**SAMPLE 1**

1. Develop an R script to simulate and critically analyze basic arithmetic operations by comparing their outputs with mathematical calculations. Evaluate the precision and efficiency of these operations, especially in handling edge cases, and discuss the implications for computational accuracy in R.
2. (i) Analyze the summary statistics of the air quality dataset to uncover patterns, trends, and potential outliers. How do these insights inform environmental or health-related conclusions?

(ii) Transform the airquality dataset into a long format using the melt function. Evaluate the advantages of using long-format data for analysis and visualization in R.

1. Randomly partition the iris dataset into 80% training and 20% testing subsets. Evaluate the implications of this sampling approach for model robustness and representativeness. Develop a logistic regression model using the training data, with Species as the target and Petal. Width and Petal.Length as predictor variables. Critically assess the model's assumptions and fit. Use the test data to predict species probabilities, and interpret the output in terms of the model’s predictive performance and reliability. What insights can be drawn from the predicted probabilities? Construct a confusion matrix for the test data predictions and analyze the model's classification accuracy, precision, recall, and F1 score. Discuss how these metrics guide the evaluation and improvement of the model.
2. Design an R program to create a multidimensional array using specified columns, rows, and tables. Analyze how the structure of the array influences data organization and accessibility. Evaluate the benefits of using arrays for multidimensional data representation in comparison to other R data structures.

**SAMPLE 2**

1. Design and construct a data frame in R with the given structure and data. Critically evaluate the dataset’s composition, including the interplay between numerical and categorical variables. Reflect on how this dataset can be utilized for various analytical purposes, such as student performance analysis or decision-making based on attempts and qualification status. Identify potential improvements or extensions to make the dataset more comprehensive for real-world applications.
2. Design an R program to create a list comprising vectors, matrices, and a custom function. Analyze the structure and behavior of the list, focusing on how the diverse data types interact within the list. Evaluate the practical applications of lists in managing and processing heterogeneous data in R.
3. Explore the USArrests dataset, which contains the number of arrests for murder, assault, and rape for each of the 50 states in 1973. It also contains the percentage of people in the state who live in an urban area.
   1. a. Explore the summary of the Data set, like the number of Features and its type. Find the number of records for each feature. Print the statistical feature of data

b. Print the state which saw the largest total number of rapes

* 1. a. Find the correlation among the features

b. Print the states that have assault arrests more than the

median of the country

1. Write an R program that:

a. Reads the sales\_data.csv file.

b. Displays its contents in the console.

c. Prints the data structure (e.g., column names, data types) to help analyze its format.

## SAMPLE-3

1. Create three vectors in R: one containing numeric data, one containing character data, and one containing logical data. What might be examples of real-world scenarios where each of these data types could be used?
2. Create below data frame exam\_data = data. frame( name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'), score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),

attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),

qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes') )

* 1. Write an R program to extract 3rd and 5th rows with 1st and 3rd columns from a given data frame
  2. Write an R program to add a new column named country in a given data frame Country<- c("USA","USA","USA","USA","UK","USA","USA","India","USA","USA")
  3. Write a R program to add new row(s) to an existing data frame new\_exam\_data = data.frame(name = c('Robert', 'Sophia'),score = c(10.5, 9), attempts = c(1, 3),qualify = c('yes', 'no'))
  4. Write an R program to sort a given data frame by name and score

e. Write a R program to save the information of a data frame in a file and display the information of the file.

1. Design an R program to create and visualize the height factor from the women's dataset, categorizing heights into meaningful intervals. Based on the dataset's distribution, how would you ensure these intervals are appropriate and justified?
2. Design an R program that demonstrates the creation of:
3. A 5 × 4 matrix filled by rows.
4. A 3 × 3 matrix with custom labels for rows and columns, filled by rows.

# SAMPLE-4

1. Write an R program to create a multi-dimensional array in R using a vector of values and dimensions. Analyze the impact of assigning meaningful names to dimensions and discuss how this enhances the readability and usability of the array in real-world applications.
2. Write an R program to call the (built-in) dataset airquality. Check whether it is a data frame or not. Order the entire data frame by the first and second columns. remove the variables 'Solar. R' and 'Wind' and display the data frame
3. Load dataset named ChickWeight,

(i). Order the data frame, in ascending order by feature name “weight” grouped by feature “diet” and Extract the last 6 records from an order data frame.

