Code No.: CS512PE | R20 | H.T.No. | | | 8 | R | | | | | |

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Regular) - January- 2024
DATA ANALYTICS USING R
(CSE)

[Time: 3 Hours] **[Max. Marks: 70]**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 20 marks. Answer all questions in Part A.
Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks.

PART-A **(20 Marks)**

1. a)	What is R and what is it used for?	[2M]
b)	Explain the purpose of the 'ls()' command in R.	[2M]
c)	What is the definition of percentile in descriptive statistics?	[2M]
d)	Why is the median often preferred over the mean in the presence of outliers?	[2M]
e)	What is the fundamental goal of linear regression?	[2M]
f)	What is the purpose of the 'lm' function in R?	[2M]
g)	What is the primary purpose of logistic regression in statistics?	[2M]
h)	Name the R function used to fit a logistic regression model.	[2M]
i)	Name the components of a decision tree representation.	[2M]
j)	Explain the R function used to create a decision tree?	[2M]

PART-B **(50 Marks)**

2.a)	Describe the role of a directory in R and why it is important.	[5M]
b)	Use the tapply() function to calculate the mean of a variable for each group in a data set.	[5M]
	OR	
3.a)	Write a command and explain the procedure to load a CSV file named "data.csv" into R.	[5M]
b)	Can you describe the role of the 'as' operator in changing the structure of data in R?	[5M]
4.a)	Examine the impact of outliers on measures like mean and median in a dataset.	[5M]
b)	Explain the concept of skewness and how it influences the shape of a distribution.	[5M]
	OR	
5.a)	How do percentiles differ from quartiles in terms of data distribution representation?	[5M]
b)	Explain exploratory data analysis (EDA) with example.	[5M]
6.	Apply the residuals() function in R to obtain the residuals of a linear regression model.	[10M]
	OR	
7.	Explain the difference between simple linear regression and multiple linear regression.	[10M]
8.a)	Explain the concept of odds ratio in the context of logistic regression.	[5M]
b)	Apply the glm() function in R to fit a logistic regression model.	[5M]
	OR	
9.	Analyze the difference between binary logistic regression and multinomial logistic regression.	[10M]

10.a)	Describe the role of split criteria in decision tree learning.	[5M]
b)	Apply the concept of entropy to measure uncertainty in a decision tree.	[5M]

OR

11.a)	Why is decision tree learning considered a top-down, greedy algorithm?	[5M]
b)	Apply pruning techniques to avoid overfitting in decision tree models.	[5M]

****