \frac{x - x_0}{h}$$

Sample Code:

```
#n = as.integer(readline(prompt = "Enter n: "))
#x = c(rep(0,n))
#y = c(rep(0,n))
#for(ii in 1:n)
#{
```

```
#  x[ii] = as.double(readline(prompt = "Enter x: "))
#  y[ii] = as.double(readline(prompt = "Enter y: "))
#}

nbd = function(x,y,a){
n = length(x)
A = matrix(c(rep(0,(n)^2)), nrow = n, ncol = n, byrow = TRUE)
for (j in1:(n-1)){
for (iin (j+1):n)
  {
if(j==1)
A[i,j]=(y[i]-y[i-1])
else
A[i,j]=(A[(i),(j-1)]-A[(i-1),(j-1)])
  }
}
#no = as.integer(readline(prompt = "Enter no to find: "))
u = (a - x[n])/(x[2]-x[1])
print(A)
sum = y[n]
for (iin1:(n-1))
{
prod = 1
for(j in0:(i-1))
prod = prod*(u+j)
prod = prod/factorial(i)
sum = sum + prod*A[n,i]
}
print(sum)
}
```

Using same tabular values used for the Newton forward difference, estimate for the year 1925 using Newton backward difference.

x	1891	1901	1911	1921	1931

y	46	66	81	93	101

Following the same procedure as the Newton forward difference, we have:

```
x = c(1891,1901,1911,1921,1931)
y = c(46,66,81,93,101)
a = 1925
nbd(x,y,a)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0    0    0    0    0
## [2,]   20    0    0    0    0
## [3,]   15   -5    0    0    0
## [4,]   12   -3    2    0    0
## [5,]    8   -4   -1   -3    0
## [1] 96.8368
```

VII. Required Resources

Sr. No	Name of Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1. Write a Program to interpolate using Newton backward interpolation method.
2. Write a Program to interpolate using Newton's Backward Interpolation for Temperature Data.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

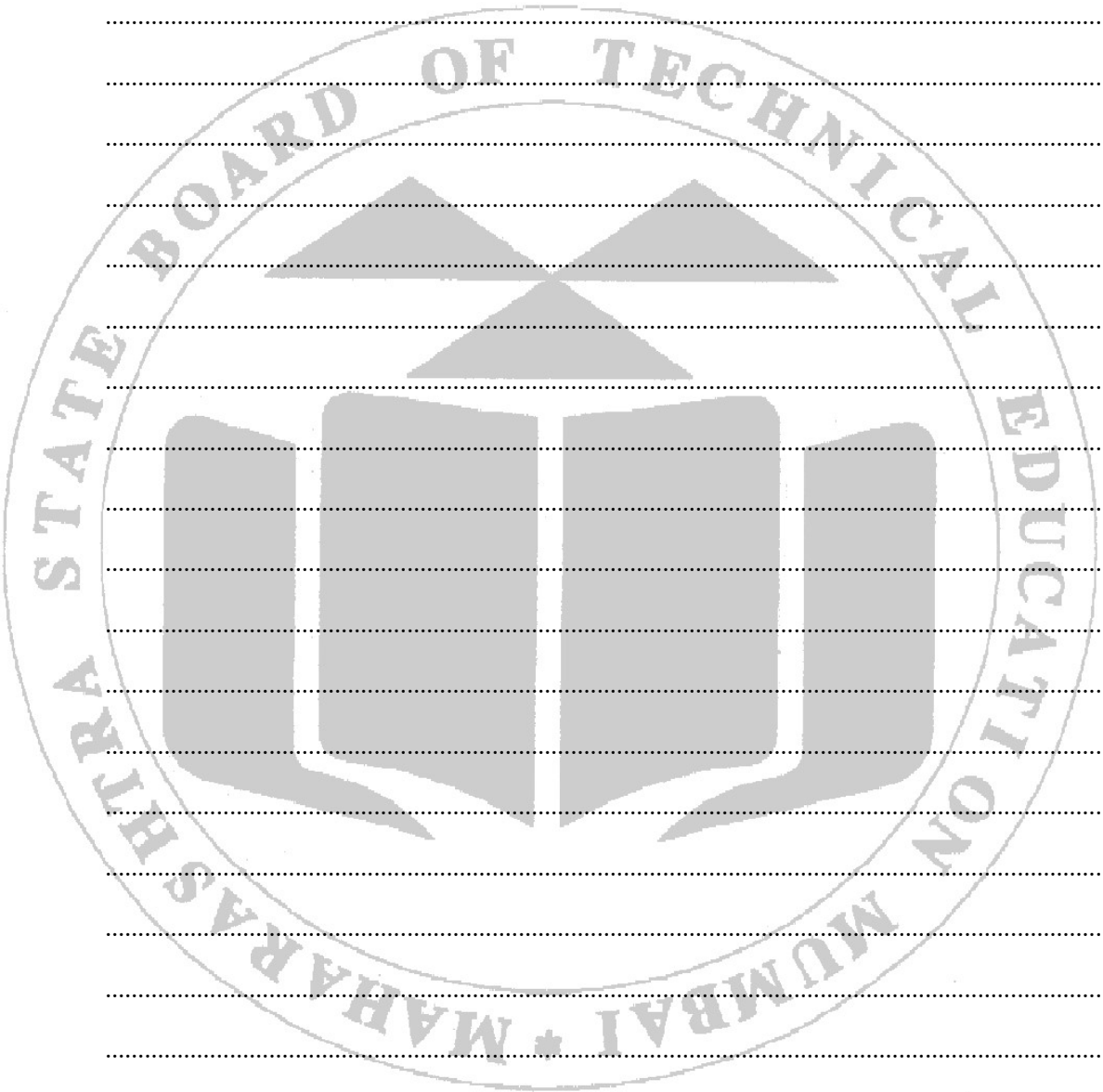
.....

