Printed Pages– 12

Roll No. 3000128 2 1095

B127475(022)

B. Tech. (Hon's) (Fourth Semester) Examination, April-May 2023

(New Scheme)

(CS Engg. Branch)

(Artificial Intelligence & Data Science)

R for DATA SCIENCE

Time Allowed : Three hours

Maximum Marks : 100

Minimum Pass Marks : 35

***Note** : Part (a) of each question is compulsory & carries 4 marks. Attempt any **two** parts from (b), (c) and (d) of each question and each part carries 8 marks.*

Unit-I

1. (a) Fill in the blanks :

R is a programming language and software environment for statistical computing and graphics. It was developed by statisticians .....

and ..... in the early 1990s. The initial version of R was based on the ..... programming language developed at .....

(b) Explain the importance and functionality of Help Files in the R language. Provide examples to illustrate their usage.

(c) A research team has conducted a study on the heights (in centimetres) of eight randomly selected individuals. The recorded heights are as follows: 160, 172, 178, 155, 183, 168, 175 and 165 cm. The team aims to analyse the data to gain insights into the height distribution of the individuals. They want to determine the average height, the minimum and maximum heights, and the height range (difference between the highest and lowest heights). Additionally, they wish to create a new vector to store the heights of the first four individuals and find the length of this vector using R.

Using R, perform the following tasks and suggest outputs

1. Create a vector containing the heights of the eight

individuals.

2. Calculate the average height of the eight individuals.
3. Determine the minimum and maximum heights.
4. Calculate the height range.
5. Assign the heights of the first four individuals to a new vector.
6. Extract the length of the new vector.

(d) Discuss the installation process of R programming language and RStudio, emphasizing the key concepts and steps involved. Explain the significance of each installation and their roles in the development and execution of R programs. Provide a step-by-step guide for installing R and R Studio, highlighting any potential challenges that users may encounter during the process. Finally, discuss how to verify the successful installation of both R and RStudio and ensure their proper functioning on the system.

Unit-II

2. (a) What is matrix transposition? Explain how to obtain the transpose of a given matrix and discuss its properties. Provide an example to illustrate the

process in R language

(b) Perform the following operations in R, and provide the resulting matrices

(i) Create a new matrix 'mat1 & mat2' with dimensions 2×3 containing the numbers 1 to 6 and 6 to 12.

(ii) Combine mat1 and mat2 vertically (row-wise) to create a new matrix.

(iii) Combine mat1 and mat2 horizontally (column-wise) to create a new matrix.

(iv) Find the transpose of the matrix 'mat'.

(v) Create a 3×3 identity matrix 'identity_mat'.

(vi) Calculate the element-wise addition of mat1 and mat2.

(vii) Calculate the element-wise subtraction of mat1 and mat2.

(viii) Perform matrix multiplication between mat1 and the transpose of mat2.

(c) Given two 3×3 matrices A and B

A = [1,2,3] B = [9,8,7]

[4,5,6] [4,5,6]

[7,8,9] [3,2,1]

(i) Create a new matrix C by horizontally concatenating A and B.

(ii) Extract the first row and third column of C and calculate their sum.

(iii) Replace the diagonal elements of A with the diagonal elements of the 3×3 identity matrix.

(iv) Calculate the transpose of the modified A.

(v) Perform element-wise addition and subtraction of A and B, and then multiply the resulting matrices.

(vi) Invert the resulting matrix from the previous step, if possible.

(vii) Create a 3×3×2 multidimensional array using A and B as the first and second slices, respectively.

(viii) Extract the element in the first row, second column, and second slice of the multidimensional array.

(d) Given a square matrix of size N × N, you are allowed to perform the following operation on the matrix :

(i) Choose a submatrix within the given matrix defined by the top-left corner (row1, col1) and the bottom-right corner (row2, col2)

(ii) Increase all the elements in the chosen submatrix by 1.

Your task is to find the minimum number of operations required to make all elements in the matrix equaL If it is not possible to make all elements equal, return -1.

Write a function minOperations(matrix) that takes a matrix as input and returns the minimum number of operations required.

