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

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Supply) - December- 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 the 'as' operator in R?	[2M]
b)	Analyze the challenges of data preprocessing and suggest possible solutions.	[2M]
c)	List two measures of central tendency commonly used in statistics.	[2M]
d)	Analyze the skewness and kurtosis of a histogram.	[2M]
e)	Identify a method for checking the linearity assumption in linear regression.	[2M]
f)	Define the term "model fitting" in linear regression.	[2M]
g)	Calculate the odds ratio for a binary logistic regression model.	[2M]
h)	Define binary logistic regression.	[2M]
i)	Analyze the importance of feature measurement in decision tree learning.	[2M]
j)	Define the concept of hypothesis space in decision tree learning.	[2M]

PART-B **(50 Marks)**

2.a.	Construct an R code to load a dataset and handle the missing values in it.	[7M]
b.	What are some commonly used methods for reading data into R?	[3M]
	OR	
3.	Explain the process of handling datasets in R .	[10M]
4.	Evaluate the effectiveness of using percentiles and quartiles in analyzing data.	[10M]
	OR	
5.a.	Develop a comprehensive strategy for identifying and handling outliers in a large data set.	[6M]
b.	How can histograms be used to display data in a meaningful way?	[4M]
6.	Explain the difference between the dependent and independent variables in a linear regression model.	[10M]
	OR	
7.a.	Compare and contrast linear regression models with other types of regression models.	[7M]\
b.	What is residual analysis in linear regression?	[3M]
8.a.	Explain the basic concepts of logistic regression and its applications in statistics.	[5M]
b.	How do you diagnose logistic regression models for potential issues?	[5M]
	OR	
9.	Develop a strategy for handling multicollinearity in a logistic regression model.	[10M]
10.	Evaluate the effectiveness of pruning techniques for improving decision tree models.	[10M]
	OR	
11.a.	Analyze the impact of overfitting on decision tree models.	[5M]
b.	Why do we prefer short hypotheses in decision tree learning?	[5M]

****