.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write an R program to implement Newton's Backward Interpolation method for Stock Prices.
2. Write an R program to implement Newton's Backward Interpolation method for Polynomial Data.
3. Write an R program to implement Newton's Backward Interpolation method for Experimental Data.

.....



.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15 Marks)		**70%**
1	Logic Formulation	1
2	Debugging Ability	2
3	Follow ethical practices	3
Product related (10 Marks)		**30%**
4	Expected output	4
5	Timely Submission of report	5
6	Answer to sample questions	6
	Total (25 Marks)	

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 21

Write a program for the implementation of extrapolation

I. Practical Significance

Extrapolation is the process in mathematics where the required value is estimated beyond the range of the given variable range. Extrapolation is often used to estimate the data of some observation below or above the given range. Extrapolation is also referred to as a mathematical prediction to predict values by observing the relationship between the given variables. In this practical, students will learn the implementation of extrapolation in R programming.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO4 - Implement appropriate method based on the Interpolation.

IV. Laboratory Learning Outcome(s)

LLO.21.1 Find Extrapolation using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

What is Extrapolation?

Extrapolation is the process in mathematics where the required value is estimated beyond the range of the given variable range. Extrapolation is often used to estimate the data of some observation below or above the given range. Extrapolation is also referred to as a mathematical prediction to predict values by observing the relationship between the given variables. There are many processes of Extrapolation. This process was first described by Thomas D. Clares on in 1959 in his book of science. He referred to it as a meaningful prediction by understanding the given data.

How to calculate Linear Extrapolation?

The method is useful when the linear function is given. It is done by drawing a tangent and extending it beyond the limit. Linear Extrapolation gives a very good result when the point to be predicted is not very far from the rest of the points.

Extrapolation formula:

