QP Code: SN23BTECH229 Reg. No | | | | | | | | | | | AY 23

Gandhi Institute of Engineering and Technology University, Odisha, Gunupur (GIET University)

B. Tech (Fifth Semester - Regular) Examinations, November 2025

23BCDPE35011– R for Data Science

(CSE-DS)

Time: 3 hrs Maximum: 60 Marks

Answer ALL questions

(The figures in the right hand margin indicate marks)

PART – A **(2 x 5 = 10 Marks)**

Q.1. Answer ***ALL*** questions

		CO #	Blooms Level
a.	What types of loops exist in R, and what is the syntax of each type?	CO2	K2
b.	Differentiate between print() and cat() functions?	CO2	K2
c.	Recall the usage of factor with a suitable example?	CO2	K1
d.	List the layers of the Grammar of Graphics used in ggplot2 in R?	CO3	K1
e.	What is the use of the tidyverse package?	CO3	K2

PART – B **(10 x 5 = 50 Marks)**

<u>Answer ***ALL*** the questions</u>

		Marks	CO #	Blooms Level
2. a.	Discuss the various miscellaneous operators used in R with example?	5	CO1	K2
b.	Write an R program to compute the result of two numbers based on an operator (+, -, , /) *using the* switch() *function.*	5	CO2	K3
	(OR)			
c.	Explain the usage of list, vector, and factor in R. Write appropriate R code examples to demonstrate the creation and usage of each data structure.	5	CO2	K2
d.	Define data frame? Write the R codes to create the below data frame?	5	CO2	K3

Id	Name	Age	Gen	Marks
E001	John	22	M	28
E002	Smith	22	M	25
E003	Stephen	23	M	26
E004	Queen	22	F	29
E005	Albert	23	M	23

Write R code to:

a) List all the records whose age > 22.
b) Extract the id of all students.
c) Find the average age of the students.
d) Add a new row with your own data

Add a new column Result with values "Pass" if Marks ≥ 80, otherwise "Fail".

		Marks	CO #	Blooms Level
3.a.	Demonstrate an R code to check whether a given number is prime using recursion.	5	CO2	K3
b.	Explain subsetting in R. Write R code examples to subset elements from a vector, matrix, and data frame.	5	CO2	K3
	(OR)			
c.	Write the steps to connect R with an SQLite database using the RSQLite package. Write the codes for the following: a) Create two tables called Employee (Empid, Ename, Esal, Edept) and Department (Did, Dname, Dstartdate) b) Insert 3 records into Employee and Department table. c) Perform leftouter, rightouter and fullouter join on both Employee and Department table.	5	CO3	K4
d.	Compile basic syntax for creating a histogram using R on data of own choice.	5	CO3	K3
4.a.	Explain how to import and export JSON and text files in R with suitable code examples.	5	CO3	K3
b.	Explain various types of charts and graphs in R using the ggplot2 package. Describe the purpose of at least three types of plots and write suitable R code examples using ggplot2 to illustrate each.	5	CO3	K2

(OR)

c. What is a Web API HTTP interface in R? Explain how HTTP methods (GET, POST) are used to interact with web APIs using suitable R packages and examples. 5 CO4 K4

d. Define a Box Plot and explain its key components. Using the dataset below, create a Box Plot in R to visualize the distribution of marks in three subjects (Math, Science, and English). 5 CO3 K3

Student	Math	Science	English
1	78	85	72
2	88	79	90
3	92	83	88
4	70	75	65
5	80	82	78

5.a. Define simple linear regression. From the data below, fit the least-squares regression line of Y on X, Compute the Standard Error of Estimate, and predict Y when $X = 15$. 5 CO4 K4

Experience (Years) (X)	3	5	7	9	12
Salary (₹ in thousands) (Y)	40	55	65	70	90

b. A candy manufacturer wants to check if preference for chocolate flavors is independent of gender among teenagers. A sample of 120 teenagers is surveyed, and the results are as follows: 5 CO4 K4

Gender	Vanilla	Strawberry	Chocolate	Mint	Total
Male	12	15	18	5	50
Female	8	22	25	15	70
Total	20	37	43	20	120

Using a chi-square test for independence at the 5% significance level, and χ^2_tab = 7.815 (for 3 degrees of freedom), test whether flavor preference is independent of gender.

(OR)

c. Define multiple linear regression and explain its purpose in data analysis. Given the following dataset showing hours of study (X_1), attendance percentage (X_2), and exam scores (Y) for five students, develop a multiple linear regression equation to predict exam scores.Also, estimate the predicted exam score for a student who studies 9 hours and has 90% attendance. 5 CO2 K2

Student	Study Hours (X_1)	Attendance (%) (X_2)	Exam Score (Y)
1	2	60	50
2	3	65	55
3	5	70	65
4	6	80	75
5	8	85	85

d. Define Correlation Analysis. Using the dataset below, calculate Karl Pearson's Coefficient of Correlation between Online Learning Hours (X), Social Media Hours (Z), and Exam Scores (Y), and visualize the correlation using a Heatmap in R. 5 CO5 K5

ID	X	Y	Z
1	2	8	40
2	3	7	50
3	4	6	60
4	5	5	70
5	6	4	80

6.a List out the commonly used functions in dplyr package. 5 CO2 K2

b. Explain boxplot and Describe how boxplots are used for outlier detection. Given the following dataset representing the ages of employees in a company: Data (Ages): 21, 22, 25, 25, 27, 28, 29, 30, 32, 34, 36, 38, 40, 58, 64 Construct a boxplot for the given data and write R code to visualize it. 5 CO3 K3

(OR)

c. Differentiate between Normal distribution and Binomial Distribution? 5 CO2 K2

d. Define Time Series Analysis and explain its significance in business forecasting. Using a 3-month Weighted Moving Average (WMA), forecast the revenue for July if the weights for the most recent three months are: 5 CO3 K3

June = 0.5, May = 0.3, April = 0.2

--- End of Paper ---