VR17

Reg. No:

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

II/IV B.Tech. DEGREE EXAMINATION, SEPTEMBER, 2021

Fourth Semester

INFORMATION TECHNOLOGY

17IT3401 STATISTICS WITH R

Time: 3 hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part - B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. Write a short note on command line interface.
   b. Define calling function.
   c. Explain aggregate function in R programming.
   d. Write a short note on Calculating a Probability.
   e. Define Weighted mean.
   f. Explain normal distribution.
   g. Mention any two applications of t-tests.
   h. Define simple linear Regression.
   i. Define Multiple Regression.
   j. Write a short note on non linear least squares.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Explain about basic math and variables of R. **7M**
   b. Discuss about matrices in R. **8M**

(or)

3. a. Explain about return values in R programming. **7M**
   b. Write about control statements in R. **8M**

UNIT-II

4. a. What are set operations? Give examples of each. **7M**
   b. Explain apply method in R. Write about lapply, sapply with suitable examples? **8M**

(or)

5. a. Write about the following with suitable example **7M**
   a) Cumulative Sum b) Cumulative Max
   c) Cumulative Min d) Cumulative product
   b. Write about sort, rank and order functions with examples. **8M**

UNIT–III

6. a. Fit a Binomial distribution to the following data **7M**

   x= 0 1 2 3 4 5

   f= 2 16 28 12 9 3

   b. What is Poisson distribution? Explain with examples. **8M**

(or)

7. a. Explain concepts of correlation and covariance with examples. **7M**
   b. Discuss about ANOVA. **8M**

UNIT-IV

8. a. Using the least square method fit a straight line for the following data: **7M**

x	-1	0	1	2	3	4
f(x)	1	0	1	4	5	5

   b. Write in detail about Random Forest. **8M**

(or)

9. a. Explain Autoregressive Moving Average model in R programming. **7M**
   b. Discuss about GARCH family of models. **8M**

* * *

VR17 Reg. No: | | | | | | | | | | |

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

II/IV B.Tech. DEGREE EXAMINATION, NOVEMBER, 2020

Fourth Semester

INFORMATION TECHNOLOGY

17IT3401 STATISTICS WITH R

Time: 3hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part-B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. Write any three linear algebra math operations.
   b. What is data reshaping?
   c. Write the syntax of strsplit() function.
   d. How missing values are represented in R language?
   e. List any three base plot functions.
   f. Compute the intersection of {1, 2...10} and {5, 6, . . , 15}
   g. What is the use of dnorm() function?
   h. Define decision tree.
   i. Write any two disadvantages of R programming language.
   j. What are the applications of t-distribution?

PART-B

4 x 15 = 60M

UNIT-I

2. a. Explain abut vectors and data.frames in R with suitable examples. **7M**
   b. Explain various ways of reading data into R programming. **8M**

(or)

3. a. Write R code to define the function by using if-else **7M**

   f(x) = x if x<1/2
   = (1-x) if ½<x<1
   = 0 otherwise

   b. Write about nested if in R. **8M**

UNIT-II

4. a. What are the apply family functions available in R? Explain with examples. **9M**
   b. Explain how to generate a random variable, with an example? **6M**

(or)

5. a. Write an example for a combinatorial simulation in R. **8M**
   b. Explain about built in random variable generators. **7M**

UNIT-III

6. a. Fit a Binomial distribution to the following data **8M**

x	0	1	2	3	4	5
f	2	16	28	12	9	3

   b. Discuss about correlation and covariance. **7M**

(or)

7. a. Explain in detail about Poisson distribution. **6M**
   b. Input a data set and explain the procedure of performing one way ANOVA test. **9M**

UNIT-IV

8. a. Explain how K-means is implemented in R? **8M**
   b. Discuss about logistic regression. **7M**

(or)

9. a. What are the disadvantages of the linear model? **3M**
   b. Discuss briefly about decision trees and how to implement in R? **12M**

* * *

VELAGAPUDI RAMAKRISHNA
SIDDHARTHA ENGINEERING COLLEGE
(AUTONOMOUS)
II/IV B.Tech. DEGREE EXAMINATION, APRIL, 2019
Fourth Semester
INFORMATION TECHNOLOGY
17IT3401 STATISTICS WITH R