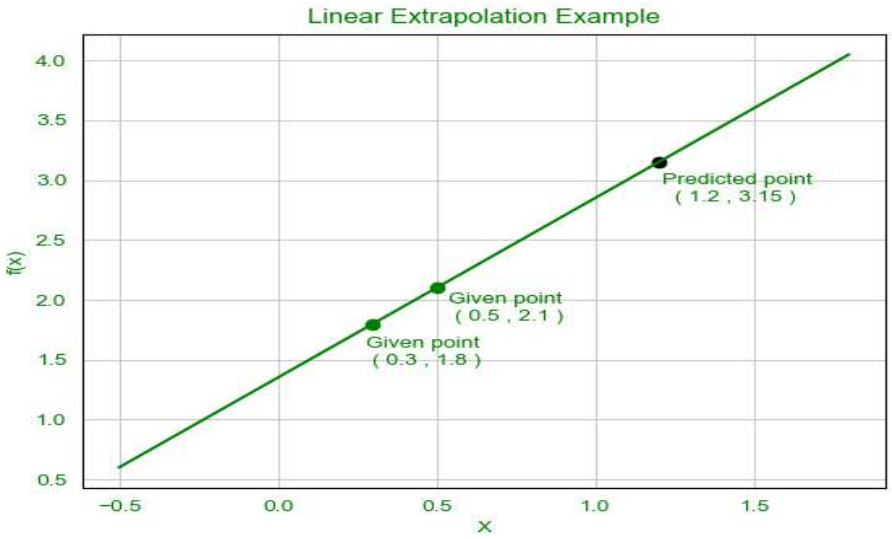
$$y(x) = y_1(x) + \frac{x - x_1}{x_2 - x_1}(y_2 - y_1)$$

Here (x_1, y_1) and (x_2, y_2) are two given points and x is the point for which we want to predict the value of y.

Example:

Input: x_1=0.3, y_1=1.8 , x_2=0.5, y_2=2.1 , x = 1.2

Output: y = 3.15



VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1) Write a R program to demonstrate implementation of extrapolation method.

.....

.....

.....

.....

.....

.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1) Write a R program to demonstrate implementation of extrapolation method for Experimental Data. .

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 22

Write a program to generate Samples using the Sampling Functions

I. Practical Significance

Sampling is the process of selecting a small subset from the available population. This small subset is used as a representation of the entire population. Sampling is helpful because it saves the time and resources that would otherwise be used to collect data from the entire population. The sample function in R is a tool used to generate random samples from a specified set of elements. This practical will let learner to develop program to generate samples using the sample function.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO.22.1Generate Samples for the given dataset using R Program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

How to Generate a Sample Using the Sample Function in R?

Sample() function is used to generate the random elements from the given data with or without replacement.

Syntax:

sample(data, size, replace = FALSE, prob = NULL)

where,

data can be a vector or a dataframe

size represents the size of the sample

replace is used to set the values again repeated if it is set to true

prob: a vector of probability weights for obtaining the elements of the vector being sampled

Example: Generate sample data from the vector

Here, we will generate the n sample data from the given vector with 11 elements using the sample

function.

Sample code:

```
# consider the vector
data=c(23,45,21,34,5,6,7,8,86,45,3)
# get 4 random elements
print(sample(data,4))
# get 1 random element
print(sample(data,1))
# get 6 random elements
print(sample(data,6))
```

Output:

```
[1] 45  7  5 34
[1] 3
[1]  5 23  8 21  6 45
```

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1) Write a program to create a vector with 11 elements and generate the sample data with a replacement.
2) Write a program for sampling with uneven probabilities using sample function in R.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1) Write a program to sample the data in the list with size 4.
2) Write a program to sample of dataframe rows using sample function in R.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 23

Write programs to perform following types of sampling
a. Simple Random Sampling
b. Stratified Sampling
c. Systematic Sampling
d. Biased Sampling

I. Practical Significance

Sampling is the process of selecting a small subset from the available population. This small subset is used as a representation of the entire population. Sampling is helpful because it saves the time and resources that would otherwise be used to collect data from the entire population. This practical will let learner to develop program to generate samples using the various types of sampling methods.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO.23.1Perform the given type of sampling using R program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

What is sampling?
Let's say that we have a population of size N, a sample is nothing but a subset of data taken from that population. The process of selecting a sample is known as sampling.

- **Simple Random Sampling in R:**

Simple random sampling can be done in R using the function 'sample()'. It accepts an argument called replace to decide whether it has to be with replacement or without replacement of values in the original set. Basically, whether we can repeat values in the selection or not.

replace=TRUE mean repetitions of values are allowed.
replace=FALSE mean repetitions of values are NOT allowed.

Let us look at the below example. From the first 100 numbers, if you want to select 12 numbers randomly without repetition.

