

**17IT3401**

- 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**

\* \* \*

**VR17**

Reg. No: 

--	--	--	--	--	--	--	--	--	--

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?  
 $\text{Emp\_sal} = 2000 + 2.5(\text{emp\_age})^2$
- 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)

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**