Code No.: (R22DS502PC) | R22 | H.T.No. | | | 8 | R | | | | | |

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Regular) - December- 2024
INTRODUCTION TO DATA SCIENCE
(CSD)

[Time: 3 Hours] **[Max. Marks: 60]**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 10 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 **(10 Marks)**

1. a)	What is the purpose of fitting a model?	[1M]
b)	What is overfitting?	[1M]
c)	Define a discrete attribute.	[1M]
d)	What does the interquartile range (IQR) represent?	[1M]
e)	How is an array different from a matrix in R?	[1M]
f)	What function is used to merge two lists in R?	[1M]
g)	Which statement is used for conditional execution in R?	[1M]
h)	Which keyword is used to define a function in R?	[1M]
i)	What is parametric data reduction?	[1M]
j)	What type of data is best suited for icon-based visualization?	[1M]

PART-B **(50 Marks)**

2.	Explain the term "Datafication." How data has to transform in industries and everyday life?	[10M]
	OR	
3.	What are the basic data types in R? Provide examples for each data type.	[10M]
4.	How do we measure the variance and standard deviation of a dataset? Calculate these for the following data: [3, 7, 7, 10, 15].	[10M]
	OR	
5.	Discuss the importance of graphic displays in representing basic statistical descriptions.	[10M]
6.	What is vector subsetting? Explain the different types of subsetting with appropriate examples	[10M]
	OR	
7.	Describe the steps to create and name a list in R. Provide an example of accessing and manipulating list elements	[10M]
8.	Explain how relational operators work with vectors in R. Illustrate with a code example that compares two vectors element-wise.	[10M]
	OR	
9.	Write an R program that uses a for loop to calculate the factorial of a given number.	[10M]
10.	Describe the concept of wavelet transforms and how they are used for data reduction. Provide a simple example.	[10M]
	OR	
11.	How does clustering help in reducing data complexity? Compare it with sampling as a data reduction strategy.	[10M]

****