

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

**CLASS: BBA
BRANCH: BBA**

**SEMESTER :VI
SESSION : SP/2025**

SUBJECT: MT334 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(a) Write an R program to find whether a number is prime using a loop. | [5] | CO1 | BL4 |
| Q.1(b) How do you create a factor in R? What is it used for? Explain with suitable example. | [5] | CO3 | BL2 |
| | | | |
| Q.2(a) How does skewness affect the mean and median of a dataset? Differentiate between mesokurtic, leptokurtic, and platykurtic distributions. | [5] | CO2 | BL3 |
| Q.2(b) Why are Contingency Tables Important in Data Analysis? What is the difference between a summary table and a contingency table? | [5] | CO2 | BL1 |
| | | | |
| Q.3(a) Explain the difference between a histogram and a bar plot in terms of data visualization. | [5] | CO1 | BL2 |
| Q.3(b) How does a box plot help in identifying outliers in the data? Explain with suitable example. | [5] | CO3 | BL2 |
| | | | |
| Q.4(a) Write R code to filter and summarize data using multiple conditions (e.g., age > 30 & gender == "Male"). | [5] | CO4 | BL5 |
| Q.4(b) How would you handle missing data (NA values) when using apply(), lapply(), and tapply() functions? | [5] | CO4 | BL4 |
| | | | |
| Q.5(a) Entropy is an important concept in information theory. Explain its significance in mining context. | [5] | CO4 | BL3 |
| Q.5(b) What is the difference between "supervised" and unsupervised" learning scheme. State and explain Bayes Theorem. | [5] | CO1 | BL1 |

:02/05/2025:M