(ii). a Perform melting function based on “Chick", "Time", and "Diet"

features as ID variables

b. Perform cast function to display the mean value of weight

grouped by Diet

c. Perform cast function to display the mode of weight

grouped by Diet

1. Randomly Sample the iris dataset such as 80% data for training and 20% for testing and create Logistics regression with train data, use species as target and petals width and length as feature variables, Predict the probability of the model using test data, Create Confusion matrix for above test model.

# SAMPLE-5

1. Design an R program to dynamically extract a subset of specific levels from a

Factor created using a random sample from LETTERS, ensuring the solution can adapt to varying subset sizes and conditions.

2. Consider the following data present. Create this file using Windows Notepad.

Save the file as input.csv.

i.Use appropriate R commands to read the input.csv file.

ii. Analyze the CSV file and compute the following.

a.Get the maximum salary

b.Get the details of the person with the maximum salary

c.Get all the people working in IT department

d.Get the persons in IT department whose salary is greater than 600

e.Get the people who joined on or after 2014

1. Write an R Program to create a vector in R that generates 10 random integers

between -50 and +50. Once created, analyze the distribution of the values

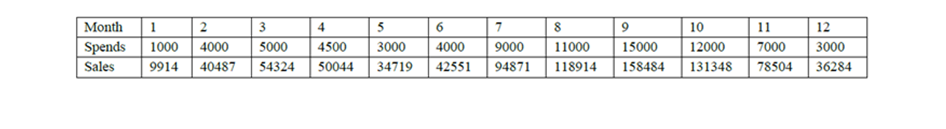
and discuss how you might modify the code to ensure specific characteristics,

such as all positive values or a specific range of variance.

1. How can you design and implement an R program that efficiently generates a sequence of numbers from 1 to 100, while incorporating conditional logic to identify and categorize multiples of 3, 5, and both? Additionally, reflect on how you can optimize the program to handle larger ranges of numbers and maintain clarity in the code structure.

# SAMPLE-6

1. Write an R script that uses a for loop to iterate over a vector, uses an if-else statement to check for a condition, and uses a function to perform an arithmetic operation. Ensure the function has default argument values and returns a complex object.
2. For this exercise, use the (built-in) dataset Titanic.
   1. Draw a Bar chart to show details of “Survived” on the Titanic based on passenger Class
   2. Modify the above plot based on gender of people who survived
   3. Draw a histogram plot to show the distribution of the feature “Age”
3. Write an R program to dynamically combine multiple arrays, ensuring that the rows are interleaved (i.e., the first row of each array, followed by the second row, and so on).
4. a. Create a data frame based on the below table



b. Create a regression model for that data frame table to show the amount of sales (Sales) based on how much the company spends (Spends) in advertising c. Predict the Sales if Spend=13500.

.

# SAMPLE-7

1. Write a R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from two given two vectors. Print the second row of the second matrix of the array and the element in the 3rd row and 3rd column of the 1st matrix.
2. Given two vectors, A=c(1,2,3,4,5,6)A = c(1, 2, 3, 4, 5, 6)A=c(1,2,3,4,5,6) and B=c(7,8,9,10,11,12)B = c(7, 8, 9, 10, 11, 12)B=c(7,8,9,10,11,12), write an R program to combine them into a 3x3x2 array. Print the array.
3. Explore the airquality dataset. It contains daily air quality measurements from New York during five months:
   * Ozone: mean ozone concentration (ppb),
   * Solar.R: solar radiation (Langley),
   * Wind: average wind speed (mph),
   * Temp: maximum daily temperature in degrees Fahrenheit,
   * Month: numeric month (May=5, June=6, and so on),
   * Day: numeric day of the month (1- 31).
4. Compute the mean temperature (don’t use built-in function)
5. Extract the first five rows from airquality.
6. Extract all columns from air-quality except Temp and Wind
7. Design an R program that not only sorts the data frame by name and score but also handles missing values by pushing records with missing scores to the end. How would you ensure flexibility for users to customize the handling of missing values?