Time: 3 hours ***Max. Marks: 70***
Part-A is compulsory
Answer One Question from each Unit of Part - B
Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. When do you use t-test?
   b. What is the use of With () function in R?
   c. What is the use of subset() and sample() function in R?
   d. How transpose works in R?
   e. What is the function used for adding datasets in R?
   f. How can you produce correlation and covariance?
   g. What is the difference between matrix and dataframes?
   h. What is the function used for merging of data frames horizontally in R?
   i. What is the use of ggplot2?
   j. Define moving average.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Create a data frame that stores the name, age, designation of the employee. Find how many employees are working in each designation? **8M**
   b. Create a function to calculate the average, median and mean for a numeric vector age in employee database. **7M**

(or)

3. a. Find the cube and square of a given number using while and for loop. **7M**
   b. Create two vectors that stores the details of name and gender of the employees. Find how many 'male' and 'female' employees are present? **8M**

UNIT-II

4. a. i) What is the significance of regular expression in R? List any four functions in it with examples. **5M**
   ii) What is use of llply() and lapply () in R? **5M**
   b. Discuss about merge () function in R. **5M**

(or)

5. a. Discuss about the following with an example **10M**
   i) cbind ii) rbind
   b. Explain how set operations are handled in R with example? **5M**

UNIT-III

6. a. If there are twelve cars crossing a bridge per minute on average, find the probability of having seventeen or more cars crossing the bridge in a particular minute. **10M**
   b. Explain about usage of summary() function in R with examples. **5M**

(or)

7. a. Assume that the test scores of a college entrance exam fits a normal distribution. Furthermore, the mean test score is 72 and the standard deviation is 15.2. What is the percentage of students scoring 84 or more in the exam? **10M**
   b. Discuss about the significance of ANOVA test. **5M**

UNIT-IV

8. a. 'stackloss' is a dataset that records the observations of a chemical plant operation. Consider stackloss as the dependent variable, and other variables such as Air.Flow (cooling air flow), Water. Temp (inlet water temperature) and Acid.Conc. (acid concentration) are considered as independent variables. Develop multiple regression model for the above data set and predict the stackloss if the air flow is 72, water temperature is 20 and acid concentration is 85. **8M**
   b. Decide which of the independent variables in the multiple linear regression model of the data set stackloss (given in 8.(a)) are statistically significant at 0.05 significance level. Use the below table for inference. **7M**

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-39.920	11.896	-3.36	0.0038 **
Air.Flow	0.716	0.135	5.31	5.8e-05 ***
Water.Temp	1.295	0.368	3.52	0.0026 **
Acid.Conc.	-0.152	0.156	-0.97	0.3440

(or)

9. a. Design PAM algorithm in R and explain with an example. **10M**
   b. Write short notes on Hierarchical clustering. **5M**

* * *

b. What are the limitations of k-means clustering method? How to perform Partitioning Around Medoids in R with the help of 'pam()' function? **8M**

* * *

Reg. No: | | | | | | | | | | |

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

II/IV B.Tech. DEGREE EXAMINATION, MARCH, 2021

Fourth Semester

INFORMATION TECHNOLOGY

17IT3401 STATISTICS WITH R

Time: 3hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part-B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. Write the advantages of R programming language.
   b. What is a compound test in R?
   c. What is a regular expression in R?
   d. What is the difference between cbind and rbind?
   e. Define covariance.
   f. What is the use of regression analysis?
   g. What are the different sorting algorithms available in R?
   h. What is recursion function in R?
   i. What is the difference between data frame and matrix in R?
   j. Compute the mean of a vector of 100 random numbers.

