



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

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 09-10-2025

Time: 09.30am to 11.00am

School: SOCSE&SOIS	Program: B.Tech(CSI)	
Course Code : CSD2007	Course Name: R Programming for Data Science	
Semester: V	Max Marks: 50	Weightage: 25%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	26	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.

5Q x 2M=10M

1	Differentiate between Data Analysis and Data Analytics with suitable examples.	2 Marks	L2	CO1
2	Write the syntax of a basic ggplot() call.	2 Marks	L1	CO1
3	Differentiate between data.frame and matrix in R.	2 Marks	L2	CO1
4	Define the meaning of “validating linear assumptions” in regression analysis.	2 Marks	L1	CO2
5	State the importance of covariation in EDA.	2 Marks	L1	CO2

Part B

Answer the Questions.

Total Marks 40M

6.	a. Explain Data Science and its life cycle with example.	10 Marks	L2	CO1
	b. Explain any four types of plots with suitable examples that can be created using the Base R package for data visualization. Additionally, illustrate one example of data visualization using the ggplot2 package in R.	10 Marks	L2	CO1

Or

7.	a. Explain the significance of Data Transformation in data processing.	10 Marks	L2	CO1
	b. Illustrate the Binary encoding and one hot encoding data transformation techniques with example	10 Marks	L2	CO1

8.	a. What are missing values in datasets? Explain different strategies to handle missing values in R.	10 Marks	L2	CO2
	b Write short notes on (a) outliers, (b) skewness, and (c) Standard deviation, d) Covariance with R illustrations.	10 Marks	L2	CO2
	Or			
9.	a. Design an end-to-end EDA workflow in R for a given dataset, including loading, cleaning, handling missing values, visualization, and model assumptions.	20 Marks	L2	CO2