# SAMPLE-8

1. Write an R program to draw an empty plot with specified axis limits and customize the axis labels to include mathematical notations or special characters using expression () and other R features.
2. Write an R program to create an array, passing in a vector of values and a vector of dimensions. Also, provide names for each dimension.
3. Get the Summary Statistics of the air quality dataset

i)Melt air-quality data set and display as a long–format data?

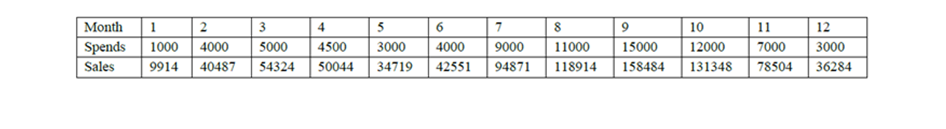
ii)Melt air-quality data and specify month and day to be “ID variables”?

iii)Cast the molten air-quality data set concerning month and date features

iv)Use the cast function appropriately and compute the average of Ozone,

Solar. R, Wind, and temperature per month?

1. a. Create a data frame based on the below table



b. Create a regression model for that data frame table to show the amount of sales (Sales) based on how much the company spends (Spends) based on expenditure c. Predict the Sales if the expenditure is Rs.13,500.

# SAMPLE-20

1. Given a dataset of daily average temperatures (in degrees Celsius) for a particular city over a month, perform a Univariate Exploratory Data Analysis (EDA) using R. The dataset is represented as an atomic vector daily\_temps with 30 entries, one for each day of the month. Your tasks are to:

i) Calculate the mean, median, and mode of the temperatures

ii) Compute the range, interquartile range (1QR), and standard

deviation to understand the dispersion.

2. Create two data frames: one with student personal information (StudentID, Name) and another with their scores (StudentID, MathScore, ScienceScore, EnglishScore). Merge these data frames based on "StudentID”. Show the merged Dataframe. Write the R code snippet necessary to perform these tasks.

3. Write an R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from two given two vectors. Print the second row of the second matrix of the array and the element in the 3rd row and 3rd column of the 1st matrix.

4. For this exercise, use the (built-in) dataset Titanic.

a. Draw a Bar chart to show details of “Survived” on the Titanic based

on passenger Class

b. Modify the above plot based on gender of people who survived

c. Draw a histogram plot to show the distribution of the feature

“Age”

# SAMPLE-9

1. Explore the USArrests dataset, which contains the number of arrests for murder, assault, and rape for each of the 50 states in 1973. It also contains the percentage of people in the state who live in an urban area.

(i) a. Explore the summary of the Data set, like the number of Features and its type. Find the number of records for each feature. Print the statistical feature of the data

b. Print the state which saw the largest total number of rape c. Print the states with the maximum & minimum crime rates for murder

1. a. Find the correlation among the features

b. Print the states that have assault arrests more than the median of the country

c. Print the states that are in the bottom 25% of murder

1. a. Create a histogram and density plot of murder arrests by US stat

b. Create a plot that shows the relationship between the murder arrest rate and the proportion of the population that is urbanized by the state. Then enrich the chart by adding assault arrest rates (by coloring the points from blue (low) to red (high)).

c. Draw a bar graph to show the murder rate for each of the 50 states.

1. Randomly Sample the iris dataset such as 80% data for training and 20% for testing and create Logistics regression with train data, use species as target and petals width and length as feature variables, Predict the probability of the model using test data, Create Confusion matrix for above test model
2. Write an R program to save the information of a data frame in a file and display the information of the file.
3. Write an R program to call the (built-in) dataset airquality. Check whether it is a data frame or not. Order the entire data frame by the first and second columns. remove the variables 'Solar. R' and 'Wind' and display the data frame

# SAMPLE-10

# 

* 1. Create a dataset containing information about products, including their

names, prices, and quantities in stock. Apply your knowledge of data frames

in R to create a data frame named "product\_data" from individual vectors

representing each product attribute. Ensure that the data frame is organized

and easily accessible. Write R code to create the "product\_data" data frame

and demonstrate how to access specific information, such as the average

price.

2. Write an R script to create a data frame with employee details (ID, Name,

Age, Salary). Perform operations to add a new column, filter rows based on a

condition, and sort the data frame by salary.

1. Create a logical vector in R Script based on conditions applied to a numeric

vector. Use logical operations (AND, OR, NOT) to perform complex filtering

on the vector.