Here, the function parameters are explained below.
x: Range of values from where random selection has to be performed
size: number of values to select
replace: whether to allow repetition or not.
Simple Random Sampling Without Replacement

```
# Defining range of values from 1 to 100
SampleRange=c(1:100)
# Choosing a sample of size 12 without repetition
sample(x=SampleRange,size=12, replace=FALSE)
```

Output:

```
> sample(x=SampleRange,size=12, replace=FALSE)
 [1] 66 57 12 58 35 41  5 25 37 77 78 55
```

- **Stratified Sampling in R:**

Stratified sampling needs a grouping reference. Based on this column data rows can be selected from each group.
There are many libraries in R to perform stratified sampling, two of the easiest to use are listed below,
library(sampling) —> strata() function
library(caret) —>createDataPartition() function

```
# Defining Sample Data for Stratified Sampling
GroupName=c('A','A','B','B','B','B','C','C','C')
Value=c(2,1,50,55,57,58,201,205,207)
SampleData=data.frame(GroupName,Value)
print(SampleData)
# Performing Stratified Sampling using library caret
library(caret)
SampleIndex=createDataPartition(SampleData$GroupName, p= 0.5, list=FALSE)
# SampleIndex will have rows selected from each group
print(SampleIndex)
# Performing Stratified Sampling using library sampling
library(sampling)
SampleIndexNew=strata(SampleData, c('GroupName'), size=c(1,2, 3), method="srswor")
# SampleIndexNew will have mentioned rows selected from each group
print(SampleIndexNew)
```

Output:

```
> print(SampleData)
  GroupName Value
1         A     2
2         A     1
3         B    50
4         B    55
5         B    57
6         B    58
7         C   201
8         C   205
9         C   207
```

Sample Data for Stratified Sampling

Output (SampleIndex):

```
> print(SampleIndex)
     Resample1
[1,]         2
[2,]         3
[3,]         6
[4,]         7
[5,]         9
```

Stratified sampling using createDataPartition () function from library(caret)

Output (SampleIndexNew):

```
> print(SampleIndexNew)
  GroupName ID_unit Prob Stratum
1         A       1  0.5       1
3         B       3  0.5       2
5         B       5  0.5       2
7         C       7  1.0       3
8         C       8  1.0       3
9         C       9  1.0       3
```

Stratified sampling using strata() function from library(sampling)

- **Systematic Sampling in R:**

seq() function in R helps to generate the ith index

Systematic Sampling means just select every ith value from the dataset.

Generate every 5th index between range 0 to 50

seq(from = 0, to = 50, by =5)

Output:

```
> seq(from = 0, to = 50, by =5)
 [1]  0  5 10 15 20 25 30 35 40 45 50
```

- **Biased Sampling in R:**

Biased sampling does not require any special function in R. One can select any index of value as per the need and inspect it.

Selecting first five index for inspection

selectedIndex=c(1:5)

print(selectedIndex)

Output:

```
> selectedIndex=c(1:5)
> print(selectedIndex)
[1] 1 2 3 4 5
```

Biased sampling in R using manual indexing

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1) Write a program to demonstrate simple random sampling in R.
2) Write a program to demonstrate stratified sampling in R.