PART-B

4 x 15 = 60M

UNIT-I

2. a. Write R function to display prime numbers up to a given range. **7M**

   b. What is a vector? How to create it? Create a vector A of elements 5, 2, -2, 6, 7, 10, 12, 14, 15 and from it create a vector Y containing elements of A > 6. **8M**

(or)

3. a. How to create variables in R? Explain various data types available. **8M**

   b. Explain the following sentence:
   Functions in R have (almost) no side effects. **7M**

UNIT-II

4. a. Explain
   i) Different set operations
   ii) Minima and maxima with suitable examples **7M**

   b. Explain how to implement simulations in R with examples? **8M**

(or)

5. a. Explain various joins with suitable examples. **7M**

   b. Explain about ddply and llply with necessary formulas and examples. **8M**

UNIT-III

6. a. Explain
   i) Two-Sample t-Test and
   ii) Paired Two-Sample t-Test functions **8M**

   b. Explain Gaussian distribution for the random normal variables and also draw the corresponding plot. **7M**

(or)

7. a. Explain about summary function and quantile function with examples for a given sample data. **8M**

   b. If only 5% kids can secure A grade in a paper, find the probability of at most 2 out of 10 kids getting A grade in that paper. **7M**

UNIT-IV

8. a. Briefly explain about non linear least squares. **8M**

   b. Discuss about random forests. **7M**

(or)

9. a. Explain vector autoregressive model and GARCH model for fitting. **7M**

VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS)

II/IV B.Tech. DEGREE EXAMINATION, NOVEMBER, 2019

Fourth Semester

INFORMATION TECHNOLOGY

17IT3401 STATISTICS WITH R

Time: 3hours ***Max. Marks: 70***

Part-A is compulsory

Answer One Question from each Unit of Part-B

Answer to any single question or its part shall be written at one place only

PART-A

10 x 1 = 10M

1. a. Which command is used for storing R object into a file?
   b. What is ggplot2?
   c. Explain about the significance of transpose in R language.
   d. dplyr package is used to speed up data frame management code. Which package can be integrated with dplyr for large fast tables?
   e. What will be the class of the resulting vector if you concatenate a number and a character?
   f. How can you verify if a given object 'X' is a matrix data object?
   g. What will be the output of runif(7)?
   h. Write a function to extract the first name from the string 'Mr. Tom White'.
   i. State whether the equation given below is linear or not?
   Emp_sal= $2000+2.5(emp\_age)^2$
   j. How can you merge two data frames in R language?

PART-B

4 x 15 = 60M

UNIT-I

2. a. Write a program to check for eligibility to vote. Consider age value of person to be a vector field. If a person age is greater than 18 then he is eligible to vote otherwise he is not eligible. Show how to use a while loop to achieve the same result? **10M**

   b. Create a function that calculates the multiples of a given number. **5M**

(or)

3. a. Discuss how arrays are created in R? **8M**

   b. Create a function that finds the sum of digits for a given number. **7M**

UNIT-II

4. a. Consider the following database **10M**
   Flight (Year, Month, DayofMonth, DayofWeek, DepTime, ArrTime, UniqueCarrier)
   i) Convert the above database into dataframe
   ii) Retrieve the January month flights
   iii) Sort the database based on DayofMonth
   iv) Display the Year, month and DayofMonth of the flight database
   v) Find the number of flights run for each year.

   b. Briefly discuss about regular expression. **5M**

(or)

b. Discuss about the significance of t-test. **5M**

UNIT-IV

8. a. Apply the simple linear regression model for the data set faithful of question 6(b) and estimate the next eruption duration if the waiting time since the last eruption has been 80 minutes. **10M**

   b. Decide whether there is a significant relationship between the variables in the linear regression model of the data set faithful of question 6(b) at 0.05 significance level. **5M**

```
Coefficients:
             Estimate Std. Error   t value   Pr(>|t|)
(Intercept) -1.87402     0.16014    -11.7     <2e-16 ***
waiting      0.07563     0.00222    34.1      <2e-16 ***
```

(or)

9. Design K-means algorithm in R and explain with an example. **15M**

* * *

5. a. Elaborate on the math functions that calculates probability, cumulative sums and products with example. **8M**

   b. Discuss with an example, how random numbers are generated in R? **7M**

UNIT-III

6. a. Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random. **8M**

   b. Assume the dataset 'faithful' given below,

	eruptions	waiting
1	3.600	79
2	1.800	54
3	3.333	74
4	2.283	62
5	4.533	85
6	2.883	55

   Find the covariance of eruption duration and waiting time from the above data set 'faithful'. Specify whether there is any linear relationship between the two variables. **7M**

(or)

7. a. Assume that the test scores of a college entrance exam fits a normal distribution. Further more, the mean test score is 72 and the standard deviation is 15.2. What is the percentage of students scoring 84 or more in the exam? **10M**