Code No: R1621051 **R16** SET - 1

II B. Tech I Semester Regular/Supplementary Examinations, October/November - 2018
STATISTICS WITH R PROGRAMMING
(Com to CSE & IT)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B)**
2. Answer **ALL** the question in **Part-A**
3. Answer any **FOUR** Questions from **Part-B**

~~~~~~

PART –A

1. a) Write the Advantages of R Programming language.
   b) Two vectors X and Y are defined as follows :
   X <- c(3, 2, 4) and Y <- c(1, 2).
   What will be output of vector Z that is defined as Z <- X*Y.
   c) How missing values and impossible values are represented in R language?
   d) List some Base plot functions
   e) Define Correlation and Covariance
   f) What is Regression & explain Simple Linear Regression

PART -B

2. a) Create the following vectors in R.
   a = (5, 10, 15, 20, ..., 160) b = (87, 86, 85, ..., 56)
   Use vector arithmetic to multiply these vectors and call the result 'd'. Select subsets of d to identify the following.
   i. What are the 19^{th}, 20^{th}, and 21^{st} elements of d?
   ii. What are all of the elements of d which are less than 2000?
   iii. How many elements of d are greater than 6000?
   b) Create a vector grades with numbers 1 to 10. Pass the function to each element of the vector as the argument grade, with values for on_time and optional_part always fixed at TRUE and FALSE respectively.

3. a) Create a vector with some of your friend's names
   i. Get the length of above vector
   ii. Get the first two friends from above vector
   iii. Get the 2nd and 3rd friends
   iv. Sort your friends by names using 2 methods
   b) Remove missing value from c(1, 2, NA, 4) & Pick 50 random numbers between 1 to 100, with replacement

4. a) i. Generate a multiplication table for numbers ranging from 1 to 10.
   ii. You have a set of colors to choose from:
   colors <- c("red", "blue", "green", "white", "black", "yellow")
   You have to pick 3 colors and you cant' pick the same color more than once. List all possible combinations.
   b) Create an R script that calculates the square root of a given integer vector x of length one, if the value contained in x is negative it should return NA.

|'''|'|'|''|'||'|||
~~~~~~

5. a) Explain student t-distribution, its properties and applications.

   b) Given that p(x=2)=9p(x=4) 90 p(x=46) for a Poisson variate X.
   Find: i. P(x<2) ii. p(x>4) iii. p(x 1)

6. a) Calculate coefficient of correlation between age of cars and annual maintenance.
   Age of cars (years) 2 4 6 7 8 10 12
   Annual maintenance cost (rupees) 1600 1500 1800 1900 1700 2100 2000

   b) Anil is taking part in four competitions. If the probability of him winning any competition is 0.30, find the probability of him winning at least one competition.

7. a) The values of y and their corresponding values of y are shown in the table below

x	0	1	2	3	4
y	2	3	5	4	6

   a) Find the least square regression line y = a x + b.
   b) Estimate the value of y when x = 10.

   b) Discuss the following:
   i. What does P-value signify about the statistical data?
   ii. What are the disadvantages of the linear model?
   iii. What are the possible ways of improving the accuracy of a linear regression model?