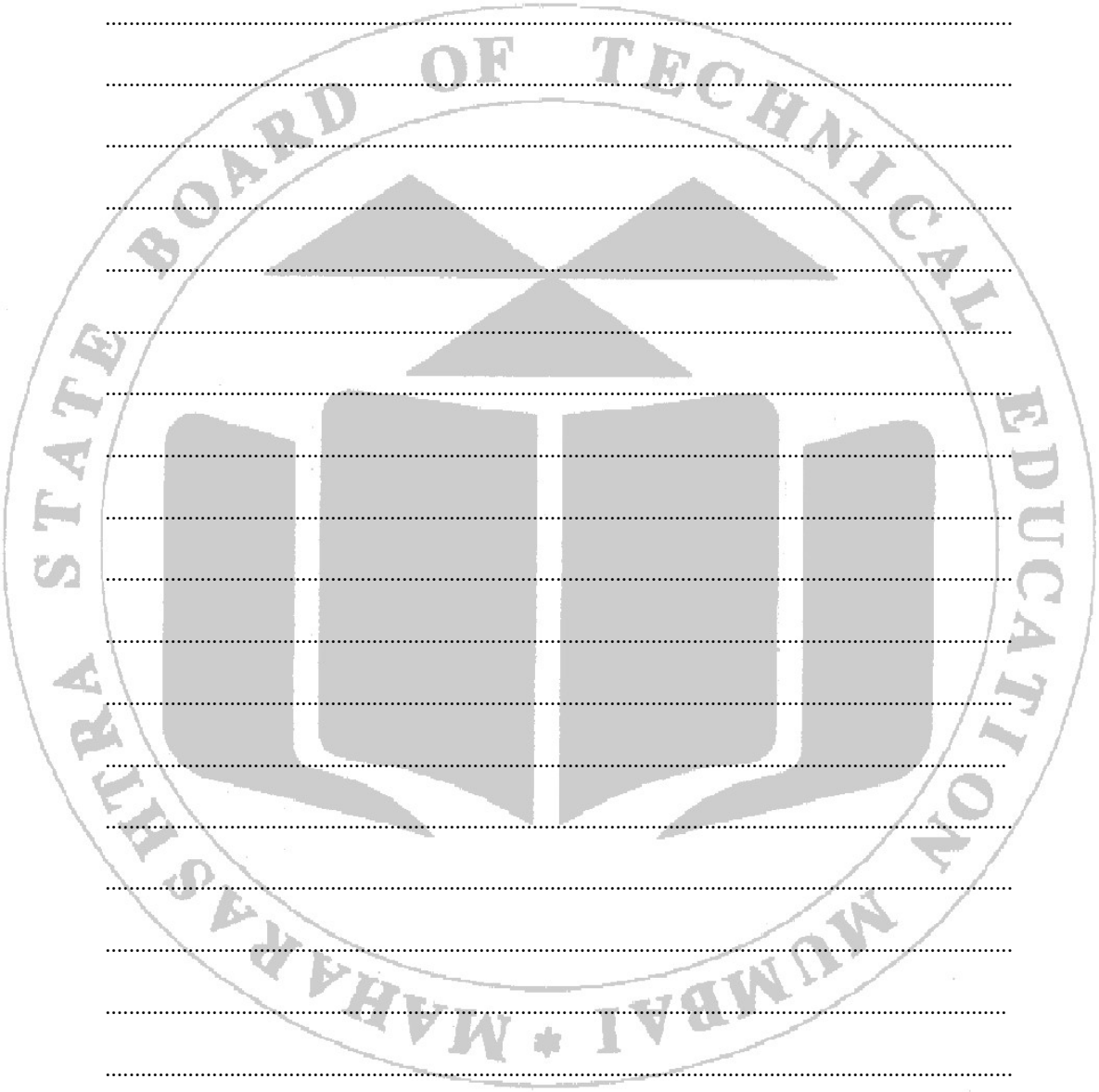
.....

.....

.....

.....

.....



IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1) Write a program to demonstrate systematic sampling in R.
2) Write a program to demonstrate biased sampling in R.

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 24

Write a program to generate a Sampling Distribution proportion

I. Practical Significance

A sampling distribution is a probability distribution of a statistic obtained from a larger number of samples drawn from a specific population. The sampling distribution of a given population is the distribution of frequencies of a range of different outcomes that could possibly occur for a statistic of a population..This practical will let learner to develop program to generate a sampling distribution proportion in R.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO.24.1Generate Sampling Distribution proportion using Rprogram.

VI. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

In statistics, a population is an entire pool from which a statistical sample is drawn. A population may refer to an entire group of people, objects, events, hospital visits, or measurements. A population can thus be said to be an aggregate observation of subjects grouped together by a common feature.

- A sampling distribution is a statistic that is arrived out through repeated sampling from a larger population.
- It describes a range of possible outcomes that of a statistic, such as the mean or mode of some variable, as it truly exists a population.
- The majority of data analyzed by researchers are actually drawn from samples, and not populations.

Steps to Calculate Sampling Distributions in R:

Step 1: Here, first we have to define a number of samples(n=1000).

```
n<-1000
```

Step 2: Next we create a vector(sample_means) of length 'n' with Null(NA) values [rep() function is used to replicate the values in the vector

Syntax: rep(value_to_be_replicated,number_of_times)

Step 3: Later we filled the created sample_means null vector with sample means from the considered population using the mean() function which are having a sample mean of 10(mean) and standard deviation of 10(sd) of 20 samples(n) using rnorm() which is used to generate normal distributions.

