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

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Supply) - May- 2023
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)	Identify the various data types in R.	[2M]
b)	Compare and contrast the various R GUIs for data input.	[2M]
c)	Difference between variance and standard deviation.	[2M]
d)	Define measures of central tendency.	[2M]
e)	Define linear regression in statistics.	[2M]
f)	Identify the assumptions of linear regression and explain why they are important.	[2M]
g)	Define binary logistic regression and its application.	[2M]
h)	Analyze the goodness of fit for a binary logistic regression model.	[2M]
i)	How to Identify the appropriate problems for decision tree learning.	[2M]
j)	Analyze the representation of decision tree in R.	[2M]

PART-B **(50 Marks)**

2.a)	Evaluate the effectiveness of the 'as' operator in changing the structure of data in R.	[7M]
b)	How do you install and load packages in R?	[3M]
	OR	
3.	Using R, create a matrix and perform basic arithmetic operations on it.	[10M]
4.	Compare and contrast the use of Mean, Median, and Mode as measures of central tendency in different scenarios.	[10M]
	OR	
5.a)	Evaluate the effectiveness of using Histograms versus Box plots for displaying data in statistics.	[5M]
b)	What is the difference between skewed and symmetric data?	[5M]
6.	Evaluate the effectiveness of different methods for model validation in linear regression.	[10M]
	OR	
7.a)	Explain the difference between correlation and linear regression.	[5M]
b)	Explain the process of fitting a linear regression model in R.	[5M]
8.	Describe the role of maximum likelihood estimation in Logistic Regression and how it is used to fit the model.	[10M]
	OR	
9.	Evaluate the effectiveness of using multinomial Logistic Regression models for predicting multiple outcomes.	[10M]
10.a)	Analyze the impact of data preprocessing on Decision Tree learning.	[7M]
b)	What is inductive biasing in decision tree learning?	[3M]
	OR	
11	Analyze the advantages and disadvantages of using different measures of impurity, such as Entropy and Gini index, in Decision Tree learning.	[10M]

****