

USN

Course Code M B A B A 4 1 1

Fourth Semester MBA Degree Examinations, August 2026

R PROGRAMMING FOR MANAGERS

Duration: 3 hrs

Max. Marks: 100

- Note:**
1. Answer any FOUR full questions from Question No. 1 to 7.
 2. Question No. 8 is compulsory
 3. Missing data, if any, may be suitably assumed

** ** *

<u>Q. No</u>	<u>Question</u>	<u>Marks</u>	<u>(RBTl:CO:PO)</u>
1	a. Explain the impact of replication factors on fault tolerance in HDFS. b. Demonstrate the use of generic HDFS commands for listing files, creating directories, uploading files to HDFS, and viewing file contents. c. Explain the architecture of the Hadoop Distributed File System and how MapReduce programming integrates with it.	03 07 10	(2:1:2) (5:1:2) (4:1:1)
2	a. Describe what happens when HDFS enters "Safe Mode." b. Evaluate the role of fault tolerance mechanisms and speculative execution in improving the reliability and performance of Hadoop Framework. c. Design a logical MapReduce pipeline to count the frequencies of words in a below given dataset. Detail the role of the Input Split, Mapper, Combiner, Shuffle, and Reducer phases in your design. Dataset (input): COMPUTER MBA CSE CSE STUDENT MBA MBA STUDENT BITM	03 07 10	(2:1:1) (5:1:2) (6:1:2)
3	a. Outline how the key features of R programming can be used for data analysis with examples. b. Apply different operators in R programming and explain their usage with examples. c. Analyze the effectiveness of vectors and built-in functions in R for data manipulation and computational tasks with examples.	03 07 10	(2:2:2) (3:2:1) (4:2:2)
4	a. Explain the concept of factors in R and examine their significance in organizing and analyzing categorical data within data analysis applications. b. Evaluate the data frame in R to store student information and demonstrate the methods used to access specific rows and columns for data manipulation and analysis.	03 07	(2:3:1) (5:3:2)

Student ID	Name	Age	Marks	Branch
101	Rahul	20	85	MBA
102	Priya	21	90	MBA
103	Arjun	22	78	MBA
104	Sneha	20	88	MBA
105	Kiran	21	75	MBA

- c. Write an R script using the vector-based `ifelse()` function to classify these grades. The code should return "Pass" if the grade is 70 or higher and "Fail" if it is below 70. **10** **(6:3:2)**
 Given a list of student grades `c(85, 92, 78, 60, 75)`
- 5 a. Describe the exact matching and partial matching of arguments in R functions by illustrating their behaviour with suitable coding examples. **03** **(2:3:2)**
 b. Evaluate the structural differences among `for`, `while`, and repeat loops in R, and demonstrate the implementation of a repeat loop that terminates using a conditional break statement. **07** **(5:3:2)**
 c. Write an R program to create the following data frame. **10** **(6:3:2)**
 (i) Write the contents of data frame from Data frame to CSV/textfile.
 (ii) Read the contents from CSV/textfile to a new dataframe and display the same.

Dataset: Product Sales Information

Product_ID	Product_Name	Units_Sold	Selling_Price (₹)	Total_Sales (₹)
P101	Rice	50	320	16000
P102	Cooking Oil	35	180	6300
P103	Sugar	60	45	2700
P104	Wheat Flour	40	250	10000
P105	Milk	80	60	4800
P106	Tea Powder	25	220	5500
P107	Biscuits	90	30	2700
P108	Soap	70	40	2800
P109	Shampoo	45	150	6750
P110	Detergent	55	120	6600

6. a. Use of the `plot()` function in R? **03** **(2:4:1)**
 b. Construct an R program to calculate mean, median, and standard deviation for a dataset. **07** **(6:4:2)**

Student_ID	Marks
1	45
2	56
3	62
4	70
5	75
6	80
7	68
8	72
9	85
10	90

- c. Analyze the dataset using R statistical functions and graphical techniques to identify trends and generate insights. **10** **(4:4:4)**

Patient_ID	Age	Gender	Blood_Pressure	Blood_Sugar	BMI	Cholesterol
P101	25	Male	120	90	22.5	180
P102	40	Female	135	110	27.8	210
P103	35	Male	128	95	24.6	190
P104	50	Female	145	130	30.2	240
P105	45	Male	138	115	28.5	220
P106	30	Female	122	88	23.4	175
P107	55	Male	150	140	31.5	250
P108	38	Female	130	100	26.0	200
P109	60	Male	155	145	32.8	260
P110	42	Female	140	118	29.1	225

7. a. Explain the concept of interfacing R to other languages. 03 (2:4:1)
- b. Compare Generalized Linear models and non-linear models in R. 07 (4:4:2)
- c. Elucidate how Parallel R is utilized alongside clustering and time series analysis. 10 (5:4:1)

8. Case Study

Retail Store Sales Analysis Using R Programming

Scenario

A retail supermarket wants to analyze its weekly sales performance using R programming. The sales manager has collected data on products, units sold, selling price, and total sales. As a data analyst, you need to use R programming concepts such as data frames, vectors, statistical functions, indexing, and conditional statements to answer the following questions.

Dataset

Product ID	Product Name	Units Sold	Selling Price (₹)	Total Sales (₹)
P101	Rice	50	320	16000
P102	Cooking Oil	35	180	6300
P103	Sugar	60	45	2700
P104	Wheat Flour	40	250	10000
P105	Milk	80	60	4800
P106	Tea Powder	25	220	5500
P107	Biscuits	90	30	2700
P108	Soap	70	40	2800
P109	Shampoo	45	150	6750
P110	Detergent	55	120	6600

- a. Construct an R program to: 04 (3:5:1)
 - (i) Create the above dataset as a data frame.
 - (ii) Display the first five rows of the dataset.
- b. Evaluate and write an R program to: 06 (5:5:4)
 - (i) Use an **if-else statement** to classify products as:
 - **High Demand** if Units Sold ≥ 60
 - **Low Demand** otherwise
 - (ii) Display the Product Name and Demand Category.
- c. Design a R program to: 10 (6:5:2)
 - (i) Display products with Units Sold greater than 50.
 - (ii) Find the product with the highest Total Sales.
 - (iii) Add a new column named Profit, assuming Profit = 20% of Total Sales.
 - (iv) Display the updated dataset.