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**

* * *

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

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