1. Write a recursive function in R to generate the nth Fibonacci number. Prompt

the user to input a number “n” and use your function to calculate and print

the nth Fibonacci numbers.

# SAMPLE-11

1. Randomly Sample the iris dataset such as 80% data for training and 20% for testing and create Logistics regression with train data, use species as target and petals width and length as feature variables, Predict the probability of the model using test data, Create Confusion matrix for above test model.
2. Get the Summary Statistics of the air quality dataset

i)Melt air-quality data set and display as a long–format data?

ii)Melt air-quality data and specify month and day to be “ID

variables”?

iii)Cast the molten air-quality data set concerning month and date

features

iv)Use the cast function appropriately and compute the average of

Ozone, Solar. R, Wind, and temperature per month?

1. Design an R program to create and visualize the height factor from the women's dataset, categorizing heights into meaningful intervals. How would you ensure these intervals are appropriate and justified based on the dataset's distribution?
2. Design an R program to create a list comprising vectors, matrices, and a custom function. Analyze the structure and behavior of the list, focusing on how the diverse data types interact within the list. Evaluate the practical applications of lists in managing and processing heterogeneous data in R.

# SAMPLE-12

* 1. Design an R program to dynamically extract a subset of specific levels from a factor created using a random sample from LETTERS, ensuring the solution can adapt to varying subset sizes and conditions.
  2. Consider the following data present. Create this file using Windows Notepad. Save the file as input.csv.

i.Use appropriate R commands to read the input.csv file.

ii. Analyze the CSV file and compute the following.

a. Get the maximum salary

b. Get the details of the person with the maximum salary

c.Get all the people working in IT department

d.Get the persons in IT department whose salary is greater than 60.

3. Write an R Program to create a vector in R that generates 10 random integers

between -50 and +50. Once created, analyze the distribution of the values and

discuss how you might modify the code to ensure specific characteristics, such

as all positive values or a specific range of variance.

4. How can you design and implement an R program that efficiently generates a sequence of numbers from 1 to 100, while incorporating conditional logic to identify and categorize multiples of 3, 5, and both? Additionally, reflect on how you can optimize the program to handle larger ranges of numbers and maintain clarity in the code structure.

# SAMPLE-13

Explore the air quality dataset. It contains daily air quality measurements from New York during a period of five months:

Ozone: mean ozone concentration (ppb),

Solar.R: solar radiation (Langley),

Wind: average wind speed (mph),

Temp: maximum daily temperature in degrees Fahrenheit,

Month: numeric month (May 5, June 6, and so on),

Day: numeric day of the month (1-31).

1. a. Extract the first five rows from the database air quality

b. Compute the mean temperature (don't use built-in function)

2.a. How many days was the wind speed greater than 17 mph?

b. Find the Mean, Median, and Mode of the wind speed data

3. a. Get the Summary Statistics of air quality dataset

b. Plot the distribution of air quality"?

4. a. Find any missing values (NA) in features and drop the missing values if it's less than 10% of data else replace that with the mean of that feature.

b. Apply a linear regression algorithm using the Least Squares Method on

"Ozone" an SolaPlot

# SAMPLE-14

1.a. Write a function in R programming to print and generate a Fibonacci

sequence using Recursion in R

b. Find the sum of natural numbers up to 10, without a formula using a loop

statement.

2. Load in-build dataset "Chick Weight" in R

a. Explore the summary of the Data set, like the number of Features and

its type. Find the number of records for each feature.

b. Extract last 6 records of the dataset

3.a. Perform melting function based on "Chick", "Time", "Diet" features as

ID variables

b. Perform cast function to display the mean value of weight grouped by

Diet

4. a. Create multi regression model to find the weight of the chicken, by

"Time" and "Diet"

b. Predict weight for Time 10 and Diet-1

# SAMPLE-15

1. Design and construct a data frame in R with the given structure and data. Critically evaluate the dataset’s composition, including the interplay between numerical and categorical variables. Reflect on how this dataset can be utilized for various analytical purposes, such as student performance analysis or decision-making based on attempts and qualification status. Identify potential improvements or extensions to make the dataset more comprehensive for real-world applications.
2. Design an R program to create a list comprising vectors, matrices, and a custom function. Analyze the structure and behavior of the list, focusing on how the diverse data types interact within the list. Evaluate the practical applications of lists in managing and processing heterogeneous data in R.
3. Explore the USArrests dataset, which contains the number of arrests for murder, assault, and rape for each of the 50 states in 1973. It also contains the percentage of people in the state who live in an urban area.
   1. a. Explore the summary of the Data set, like the number of Features and its type. Find the number of records for each feature. Print the statistical feature of data

b. Print the state which saw the largest total number of rapes

c. Print the states with the maximum & minimum crime rates for murder

1. Design an R program to dynamically extract a subset of specific levels from a factor created using a random sample from LETTERS, ensuring the solution can adapt to varying subset sizes and conditions.

