Home Assignment: Individual

Linear Regression

In this assignment, we use New York City condo evaluations data for fiscal year 2011-2012, obtained through NYC Open Data[1]. The data unit (*i.e.*, each row in the data) is a building. The original data have been cleaned up and stored in "housing.csv". For this data, the response variable (*i.e.*, the dependent variable, Y) is the value per square foot (`ValuePerSqFt`). And, the predictors (*i.e.*, the independent variables, X) are everything else. However, we focus on the following predictors:

- `SqFt`: total square footage of floor area in the building
- `Units`: number of units in the building
- `Boro`: the borough in which the building is located

Before running multiple linear regression models, you need to go through the command lines in the R script file "Data Preparation and Preliminary Analysis.R" to get data ready for your analyses.

Questions:

(1) Can the value per square foot be explained in terms of "`Units`", "`SqFt`", and "`Boro`" when the predictors are considered simultaneously? Interpret the results.

(2) Given the linear regression model you've specified in question (1), does the impact of "`Units`" on "`ValuePerSqFt`" vary across "`Boro`"? Or, in other words, does "`Boro`" moderate the influence of "`Units`" on "`ValuePerSqFt`"? Interpret the results.

(3) Given the linear regression model you've specified in question (1), does the impact of "`SqFt`" on "`ValuePerSqFt`" vary across "`Boro`"? Or, in other words, does "`Boro`" moderate the influence of "`SqFt`" on "`ValuePerSqFt`"? Interpret the results.

(4) Given the linear regression model you've specified in question (1), add a quadratic term of "`Units`" to model. Interpret the results of this updated model.

(5) Given the linear regression model you've specified in question (1), add a quadratic term of "`SqFt`" to model. Interpret the results of this updated model.

(6) Given the linear regression model you've specified in question (1), add an interaction term of "`Units`" and "`SqFt`" to the model. Interpret the results of this updated model.

[1] Its website is at https://opendata.cityofnewyork.us/

Note:

This is an individual based home assignment. Each student must email her/his report to Mr. Frank Yuan by 9:00 p.m., April 10. In the report, you are required to interpret your analyses and results meaningfully. In addition to the written report, you also need to submit the R code file to Frank. Visualize your findings to support your conclusions when necessary.