Syntax: mean(x, trim = 0)

Syntax: rnorm(n, mean, sd)

Step 4: To check the created samples we used head() which returns the first six samples of the dataframe (vector,listetc,.).

Syntax: head(data_frame,no_of_rows_be_returned) #By default second argument is set to 6 in R.

Step 5: Finally to visualize the sample_mean dataset we plotted a histogram (for better visualization) using hist() function in R.

Syntax: hist(v,main,xlab,ylab,col)

where.

v is a vector containing values used in histogram.

main indicates title of the chart.

col is used to set color of the bars.

xlab is used to give description of x-axis.

ylab is used to give description of y-axis.

Step 6: Finally we found the probability of generated sample means which are having mean greater than or equal to 10.

Sample Example: In this particular example, we find the probability that the sample mean is less than or equal to 10, given that the population means is 10, the population standard deviation is 10, and the sample size is 20 is 0.506(approx 0.50).

Sample code:

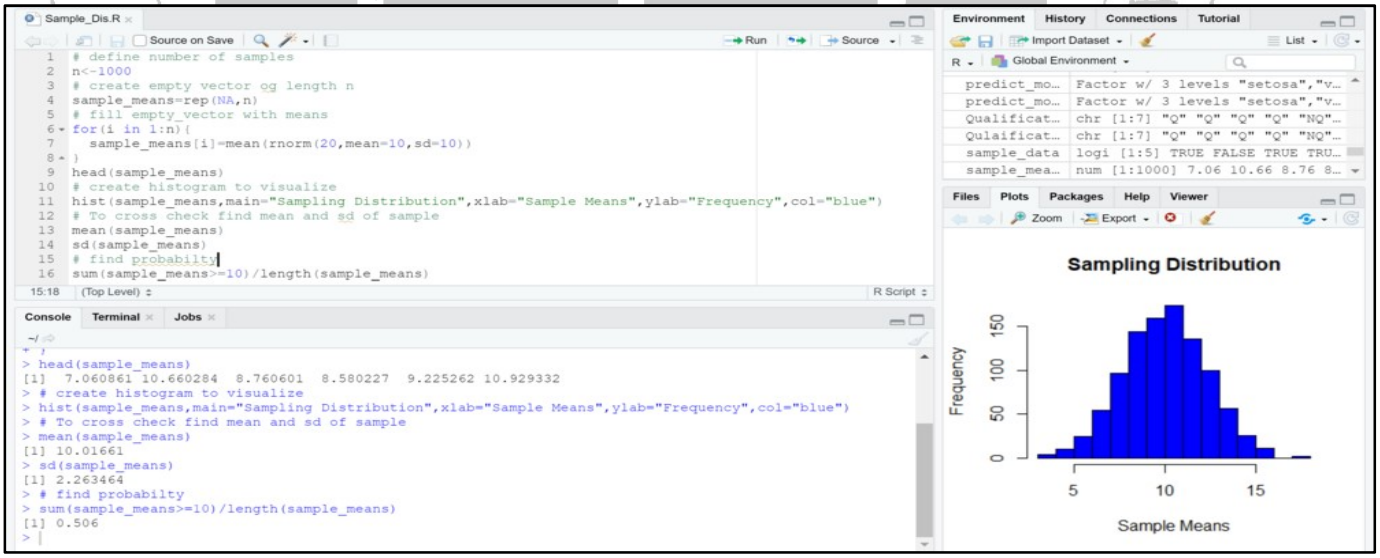
```
# define number of samples

n < -1000

# create empty vector of length n
```

```
sample_means = rep(NA, n)
# fill empty_vector with means
for(i in 1: n){
sample_means[i] = mean(rnorm(20, mean=10, sd=10))
}
head(sample_means)
# create histogram to visualize
hist(sample_means, main="Sampling Distribution",
xlab="Sample Means", ylab="Frequency", col="blue")
# To cross check find mean and sd of sample
mean(sample_means)
sd(sample_means)
# To find probability
sum(sample_means>= 10)/length(sample_means)
```

Output:



