



# PRESIDENCY UNIVERSITY

BENGALURU

Roll No.														
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--

## End - Term Examinations – MAY 2025

Date: 21-05-2025

Time: 01:00 pm – 04:00 pm

School: SOIS	Program: MCA	
Course Code: CSA4036	Course Name: R Programming for Data Science	
Semester: IV	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04	C05
Marks	25.5	25.5	24.5	24.5	-

### Instructions:

- (i) Read all questions carefully and answer accordingly.
- (ii) Do not write anything on the question paper other than roll number.

### Part A

Answer ALL the Questions. Each question carries 2marks.

10Q x 2M=20M

1.	What is R, and why is it widely used in data analysis?	2 Marks	L1	C01
2.	What is the role of the aes() function in ggplot2?	2 Marks	L1	C01
3.	What is the difference between select() and mutate() in dplyr?	2 Marks	L1	C01
4.	List the common dataset formats.	2 Marks	L1	C02
5.	Summarize the common Causes of Anomalies.	2 Marks	L1	C02
6.	How can you detect anomalies in numerical data? Name one method.	2 Marks	L1	C02
7.	What is the purpose of using a training and test dataset in regression analysis?	2 Marks	L1	C03
8.	What are outliers, and how do they affect regression models?	2 Marks	L1	C03
9.	What is the output of logistic regression in a classification problem?	2 Marks	L1	C04
10.	Write the advantages of using SVM for classification.	2 Marks	L1	C04

## Part B

### Answer the Questions.

**Total Marks 80M**

<b>11.</b>	<b>a.</b>	<p>Apply and implement a program in R to determine whether a person is eligible to vote using an If-Else statement.</p> <p>(i)The program should take the person's age as input and display the eligibility status based on the following criteria: 18 years and above → Print "Eligible to vote" Below 18 years → Print "Not eligible to vote"</p> <p>(ii)Write the R program to check voting eligibility based on these conditions.</p>	<b>20 Marks (10+10)</b>	<b>L3</b>	<b>C01</b>
<b>Or</b>					
<b>12.</b>	<b>a.</b>	<p>Build a program in R to classify student grades based on their scores using an If-Else statement.</p> <p>(i)The program should take a student's score as input and assign a grade based on the following criteria: 90 and above → Grade A 80 to 89 → Grade B 70 to 79 → Grade C 60 to 69 → Grade D Below 60 → Grade F</p> <p>(ii)Write the R program to determine the grade of a student based on these conditions.</p>	<b>20 Marks (10+10)</b>	<b>L3</b>	<b>C01</b>
<b>13.</b>	<b>a.</b>	Describe the concept of Covariation and explain how to compute the covariance matrix using a sample dataset. Provide a detailed explanation of the steps involved and demonstrate the computation with an example dataset.	<b>10 Marks</b>	<b>L2</b>	<b>C02</b>
	<b>b.</b>	Explain the concepts of Patterns and Models. Discuss their importance in data analysis and provide examples of how they are used to derive insights from data.	<b>10 Marks</b>	<b>L2</b>	<b>C02</b>
<b>Or</b>					
<b>14.</b>	<b>a.</b>	Explain various visualization techniques used in data analysis. Provide examples using a sample dataset to demonstrate how these techniques can be applied to uncover insights.	<b>10 Marks</b>	<b>L2</b>	<b>C02</b>
	<b>b.</b>	Describe missing values in a dataset with an example. Discuss the methods available for handling missing values effectively.	<b>10 Marks</b>	<b>L2</b>	<b>C02</b>
<b>15.</b>	<b>a.</b>	<p>Experiment the following tasks with the iris dataset, which contains four attributes (Sepal Length, Sepal Width, Petal Length, Petal Width) and a categorical variable (Species):</p> <p>(i)Demonstrate the R code to install and load the necessary packages. Split the dataset into 70% training and 30% testing sets.</p>	<b>20 Marks (10+10)</b>	<b>L3</b>	<b>CO 3</b>

		(ii) Demonstrate the R code to identify patterns and Models in the dataset.			
<b>Or</b>					
<b>16.</b>	<b>a.</b>	Apply the data points: 3, 6, 7, NaN, 9, treat them as values at equally spaced time intervals (e.g., t = 1 to 5). (i) Apply a simple linear regression model to estimate the missing value (NaN). (ii) Show all steps and calculations clearly.	<b>20 Marks (10+10)</b>	<b>L3</b>	<b>CO 3</b>

<b>17.</b>	<b>a.</b>	(i) Demonstrate a decision tree model using the Iris dataset. (ii) Show the tree and explain how the model splits based on attribute values.	<b>20 Marks (10+10)</b>	<b>L2 L2</b>	<b>CO 4</b>
<b>Or</b>					
<b>18.</b>	<b>a.</b>	(i) Outline logistic regression on the Iris dataset to classify flower species. (ii) Demonstrate R code and interpret the output using accuracy and a confusion matrix.	<b>20 Marks (10+10)</b>	<b>L2 L3</b>	<b>CO 4</b>