# SAMPLE-16

1.a. Write a function in R programming to print and generate a Fibonacci sequence

using Recursion in R

b. Find sum of natural numbers up to 10, without formula using loop statement.

2 a. Write suitable R code to compute the mean, median, mode of the following values

с (90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20)

b. Write R code to find 2nd highest and 3rd Lowest value of above problem.

3. Load mt cars dataset Use gg plot package to plot below EDA questions label the

plot accordingly

a. Create weight (wt) vs displacement (disp) scatter plot factor by Engine

Shape(vs)

b. Create horsepower (hp) vs mileage (mgp) scatter plot factor by Engine

Shape(vs)

c. In above(b) plot, Separate columns according to cylinders(cyl) size

4.Using linear regression analysis establish a relationship between height and weight

of a person using the input vector given below.

#Values of height

151, 174, 138, 186, 128, 136, 179, 163, 152, 131

#Values of weight.

63, 81, 56, 91, 47, 57, 76, 72, 62, 48

Predict the weight of a person with height 170. Visualize the regression

graphically.

# SAMPLE-17

1. a.Write a function in R programming to print and generate Fibonacci sequence using Recursion in R

b. Find sum of natural numbers up-to 10, without formula using loop

statement.

2. Load in-build dataset "Chick Weight" in R

a.Explore the summary of Data set, like number of Features and its type.

Find the number of records for each features

b. Extract last 6 records of dataset

c. order the data frame, in ascending order by feature name "weight"

grouped by feature "diet".

3.a. Perform melting function based on "Chick", "Time","Diet" features as ID

variables

b. Perform cast function to display the mean value of weight grouped by Diet

c. Create a Histogram for "Weight" features belong to Diet-2 category

4. a. Create multi regression model to find a weight of the chicken, by "Time"

and "Diet"

b. Predict weight for Time 10 and Diet-1

c. Find the error in model for same.

# SAMPLE-18

1.Write an R program create a multi-dimensional array in R using a vector of values and dimensions. Analyze the impact of assigning meaningful names to dimensions and discuss how this enhances the readability and usability of the array in real-world applications.

1. a.Write a function in R programming to print and generate Fibonacci sequence using Recursion in R

b. Find sum of natural numbers up-to 10, without formula using loop

statement.

1. Load dataset named ChickWeight,

(i). Order the data frame, in ascending order by feature name “weight” grouped by feature “diet” and Extract the last 6 records from an order data frame.

(ii). a Perform melting function based on “Chick", "Time", and "Diet" features as ID variables

1. Randomly Sample the iris dataset such as 80% data for training and 20% for testing and create Logistics regression with train data, use species as target and petals width and length as feature variables, Predict the probability of the model using test data, Create Confusion matrix for above test model.

# SAMPLE-19

1. Design an R program to dynamically extract a subset of specific levels from

a factor created using a random sample from LETTERS, ensuring the

the solution can adapt to varying subset sizes and conditions.

2. Consider the following data present. Create this file using Windows

Notepad. Save the file as input.csv.

i. Use appropriate R commands to read the input.csv file.

ii. Analyze the CSV file and compute the following.

a. Get the maximum salary

b. Get the details of the person with the maximum salary

c.Get all the people working in IT department

d.Get the persons in IT department whose salary is greater than 60.

3. Write an R Program to create a vector in R that generates 10 random integers

between -50 and +50. Once created, analyze the distribution of the

values and discuss how you might modify the code to ensure specific

characteristics, such as all positive values or a specific range of variance.

4. a. Write a function in R programming to print and generate a Fibonacci sequence

using Recursion in R

b. Find sum of natural numbers up to 10, without formula using loop statement.