VII. Required Resources:

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1) Write a program to implement sampling distribution proportion in R.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1) Write a program to demonstrate implementation of sampling distribution proportion on experimental data in R.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 25

Write a program based on t-Distribution using dt, pt, qt & rt functions

I. Practical Significance

The t-distribution is a type of probability distribution that arises while sampling a normally distributed population when the sample size is small and the standard deviation of the population is unknown. It is also called the Student's t-distribution. This practical will let learner to develop program to generate a t-distribution in R.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO.25.1Perform Hypothesis Testing by t-Distribution using R program.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

The t-distribution is a type of probability distribution that arises while sampling a normally distributed population when the sample size is small and the standard deviation of the population is unknown. It is approximately a bell curve, that is, it is approximately normally distributed but with a lower peak and more observations near the tail.

Student's t-distribution in R

Functions used:

To find the value of probability density function (pdf) of the Student's t-distribution given a random variable x, use the dt() function in R.

Syntax: dt(x, df)

Parameters:

x is the quantiles vector

df is the degrees of freedom

pt() function is used to get the cumulative distribution function (CDF) of a t-distribution

Syntax: pt(q, df, lower.tail = TRUE)

Prameter:

q is the quantiles vector

df is the degrees of freedom

lower.tail – if TRUE (default), probabilities are P[X ≤ x], otherwise, P[X > x].

The qt() function is used to get the quantile function or inverse cumulative density function of a t-distribution.

Syntax: qt(p, df, lower.tail = TRUE)

Parameter:

p is the vector of probabilities

df is the degrees of freedom

lower.tail – if TRUE (default), probabilities are P[X ≤ x], otherwise, P[X > x].

rt() function in R Language is utilized to make an irregular succession of values from Student t-distribution.

Syntax: rt(n, df, ncp)

parameters:

n: is a number of observation

df: level of opportunity or degree of freedom

ncp: Non- Centrality parameters

Example:

Here, we find the value of the Student t distribution pdf by passing different x and the degree parameter to the df() function in the R language.

```
# here we find the value of the Student

# t distribution pdf at x = 1 with 30

# degrees of freedom

dt(x = 1, df = 30)
```

```
# here we take by default, R assumes the
# first argument is x and the second
# argument is df
dt(1, 30)
# find the value of the Student t distribution
# pdf at x = 2 with 40 degrees of freedom
dt(2, 40)
```

Output:

```
#[1] 0.2379933
#[1] 0.2379933
#[1] 0.05618831
```

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating System	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1) Write a program to calculate t-distribution using dt function in R.
2) Write a program to calculate t-distribution using pt function in R.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1. Write a program to calculate t-distribution using qt function in R.
2. Write a program to calculate t-distribution using rt function in R.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	

Practical No. 26

Write a program based on Chi-Square Distribution using dchisq, pchisq, qchisq & rchisq functions

I. Practical Significance

The chi-squared distribution with df degrees of freedom is the distribution computed over the sums of the squares of df independent standard normal random variables. This distribution is used for the categorical analysis of the data. This practical will let learner to develop program to generate a chi-squared distribution in R.

II. Industry / Employer Expected Outcome(s)

Apply Mathematics to solve real-world problems using AI/ML concept and principles to enhance decision-making, design and innovation with precision and efficiency.

III. Course Level Learning Outcome(s)

CO5 - Apply Sampling Methods to solve given problem using R-Programming.

IV. Laboratory Learning Outcome(s)

LLO.26.1Perform Hypothesis Testing by Chi-Square Distribution using Rprogram.

V. Relevant Affective Domain related Outcomes

- Follow safety practices.
- Develop a positive and proactive mindset towards programming in R.
- Follow ethical practices.

VI. Relevant Theoretical Background

The chi-squared distribution in R: Functions

- **qchisq() function:**

qchisq gives the quantile function. When we supply the value of ncp = 0, the algorithm for the non-central distribution is used. The value of this method is equivalent to the value of x at the qth percentile (lower.tail = TRUE).

Syntax:

qchisq(p, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)

Parameter :

p – vector of probabilities

df – degrees of freedom

ncp – non-centrality parameter (non-negative).

log.p – logical; if TRUE, probabilities p are given as log(p).

lower.tail – logical; if TRUE (default), probabilities are P[X ≤ x], otherwise, P[X > x].

