CSAEC49

USN	1	M	S							

RAMAIAH Institute of Technology

(Autonomous Institute, Affiliated to VTU)
(Approved by AICTE, New Delhi & Govt. of Karnataka)
Accredited by NBA & NAAC with 'A+' Grade

SEMESTER END EXAMINATIONS – AUGUST 2024

Program	:	**B.E :- Computer Science and Engineering**	**Semester**	:	**IV**
Course Name	:	**Data Analysis using R**	**Max. Marks**	:	**50**
Course Code	:	**CSAEC49**	**Duration**	:	**2 Hrs**

Instructions to the Candidates:

- **PART – A** : Answer all the questions.
- **PART – B** : Answer one full question from each unit.

PART - A

1. What is the out put of the following code? CO1 (01)

```
list1 <- c(TRUE, 0.1)
list2 <- c(0,4+3i)
print(list1[1]||list2[1])
```

a. 0 b. 1
c. 4.1+3i d. TRUE

2. What is the out put of the following code? CO1 (01)

```
friend.data <- data.frame(
  friend_id = c(1:5),
  friend_name = c("Sachin", "Sourav",
                  "Dravid", "Sehwag",
                  "Dhoni"),
  stringsAsFactors = FALSE
)
dim(friend.data)
```

a. 5 b. 5 2
c. 2 5 d. 10

3. Among the following commands which one can Plot the matrices between 4 variables giving 12 plots? CO2 (01)

a.
```
pairs(~wt + mpg + disp +
    cyl, data = mtcars)
```

b.
```
plot(x = input$wt, y =
    input$mpg,
xlab = "Weight",
ylab = "Milage",
xlim = c(1.5, 4),
ylim = c(10, 25),
main = "Weight vs Milage"
)
```

c.
```
ggplot(mtcars, aes(x =
    log(mpg), y = log(drat))) +
  geom_point(aes(color =
factor(gear))) +
  stat_smooth(method = "lm",
  col = "#C42126", se =
FALSE, size = 1
)
```

d. none of the above

4. A box plot is a chart _____ CO2 (01)
   a. to display information in the form of distribution
   b. used to display information in the form of a series of data points.
   c. to visually represent categorical data in a structured manner.
   d. none of the above

5. Which function is used to calculate the mean of a numeric vector in R? CO3 (01)
   a. mean()
   b. average()
   c. median()
   d. central()

6. Which package in R provides functions for calculating skewness and kurtosis? CO3 (01)
   a. stats
   b. base
   c. moments
   d. summarytools

7. Which function in R is used to perform an independent t-test? CO4 (01)
   a. t.test()
   b. ttest()
   c. t_test()
   d. test.t()

8. When should you use a paired t-test? CO4 (01)
   a. Comparing means from two independent groups
   b. Comparing means from the same group at different times
   c. Testing for correlation between two variables
   d. Testing for independence in categorical data

9. Which package in R is commonly used to connect to SQL databases? CO5 (01)
   a. ggplot2
   b. DBI
   c. dplyr
   d. tidyr

10. How do you filter records in a table named sales where the amount is greater than 1000? CO5 (01)
   a. SELECT * FROM sales WHERE amount > 1000;
   b. FILTER * FROM sales WHERE amount > 1000;
   c. SELECT * FROM sales HAVING amount > 1000;
   d. FILTER * FROM sales HAVING amount > 1000;

PART - B
UNIT – I

1. a) Identify the types of operators used in R programming? Give examples for each one of them and describe. CO1 (04)
   b) How do you create a factor in R using the factor() function? Provide an example. CO1 (04)

2. a) Discuss the usages of string literals in R, distinguishing between single quotation marks (') and double quotation marks (")? CO1 (04)
   b) Explain the role of the sort() function in R. How does sort() differ from order() in terms of its output and application? CO1 (04)

UNIT – II

3.	a)	How to create a scatter plot in R using the plot() function? Include details on how to set the color of the data points, customize the plot title, and label the x and y axes.	CO2	(04)
	b)	How can you customize the axis limits of a plot in R? Create an example using the plot() function where you set the x-axis range from -10 to 10 and the y-axis range from -5 to 5.	CO2	(04)
4.	a)	Describe the steps needed to add a legend to a scatter plot in R. Provide an example where the legend is positioned at the top right of the plot and explain the parameters used in the legend() function.	CO2	(04)
	b)	How can you combine multiple scatter plots into one plot in R? Provide an example using the par(mfrow=c(n,m)) function to create a 2x2 grid of scatter plots.	CO2	(04)

UNIT – III

5.	a)	Describe the process of calculating the median of a dataset in R. Why is the median an important measure of central tendency?	CO3	(04)
	b)	Discuss variance, and how is it computed in R? Provide an example using a numeric vector.	CO3	(04)
6.	a)	What does kurtosis measure in a dataset? Provide an example of how to calculate kurtosis in R and interpret its value.	CO3	(04)
	b)	Identify the key differences between variance and standard deviation? Give example and show.	CO3	(04)

UNIT – IV

7.	a)	What is the purpose of a t-test in hypothesis testing, and how do you perform an independent t-test in R? Provide an example using a sample dataset.	CO4	(04)
	b)	Describe how to visualize the correlation between two variables in R. Which plot would you use, and how do you create it?	CO4	(04)
8.	a)	When would you use a paired t-test instead of an independent t-test? Illustrate with a scenario and corresponding R code.	CO4	(04)
	b)	Explain how to interpret the results of a Chi-Square test in R. What do the test statistics and p-value indicate?	CO4	(04)

UNIT- V

9.	a)	Explain the purpose of the WHERE clause in SQL. Provide an example of using it in an R query to select employees with the job title 'Manager'.	CO5	(04)
	b)	What is the purpose of the ORDER BY clause in SQL? Provide an example of using it in an R query to sort the results by the date column in ascending order.	CO5	(04)
10.	a)	How do you use the LIKE operator in an SQL statement in R to find all entries in a products table where the product name contains the word 'Laptop'? Provide an example.	CO5	(04)
	b)	How do you use the LIMIT clause in an SQL statement in R to retrieve only the first 10 rows from a table named orders? Provide an example.	CO5	(04)

****