

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI  
(END SEMESTER EXAMINATION)**

CLASS: MBA  
BRANCH: MBA

SEMESTER : 3  
SESSION : MO/2025

**SUBJECT: MT551 DATA SCIENCE USING R**

TIME: 3 Hours

FULL MARKS: 50

**INSTRUCTIONS:**

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

Q.1 Following data is stored in a file by the name -- data.csv, [10]    CO 3    BL 4

Marketing Expense (Rs'000)	Sales (Rs'000)
50	11
50	
55	14
60	16
65	16
56	12
63	16
	14
50	13
60	13

Using R, write a comprehensive script to (i) impute the missing values. (ii) also, extend the script to estimate the Karl Pearson correlation coefficient without using directly the cor() function.

Q.2 Data is stored in a file by the name -- data.csv. The table below just represents how the data is stored in the file. 10    4    5

X	Y
5	15
7	18
11	38
8	25
10	31

Without directly using the lm() function, write a script in R to load the data from the .csv file and then estimate the intercept and slope of the simple linear regression  $Y = a + bX$ .

Q.3(a) Explain the importance of Data Science in Business Decision Making with suitable example [5]    1    3

Q.3(b) In the context of R, write notes on, [5]    2    3  
 (1) Functions  
 (2) Data Types and Objects

Q.4 In the context of Decision Tree Explain your understanding about, [10]    5    3  
 (i) Gini Impurity  
 (ii) Entropy  
 (iii) Overfitting  
 (iv) Useful Packages in R

Q.5(a) Explain the Plot functionalities available in R with suitable examples [5]    2    3

Q.5(b) Explain the essential steps in predictive modelling using a suitable real-life example. Use a flow diagram to illustrate. [5]    2    4