Indian Institute of Management Tiruchirappalli

PGP Term IV

BSF (2019) - Assignment 1 (R Programming)

Submission Date: 18th June 2019 11:59 AM
Maximum Marks: 10

This is a group assignment. You need to use R programming for all data cleaning, manipulation, and analysis. Submit soft copy of your assignment, R codes, and other supporting documents in the Google Classroom (No hard copy/email). Show adequate reasoning, including calculations, if any, in support of your answers. State any assumption you make (your assumptions need to be reasonable).

Mention your Group No (1 or 2) and team No (1 to 10) in your submission file-name (`Assignment1_GR1_TEAM5.R`). Also, mention your team details in one your uploading files.

The following questions are based on **Housing.xlsx**. Use **readxl** r-package to load the xlsx file (or you can convert the xlsx to csv). You can find the details about the dataset here `https://vincentarelbundock.github.io/Rdatasets/doc/Ecdat/Housing.html`.

1. Return a data frame with columns: price, lotsize, bedrooms
2. What is the mean price in the 4 bedrooms segment?
3. Compute the summary of price per lotsize.
4. Create a data frame/table called stories.summary whose rows are the stories and whose columns are: number of houses, mean house price, median house price, and maximum house price. (*Hint: aggregate*)
5. How many houses have lotsize between 4000 and 6000
6. Order the house by highest lotsize and print the 10 highest lotsize apartments.
7. What is the 'coefficient of variation' of house price with and without air condition. (Hint: $c_v = \sigma/\mu$)
8. Boxplot of house prices with both preferred and non preferred area in a same graph.
9. Scatter plot of price and lotsize. Calculate the linear correlation.
10. Does the relationship between price and lotsize same for all stories segments? (*Hypothesis testing is not required;subjective arguments based on the chart or summary is enough*)