|'''|'|'|''|'||'|||

Code No: R1621051 **R16** SET - 2

II B. Tech I Semester Regular/Supplementary Examinations, October/November - 2018
STATISTICS WITH R PROGRAMMING
(Com to CSE & IT)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B)**
2. Answer **ALL** the question in **Part-A**
3. Answer any **FOUR** Questions from **Part-B**

~~~~~~

PART –A

1. a) How many data structures does R language have?
   b) How will you merge two data frames in R programming language?
   c) What are the different types of sorting algorithms available in R language?
   d) Compute the intersection of {1, 2, . . . , 10} and {5, 6, . . . , 15}
   e) Create a factor 'drinks' & Examine the representation of the factor
   f) What is correlation Analysis & T-test?

PART -B

2. a) Create three vectors x,y,z with integers and each vector has 3 elements. Combine the three vectors to become a 3×3 matrix A where each column represents a vector. Change the row names to a,b,c.
   b) How the Random vectors created with a set of functions explain with suitable example.
      i. If a vector is passed to an arithmetic calculation how it will be computed?
      ii. If the vectors involved are of different lengths what will be the resultant vector ?

3. a) Write R code to implement Binary Search Tree.
   b) What is Recursion? What is Recursive Function in R? Find Sum of Series $1^2+2^2+3^2+\ldots..+n^2$ using the Recursive function.

4. a) i. Compute the truth table for logical OR. The function R computes the logical EXCLUSIVE-OR. What is the difference between the two?
      ii. Consider the vector 1:K, where K is a positive integer. Write an R command that determines how many elements in the vector are exactly divisible by 3.
   b) Write R codes that takes the coefficients of a quadratic equation, and outputs an appropriate message for the cases of
   (i) two distinct roots (b2 − 4ac > 0)
   (ii) coincident roots (b2 = 4ac) or
   (iii) complex roots (b2 < 4ac).

|'''|'|'|''|''||'|||
~~~~~~

5. a) Please obtain the transpose matrix of B named tB .

   b) Consider A=matrix(c(2,0,1,3), ncol=2) and B=matrix(c(5,2,4,-1), ncol=2).
   a) Find **A + B**
   b) Find **A – B**

6. a) Discuss in detail about Poisson Distributions

   b) If only 5 percent 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.

7. a) The sales of a company (in million dollars) for each year are shown in the table below.

x (year)	2005	2006	2007	2008	2009
y (sales)	12	19	29	37	45

   a) Find the least square regression line y = a x + b.
   b) Use the least squares regression line as a model to estimate the sales of the company in 2012.

   b) i. Differentiate between univariate, bivariate and multivariate analysis.
   ii. What do you understand by the term Normal Distribution?

|'''|'|'|''|'||'|||

Code No: R1621051 **R16** SET - 3

II B. Tech I Semester Regular/Supplementary Examinations, October/November - 2018
STATISTICS WITH R PROGRAMMING
(Com to CSE & IT)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B)**
2. Answer **ALL** the question in **Part-A**
3. Answer any **FOUR** Questions from **Part-B**

~~~~~~

PART –A

1. a) What is the difference between data frame and a matrix in R?
   b) If you want to know all the values in c (1, 3, 5, 7, 10) that are not in c (1, 5, 10, 12, 14). Which in-built function in R can be used to do this? Also, how this can be achieved without using the in-built function.
   c) Plot 100 normal random numbers.
   d) Compute the mean of the square root of a vector of 100 random numbers.
   e) Write the Disadvantages of R Programming language.
   f) What is anova test? Explain dnorm() function

PART -B

2. a) 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
   b) Write R function to check whether the given number is prime or not

3. a) Make a script file which constructs three random normal vectors of length 100. Call these vectors x1, x2 and x3. Make a data frame called t with three columns (called a, b and c) containing respectively x1, x1+x2 and x1+x2+x3.
   Call the following functions for this data frame: plot(t) and sd(t).
   b) Write a function for Make a vector from 1 to 100. Make a for-loop which runs through the whole vector. Multiply the elements which are smaller than 5 and larger than 90 with 10 and the other elements with 0.1., so that you can feed it any vector you like (as argument). Use a for-loop in the function to do the computation with each element.

4. a) Create a character vector with length of number-of-rows-of-iris-dataset, such that, each element gets a character value – "greater than 5″ if the corresponding 'Sepal.Length' > 5, else it should get "lesser than 5″.
   Create a logic for the same problem statement using apply() function
   b) Use the functions mean() and range() to find the mean and range of: (a) the numbers 1, 2, . . . , 21 (b) the sample of 50 random normal values, that can be generated from a normaL distribution with mean 0 and variance 1 using the assignment y <- rnorm(50). (c) the columns height and weight in the data frame women

|'''|'|'|''|'||'|||
~~~~~~

5. a) There are 3 color palletes: the first one has 4 colors, the second has 6 colors and the third has 8 colors. You have to pick a pallete and then choose up to 5 (1, 2, 3, 4 or 5) colors from the chosen color pallete. How many different possibilities are there?
   b) What is Recursion? What is Recursive Function in R? Find the sum of natural numbers using the Recursive function.

