Roll No:
Date: __/__/____

ST. CLARET COLLEGE LIBRARY Bengaluru-13 AUTONOMOUS

St. Claret College

Autonomous, Bengaluru

UG END SEMESTER EXAMINATION-MAY 2026
BCA II SEMESTER
CA 2525: STATISTICAL COMPUTING AND R PROGRAMMING

TIME: 3 hours. 13 **MAX. MARKS: 80**

This paper contains THREE printed pages and FOUR parts

Instructions:
1. **Verify and ensure that the question paper is completely printed.**
2. **Any discrepancies or questions about the exam paper must be reported to the COE within 1 hour after the examination.**
3. **Students must check the course title and course code before answering the questions.**

PART-A

Answer ALL questions. Each answer carries ONE mark. **[10 x 1 = 10]**

1. Which of the following is NOT a valid R data type?
   a) numeric b) integer c) float d) character
2. Which function is used to install a package in R?
   a) library() b) require() c) install.packages() d) load()
3. Which loop in R executes at least once regardless of condition?
   a) for b) while c) repeat d) if-else
4. Which keyword is used to exit the current loop in R?
   a) stop b) exit c) break d) next
5. Bernoulli distribution is characterized by:
   a) Multiple outcomes b) Continuous values c) Two possible outcomes d) Infinite trials
6. Which of the following is a discrete probability distribution?
   a) Normal b) Normal c) Binomial d) Exponential
7. In a normal distribution, the mean, median, and mode are:
   a) Different b) Equal c) Random d) Zero
8. Linear regression is used to:
   a) Classify data b) Find linear relationship between input and output variables
   c) Find probability d) Group data
9. ANOVA is primarily used to:
   a) Compare two means b) Compare more than two means
   c) Find variance only d) Predict outcomes
10. A p-value is used to:
    a) Measure data size b) Accept null hypothesis always
    c) Test statistical significance d) Calculate mean

PART-B

Answer any FIVE questions. Each answer carries TWO marks. **[5 x 2 = 10]**

11. Explain the implicit coercion in R programming.
12. Write a R program to print whether a given number is positive or negative using if-else if-else statement
13. What is a user-defined function? Give an example.
14. Define Binomial distribution. State its assumptions.
15. Explain the properties of a normal distribution.
16. Define hypothesis testing.
17. What is simple linear regression?

PART-C

Answer any FOUR questions. Each answer carries FIVE marks. **[4 x 5 = 20]**

18. Explain R data structures (vector, matrix, list, and data frame) in detail. Write R code to create each structure and demonstrate one operation on each.
19. Write a program to find the factorial of a number using while and repeat loops.
20. A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn are blue?
21. Explain different data visualization techniques in R. Write R code to create a histogram and a boxplot for a given dataset.
22. Explain sampling distribution of the mean. State and explain the Central Limit Theorem with an example.
23. A college claims that the average score of students in a test is 60. A teacher suspects that the average marks are different from 60 based on a sample. Explain how hypothesis testing can be used in this situation. Clearly state the hypotheses, steps involved, and how the decision is made using the p-value (assume the significance level = 0.05).

PART-D

Answer any FOUR questions. Each answer carries TEN marks. **[4 x 10 = 40]**

24. Write an R program to perform the following operations and write the output:

Car	mpg	cyl	gear
A	22	4	4
B	18	6	4
C	25	4	5
D	16	8	3
E	30	4	5

(a) Create a data frame using the above data [2]
(b) Select only mpg, and gear columns [2]
(c) Filter cars with mileage (mpg) greater than 20 [2]
(d) Find the number of cars per cylinders (cyl) [4]

25. Write R programs for

a) To print the day of the week based on numeric input (1–7). [5]

b) A user defined function that accepts a data frame and returns the count of missing values in each column [5]

26. Explain measures of dispersion(range, IQR, variation and standard deviation) in detail.
27. Compare Binomial and Poisson distributions.
28. Consider the following data frame:

```
data <- data.frame(Hours_Studied = c(2, 4, 6, 8, 10),
                   Attendance = c(60, 70, 80, 90, 95),
                   Marks = c(50, 55, 65, 75, 85)
)
```

a) Identify the independent and dependent variables. [2]
b) Write an R program to: [5]
   i) Fit a multiple linear regression model to predict the dependent variable
   ii) Display the summary of the model and the coefficients.
c) If the intercept =10, coefficient of Hours_Studied = 3 and coefficient of Attendance = 0.5, construct the regression equation and write your interpretations. [3]

29. Consider the dataframe given below:

```
data <- data.frame( Product = c("A","B","C","D","E"),
                    Sales = c(120, 150, 100, 180, 130),
                    Profit = c(30, 50, 20, 60, 40))
```

Write an R program to create and customize plots by adding title, axis labels, and colors:
a) Create a bar plot for Sales of each product [4]
b) Create a scatter plot to show the relationship between Sales and Profit [4]
c) Comment on the relationship between Sales and Profit [2]