
Assignment - R Programming

Q1. Create a file that contains 1000 lines of random strings.

Q2. Create a random dataset of 100 rows and 30 columns. All the values are defined between [1,200]. Perform the following operations:

(i) Replace all the values with NA in the dataset defined between [10, 60]. Print the count of number rows having missing values.

(ii) Replace all the NA values with the average of the column value.

(iii) Find the Pearson correlation among all the columns and plot heat map. Also select those columns having correlation <=0.7.

(iv) Normalize all the values in the dataset between 0 and 10.

(v) Replace all the values in the dataset with 1 if value <=0.5 else with 0.

Q3. Create a random dataset of 500 rows and 10 columns.
Columns 1 to 4 are defined between [-10, 10];
Columns 5 to 8 are defined between [10, 20];
Columns 9 to 10 are defined between [-100, 100].
Apply following clustering algorithms; determine the optimal number of clusters and plot distance metric graph using each algorithm.

(i) K-Mean clustering

(ii) Hierarchical clustering

Q4. Create a random dataset of 600 rows and 15 columns. All the values are defined between [-100,100].
Perform the following operations:

(i) Plot scatter graph between Column 5 and Column 6.

(ii) Plot histogram of each column in single graph.

(iii) Plot the Box plot of each column in single graph.

Q5. Create a random dataset of 500 rows and 5 columns:
All the values are defined between [5,10].
Perform the following operations:

(i) Perform t-Test on each column.

(ii) Perform Wilcoxon Signed Rank Test on each column.

(iii) Perform Two Sample t-Test and Wilcoxon Rank Sum Test on Column 3 and Column 4