6. a) If a committee has 7 members, find the probability of having more female members than male members given that the probability of having a male or a female member is equal.
   b) The data are presented below. Compute the appropriate t-test.

Older Adults	Younger Adults
45	34
38	22
52	15
48	27
25	37
39	41
51	24
46	19
55	26
46	36
Mean =	Mean =
S =	S =
S^2 =	S^2 =

7. a) The table below shows the number of absences, x, in a Calculus course and the final exam grade, y, for 7 students. Find the correlation coefficient and interpret your result.

x	1	0	2	6	4	3	3
y	95	90	90	55	70	80	85

   b) Discuss the following:
      i. What is difference between simple linear and multiple linear regressions?
      ii. What is difference between regression model, and estimated regression equation?
      iii. How to find f test and t test p values

|'''|'|'|''|'||'|||

Code No: R1621051 **R16** SET - 4

II B. Tech I Semester Regular/Supplementary Examinations, October/November - 2018
STATISTICS WITH R PROGRAMMING
(Com to CSE & IT)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B)**
2. Answer **ALL** the question in **Part-A**
3. Answer any **FOUR** Questions from **Part-B**

~~~~~~

PART –A

1. a) Create a vector containing following mixed elements {1, 'a', 2, 'b'} and find out its class.
   b) How to check the frequency distribution of a categorical variable?
   c) Write a function in R language to replace the missing value in a vector with the mean of that vector
   d) Differentiate between lapply and sapply.
   e) How will you create scatterplot matrices in R language?
   f) Initialise a character vector of length 26What is the procedure to check the cumulative frequency distribution of any categorical variable?

PART -B

2. a) i. Compute the difference between 2014 and the year you started at this university and divide this by the difference between 2014 and the year you were born. Multiply this with 100 to get the percentage of your life you have spent at this university.
   ii. Compute the sum of 4, 5, 8 and 11 by first combining them into a vector and then using the function sum.
   b) Find the class of *'iris'* dataframe, find the class of all the columns of *'iris'*, get the summary of *'iris'*, get the top 6 rows, view it in a spreadsheet format, get row names, get column names, get number of rows and get number of columns.
   Get the last 2 rows in last 2 columns from iris dataset
   Get rows with Sepal.Width > 3.5 using which() from iris

3. a) applying the functions median() and sum()
   (a) the numbers 1, 2, . . . , 21
   (b) the sample of 50 random normal values, that can be generated from a normaL distribution with mean 0 and variance 1 using the assignment y <- rnorm(50).
   (c) the columns height and weight in the data frame women.
   b) Let vector y be the logarithm of a random sample from a standard normal distribution, N(0,1). Use the ifelse function to replace missing values with the value 9999.

|'''|'|'|''|'||'|||
~~~~~~

4. a) Compound interest can be computed using the formula $A = P \times (1+ R/100)^n$ where P is the original money lent, A is what it amounts to in n years at R percent per year interest. Write R code to calculate the amount of money owed after n years, where n changes from 1 to 15 in yearly increments, if the money lent originally is 5000 pounds and the interest rate remains constant throughout the period at 11.5%.

   b) Find the transpose matrix of **A.**
   Find the value of x on **Ax=b.**

5. a) Compute the correlation coefficient for the following data

A	58	68	75	45	44	67	23	33	55	44
B	52	72	68	53	65	66	35	45	66	54

   b) What is Box plot? Explain importance of box plot with suitable example?

6. a) A test is conducted which is consisting of 20 MCQs (multiple choices questions) with every MCQ having its four options out of which only one is correct. Determine the probability that a person undertaking that test has answered exactly 5 questions wrong.

   b) Discuss in detail about Poisson Distributions

7. i. A CEO is measuring how an increase in the pay of his employees (x) impacts how long they voluntarily stay at work (y). What is the regression equation for this data?

x	1	3	6	9
y	4	10	19	28

   ii. How do you interpret b1 in simple linear regression?
   iii. What is adjusted R^2?
   iv. What is logistic regression? Or State an example when you have used logistic regression recently.

|'''|'|'|''|'||'|||