Function signature: minOperations(matrix: matrix) > integer

Example

matrix = matrix(c(1, 2, 3, 4, 5, 6, 7, 8,9), nrow = 3, ncol = 3)

minOperations(matrix) => 2

Explanation

In the given example, the initial matrix is:

123

456

789

We can perform the following operations:

Choose submatrix (1, 1) to (2, 2) and increase all elements by 1. The matrix becomes:

233

566

789

Choose submatrix (2, 2) to (3, 3) and increase all elements by 1. The matrix becomes:

233

577

7810

Now, all elements in the matrix are equal to 3. It took 2 operations to achieve this, so the expected output is 2.

Note : You can assume that the input matrix will always be a square matrix of positive integers.

Unit-III

(a) Which of the following statements about characters in the R language are true?

(Select all that apply)

(i) Characters can be defined using either single quotes (' ') or double quotes (" ").

(ii) Special characters such as \n and \t can be included in character strings.

(iii) The length of a character vector can be determined using the len( ) function.

(iv) Characters can be concatenated using the paste( ) function.

(v) Characters can be converted to numeric values using the is.numeric( ) function.

(vi) Characters can be indexed using numeric or logical vectors to extract specific elements.

(b) A fashion magazine has compiled a list of fashion trends for the upcoming season. They want to analyse the trends and extract specific information. The trends on the list are as follows:

(i) "Floral prints are back in style."

(ii) "Embrace oversized sweaters for a cozy look."

(iii) "Animal prints are making a fierce comeback."

Using R, perform the following tasks:

(i) Create strings for each of the trends.

(ii) Concatenate the three trends into one string, separating each trend with a newline character.

(iii) Extract a substring from the second trend (characters 9 to 19).

(iv) Check if the word "prints" is present in each of the trends.

(c) Using R language, perform the following tasks on the given dataset and provide your code and output for each step. The dataset contains information about the favourite fruit of a group of people.

Dataset:

data <- c("apple", "banana", "cherry", "apple", "banana", "apple", "cherry", "banana", "apple", "cherry")

Tasks:

(i) Convert the dataset into a factor and display the levels.

(ii) Identify the unique categories in the dataset.

(iii) Create a new factor with custom levels: "cherry", "banana", "apple", and display the levels.

(iv) Count the frequency of each fruit in the dataset.

(v) Cut the frequency data into two intervals: "Low" for frequencies less than or equal to 3, and "High" for frequencies greater than 3.

Provide the code and output for each step.

(d) Explain, the concepts of defining and ordering levels, as well as combining and cutting data in the R language. Provide a brief explanation for each concept and include examples to illustrate their usage.

Unit-IV

4. (a) Which of the following options correctly describe the special values in R: Infinity, NaN, NA, and NULL? (Select all that apply.)

(i) Infinity: Represents an infinite or extremely large numerical value.

(ii) NaN: Represents a logical value indicating a false condition.

(iii) NA: Indicates missing values or unavailable data.

(iv) NULL: Denotes the absence of a value or an empty object.

(b) Differentiate between Lists and Data Frames in the R language. Discuss their characteristics, usage, and key distinctions. Support your answer with suitable examples.

(c) What are the advantages of using Logical Record Subsets for data manipulation in R?

(d) Compare and contrast the usage and functionality of is-dot object checking functions and as-dot coercion functions in R. Provide examples to illustrate their differences.

Unit-V

5. (a) The code snippet above intends to create a line plot using ggplot2. However, there is an error in the code. Can you identify and correct the error to generate the desired plot?

```
library(ggplot2)
data <- data.frame(x = (1, 2, 3, 4, 5), y = (3, 5, 4, 6, 8))
```

ggplot(data, aes(x = x, y = y)) + geam_line(col = "red") + geam_point (col= "blue", pkh=16) + ggtitle("My Plot") + xlab("X Values") + ylab("Y Values")

(b) Explain the concepts of reading and writing external data files in R, the significance of R -Ready data sets, and the benefits of utilizing contributed data sets in R programming. Provide relevant examples where applicable.

(c) Explain the key aspects of ggplot 2 in R for data visualization, including the basic principles, layers and aesthetics, customization and theming, as well as advanced plots and geometries. Provide examples to demonstrate the practical application of ggplot2 for creating visually appealing and informative plots in R.

(d) Explain the purpose and importance of setting appearance constants with geoms in R for data visualization. Highlight the role of geoms in defining visual elements such as colors, shapes, sizes and line types. Provide examples demonstrating how appearance constants can be applied to geoms in R to enhance the aesthetics and interpretability of plots.