Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025

Date: 15-12-2025 **Time:** 1.00pm to 04.00pm

School: SOCSE	**Program:** B.Tech(CSI)	
Course Code : CSD2007	**Course Name:** R Programming for Data Science	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4
Marks	26	26	24	24

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	CO1
2.	What is the role of the aes() function in ggplot2?	2 Marks	L1	CO1
3.	What is the difference between select() and mutate() in dplyr?	2 Marks	L1	CO1
4.	List the common dataset formats.	2 Marks	L1	CO2
5.	Summarize the common Causes of Anomalies.	2 Marks	L2	CO2
6.	Explain How can you detect anomalies in numerical data? Name one method.	2 Marks	L2	CO2
7.	What is the purpose of using a training and test dataset in regression analysis?	2 Marks	L1	CO3
8.	What are outliers, and how do they affect regression models?	2 Marks	L1	CO3
9.	What is the output of logistic regression in a classification problem?	2 Marks	L1	CO4
10.	Explain the advantage of using SVM for classification.	2 Marks	L2	CO4

Part B

Answer the Questions. **Total Marks 80M**

11.	**a.**	Discuss different data types and data structures in R with examples.	**20 Marks**	**L1**	**CO1**
	b.	Explain the role of R in Data Science. Discuss its advantages and limitations compared to Python.		**L2**	**CO1**
		Or			
12.	**a.**	Describe the process of working with directories in R. Illustrate with commands to set and get the working directory.	**20 Marks**	**L1**	**CO1**
	b.	Explain data preprocessing and the concept of data transformation using dplyr. Write R code examples to demonstrate the use of select(), filter(), arrange(), mutate(), and summarise().		**L2**	**CO1**

13.	**a.**	Discuss the assumptions of Linear Regression Model and explain why they are important.	**20 Marks**	**L2**	**CO2**
	b.	Explain Non-Linear Regression. Give examples of situations where Non-Linear Regression is more suitable than Linear Regression.		**L2**	**CO2**
		OR			
14.	**a.**	Discuss techniques for exploring a new dataset in R. Include summary statistics and visualization examples. Explain methods to identify and handle anomalies in numerical data with R examples and demonstrate how to visualize relationships between numerical and categorical variables in R.	**20 Marks**	**L2**	**CO2**

15.	**a.**	Find the Multiple linear regression equation for the following dataset. Also find the y for x1=47 x2=25	**20 Marks**	**L4**	**CO3**

Y	X1	X2
2	45	20
1	38	30
3	50	30
2	48	28
3	55	30

		Or			
16.	**a.**	Explain the concept of Regression Analysis. Discuss its importance and applications in Data Science. Derive the equation for Simple Linear Regression and explain the usage of Loss Function.	**20 Marks**	**L4**	**CO3**

<table>
<tr><td>17.</td><td>a.</td><td>Describe the Naive Bayes Machine learning model. Derive Bayes theorem from conditional probability. Use the data set to explain the prediction.

Problem: Players will play if the weather is sunny. Is this statement correct?

<table><tr><th>Weather</th><th>Play</th></tr><tr><td>Sunny</td><td>No</td></tr><tr><td>Overcast</td><td>Yes</td></tr><tr><td>Rainy</td><td>Yes</td></tr><tr><td>Sunny</td><td>Yes</td></tr><tr><td>Sunny</td><td>Yes</td></tr><tr><td>Overcast</td><td>Yes</td></tr><tr><td>Rainy</td><td>No</td></tr><tr><td>Rainy</td><td>No</td></tr><tr><td>Sunny</td><td>Yes</td></tr><tr><td>Rainy</td><td>Yes</td></tr><tr><td>Sunny</td><td>No</td></tr><tr><td>Overcast</td><td>Yes</td></tr><tr><td>Overcast</td><td>Yes</td></tr><tr><td>Rainy</td><td>No</td></tr></table></td><td>20 Marks</td><td>L3</td><td>CO4</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>18.</td><td>a.</td><td>Explain k-fold cross-validation with step-by-step workflow. Compare it with a simple train–test split. Discuss the choice of k, bias-variance trade-off, and advantages of cross-validation in regression.</td><td>20 Marks</td><td>L3</td><td>CO4</td></tr>
</table>