

**Code No: 156FD****JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD****B. Tech III Year II Semester Examinations, August/September - 2024****INTRODUCTION TO DATA SCIENCE****(Computer Engineering – Software Engineering)****Time: 3 Hours****Max. Marks: 75**

- Note:** i) Question paper consists of Part A, Part B.  
ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.  
iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

**PART – A****(25 Marks)**

- 1.a) Write an R script that installs the “ggplot2” package if it is not already installed and then loads it into the R environment. [2]
- b) Write an R script to create a new directory named “DataScienceProjects” in the current working directory and confirm its creation. [3]
- c) Define a binary attribute and give an example. [2]
- d) Differentiate between discrete and continuous attributes. [3]
- e) Write an R script to create a factor from a character vector representing the days of the week (with some repetitions) and summarize the factor. [2]
- f) Write an R script to create a named list with components “Name” (character), “Age” (numeric), and “Scores” (numeric vector). Access the “Scores” element and print it. [3]
- g) How do you prevent an infinite loop when using a while loop in R? [2]
- h) Write an R script to create two numeric vectors and use relational operators to create a logical vector that indicates where the elements of the first vector are equal to the corresponding elements of the second vector. [3]
- i) What is the main objective of Principal Components Analysis (PCA)? [2]
- j) Define stratified sampling and explain its purpose. [3]

**PART – B****(50 Marks)**

- 2.a) Write an R script to create a matrix from a given vector and print the dimensions and the data type of the matrix.
- b) Name two common probability distributions used in statistical modeling and describe the use cases. [5+5]

**OR**

- 3.a) List and briefly describe the basic data types in R.
  - b) Write an R script to create a vector of integers, a vector of characters, and a vector of logical values. Then, print the type of each vector. [5+5]
- 4.a) List and briefly describe the four different types of attributes.
  - b) Write an R function “attribute\_cardinality” that takes a vector and returns the number of unique values in the attribute. [5+5]

**OR**

- 5.a) Define mean, median, and mode, and explain when each measure is most appropriate to use.
- b) Write an R function “calculate\_dispersion” that takes a numeric vector and returns the range, variance, and standard deviation. [5+5]

- 6.a) Write an R function that takes two numeric vectors of the same length and returns their element-wise product.
- b) Write an R script to create a numeric vector and then extract all elements that are greater than a specified value. [5+5]

**OR**

- 7.a) Write an R script to create a numeric vector and then extract the elements at even positions.
- b) Write an R script to create a data frame with columns “Name”, “Age”, and “Gender”. Add a new row to the data frame and then subset it to include only rows where Age is greater than 25. [5+5]

- 8.a) Write an R function “logical\_comparison” that takes two logical vectors and returns a new vector that is the element-wise logical AND of the two vectors.
- b) Write an R script to create a numeric vector and use logical operators to identify and print elements that are either less than 10 or greater than 20. [5+5]

**OR**

- 9.a) Write an R function “check\_age” that takes a numeric vector of ages and returns a character vector indicating if each age is “Minor” (less than 18), “Adult” (18 to 64), or “Senior” (65 and above).
- b) Write an R function “nested\_sum” that contains a nested function “sum\_inner” which calculates the sum of two numbers. The outer function should take a numeric vector and use the inner function to return the sum of all elements. [5+5]

- 10.a) Implement a function in R that performs data reduction using wavelet transforms. Your function should take a dataset as input and apply wavelet transforms to reduce the dimensionality of the data.
- b) Create an R script to build a regression model for predicting housing prices based on various attributes such as square footage, number of bedrooms, and location. [5+5]

**OR**

- 11.a) Write an R script to generate histograms for multiple variables in a given dataset.
- b) Develop an R function to perform k-means clustering on a dataset with a specified number of clusters. [5+5]

---ooOoo---