Project Assignment #3   **12.09.2018**

Project objective:

As the core topic of data analytics, statistical learning is one of the most important topics to learn and develop the skill set. With this objective, in this course, we have learned the following topics:

  i.   Hypothesis testing as the non-parametric based approach to statistical inference.
  ii.   Analysis of variance.
  iii. Correlation analysis.
  iv.   Regression analysis (linear (simple and multiple), nonlinear and auto-regression analysis

The projects under this assignment are to practice the concepts on the above-mentioned topics with real life data. You are advised to implement only ONE of the following projects as assigned to you. You are advised to implement the project using R programming language. A sincere student should attempt to implement all the projects, if possible; however, only the assigned project need to be submitted and will be evaluated.

Topic 3.1

Pearson's correlation analysis
Reference: CAMERA dataset

There are 11 attributes specifying the different camera products.
  a. It will be interesting to find out if any pair of attributes is highly correlated or not. In this project, you have to find the pair which has Pearson's correlation coefficient is greater than 0.80.
  b. Also, test the significance with 5% confidence level.

Topic 3.2

Spearman's correlation analysis
Reference: SNACKS dataset.

  a. Find the attribute(s) with which the "linking score" is most highly correlated.
  b. Test the significance of the correlation at 1% confidence level.

Topic 3.3

Regression analysis
Reference: MACHINE dataset.

    a. Find if PRP and ERP are linearly or non-linearly related. To test the non-linearity, you should check up degree 3.
    b. In each case, also calculate R^2, the measure of "Quality of Fit".

Topic 3.4

Auto-Regression Analysis
Reference : STOCK dataset.

    a. For the given stock dataset, predict the stock value in the month of 2017-10-1.
    b. Consider your calculations up to order 3 and compare the results that you have observed in different order, that is, AR(1), AR(2) and AR(3).

Topic 3.5

Analysis of variance (ANOVA)
Reference: IRIS dataset.

For the given dataset find the following.
    a. Apply analysis of variance to decide whether the mean values are significantly different or not. Consider each attribute separately.
    b. Test your hypothesis at two different confidence levels namely 1% and 5%.

Submission procedure:

1. Prepare a report which should include tool used, methodology followed, reasonable assumptions, if any, etc.
2. Submit the program file.
3. You may create a tar file including the above document using any zip program and submit the same to Moodle system at https://10.5.18.110/moodle/login/index.php .
4. Plagiarism, if found, should be taken seriously.

5. **Last date of submission is: 14.10.2018, 24:00 hours (hard deadline).**