Example:

```
# defining the degrees of freedom
free = 5
qchisq(.75, df=free)
```

Output

```
[1] 6.62568
```

- **dchisq() function:**

dchisq gives the density function. That is, it is used for computing the cumulative probability (lower.tail = TRUE for left tail, lower.tail = FALSE for right tail) of less than or equal to the value of vector of quantiles, that is q.

Syntax:

```
dchisq(x, df, ncp = 0, log = FALSE)
```

Parameter :

x – vector of quantiles

df – degrees of freedom

ncp – non-centrality parameter (non-negative).

log.p – logical; if TRUE, probabilities p are given as log(p).

Example:

```
# defining degrees of freedom
df = 6
vec<- 1:4
```

```
print ("Density function values")

dchisq(vec, df = df)
```

Output

[1] “Density function values”

[1] 0.03790817 0.09196986 0.12551072 0.13533528

- **pchisq() function:**

pchisq gives the distribution function. dchisq(x, df) gives us the probability of $\chi 2$ with equivalent to a value of x when the degree of freedom is df. This method can be used to calculate the area under the curve for the specified intervals of the $\chi 2$-curve with a given number of degree of freedoms.

Syntax:

```
pchisq(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
```

Parameter :

q – vector of quantiles

df – degrees of freedom

ncp – non-centrality parameter (non-negative).

log.p – logical; if TRUE, probabilities p are given as log(p).

lower.tail – logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.

Example:

```
# defining degrees of freedom

df = 5

# calculating for the values in the interval [0,5]

print ("Calculating for the values [0,5]")

pchisq(5, df = df,lower.tail = TRUE)

# calculating for the values in the interval [5,inf)

print ("Calculating for the values [5,inf)")
```

pchisq(5, df = df,lower.tail = FALSE)

Output

[1] “Calculating for the values [0,5]”

[1] 0.5841198

[1] “Calculating for the values [5,inf)”

[1] 0.4158802

- **rchisq() function:**

rchisq(n, df) returns n random numbers from the chi-square distribution. It is therefore to generate random deviates.

Syntax:

rchisq(n, df, ncp = 0)

Parameter :

n – number of observations. If length(n) > 1, the length is taken to be the number required.

df – degrees of freedom (non-negative, but can be non-integer).

ncp – non-centrality parameter (non-negative).

Example:

```
# computing values of 50k random values with 5 degrees of freedom
x <- rchisq(50000, df = 5)
hist(x,
freq = FALSE,
xlim = c(0,16),
ylim = c(0,0.2))
curve(dchisq(x, df = 5), from = 0, to = 15,
    n = 5000, col= 'red', lwd=2, add = T)
```

VII. Required Resources

Sr. No	Name Resource	Specification	Quantity	Remarks
1	Computer System	Any desktop or laptop computer with basic configuration	One computer system for each student	
2	Operating system	Windows/LINUX	One for each computer system	
3	Software	R Studio latest version	One for each computer system	

VIII. Exercise

(Use blank space for answers or attach more pages if needed)

1) Write a program to calculate chi-square distribution using dchisq function in R.
2) Write a program to calculate chi-square distribution using pchisq function in R.

IX. Practical related questions (Note: Below given are few sample questions for reference. Teachers must design more such questions to ensure the achievement of identified CO)

1) Write a program to calculate chi-square distribution using qchisq function in R.
2) Write a program to calculate chi-square distribution using rchisq function in R.

.....

.....

X. References/Suggestions for further reading

1. https://www.w3resource.com/r-programming-exercises/basic/
2. https://www.tutorialspoint.com/r_programming_language/index.asp
3. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/
4. http://nptel.ac.in/courses/106102064/1
5. www.mathworks.com/

XI. Assessment Scheme (25 Marks)

Performance Indicators		Weightage
Process related (15Marks)		**70%**
1	Logic Formulation	10%
2	Debugging Ability	20%
3	Follow ethical practices	40%
Product related (10Marks)		**30%**
4	Expected output	10%
5	Timely Submission of report	10%
6	Answer to sample questions	10%
	Total (25Marks)	**100%**

Marks Obtained			Dated signature of Teacher
Process Related (15)	**Product Related (10)**	**Total (25)**	