Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 08/10/2026 **Time:** 11:45 AM - 01:15 PM

Course Code: CSD2007 **Course Name:** R programming for Data Science

Semester: Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	88	92

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	List and illustrate the arithmetic operators in R Programming with suitable examples.	2	CO1	BT2
2	Explain the importance of data visualization in R Programming. Name the ggplot2 library and state one type of graph that can be created using it.	2	CO1	BT2
3	State any four advantages of Exploratory Data Analysis in R.	2	CO2	BT2
4	Explain missing values in R and list any two methods for handling them.	2	CO2	BT2
5	Explain the significance of identifying anomalies in R Programming.	2	CO2	BT2

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	Explain the concept of vectors in R programming. Discuss the different types of vectors, methods of creating vectors, accessing vector elements, modifying vector elements, and common operations/functions performed on vectors with suitable examples.	10	CO1	BT2
7	Explain the concept of lists in R programming. Discuss the characteristics of lists, methods of creating lists, accessing and modifying list elements, adding and removing elements, and nested lists with suitable examples.	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	An airline company collects large amounts of data related to flight schedules, passenger numbers, ticket prices, fuel consumption, departure delays, arrival delays, and flight cancellations. The airline wants to analyze this data to understand its past operational performance, identify the major causes of flight delays, predict future passenger demand, and determine suitable strategies for improving operational efficiency. As a data analyst, explain how data analysis can be applied to this scenario. Discuss the four types of data analysis, descriptive, diagnostic, predictive, and prescriptive and explain the major steps involved in the data analysis process	10	CO1	BT2

9	An electricity company records the daily electricity consumption of 4 households for 5 days. The readings are stored in an R matrix. Write an R program using a matrix to: 1. Create the matrix with suitable row and column names. 2. Display the complete matrix. 3. Display the consumption values of the second household. 4. Display the consumption recorded on the third day. 5. Calculate the total consumption of each household. 6. Calculate the average consumption of each household. 7. Identify the highest consumption value in the matrix. 8. Calculate the overall average electricity consumption.	10	CO1	BT3

PART D: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
10	A university has student data containing Student ID, Department, Attendance, Internal Marks, Semester Marks, and Study Hours. The dataset contains some missing values due to incomplete data collection or entry errors. As a data analyst, explain: 1. What are missing values in R and why is handling them important? 2. How can missing values be detected and counted using is.na(), sum(is.na()), colSums(is.na()), and which()? 3. Explain suitable methods for handling missing values using: o na.omit() o Mean/median replacement o Predefined constant replacement	10	CO2	BT2
11	A tourism company has visitor data containing Visitor ID, Destination, Travel Duration, Number of Activities, Travel Expenditure, and Season. The company wants to understand visitor behavior before predicting travel expenditure. As a data analyst, explain how EDA can be used to: 1. Inspect and understand the structure and quality of the dataset. 2. Calculate descriptive statistics to summarize visitor data. 3. Identify missing values and anomalies. 4. Use visualizations to understand data distributions. 5. Analyze relationships and patterns among travel duration, activities, expenditure, and season. Explain how EDA helps prepare the dataset for predictive analysis.	10	CO2	BT2

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	A university has collected student academic data containing Student ID, Attendance Percentage, Study Hours, Internal Marks, and Semester Marks. During Exploratory Data Analysis, most students have semester marks between 55 and 90, while one student has an unusually high value of 198. As a data analyst, explain how anomalies in the student dataset can be detected using summary statistics, box plots, and scatter plots. Discuss how each method helps identify unusual observations and explain the importance of detecting such anomalies before performing further data analysis or developing a predictive model.	10	CO2	BT3
13	An agricultural organization has collected farm data containing Farm ID, Crop Type, Rainfall, Temperature, Fertilizer Usage, and Crop Yield. During EDA, the analyst finds that some farms have missing rainfall and fertilizer usage values because of incomplete data collection and sensor failures. As a data analyst, explain the concept of missing values in R and discuss why handling missing values is important before performing data analysis or developing predictive models. Explain how missing values can be detected, counted, and located using functions such as is.na(), sum(is.na()), colSums(is.na()), and which(). Also discuss suitable approaches for handling missing values, including na.omit(), mean or median replacement, and replacement with a predefined constant. (10 Marks)	10	CO2	BT3