Final Project

Math 2200-Advanced Statistics

March 28, 2021

Project Description

Overview

This project is meant to be a capstone experience, meaning that you are expected to apply what you have learned in the class this semester when analyzing a real set of data.

You may work alone or in a group of up to 3 people. Each group will submit a single report. Members in groups of more than one are expected to contribute equally to the analysis of the data and to the writing of the report. **All members in a group will receive the same final project grade.**

I will provide you with at least one multivariate data set to analyze, but if you have your own data to analyze (perhaps data arising from your own research or consulting work or from the literature or the web), you may use this as long as you check with me to make sure it is appropriate for the purposes of this project.

Analysis

In addition to looking at the basic descriptive statistics of the data, you will analyze the data set with at least two methods(Multiple linear regression, logistic regression, or nonparametric test) we have already learned (Chapters 1,2,3, 6, and 7) . When applying a specific method, make sure to state any conditions or assumptions of this method and check to see if they are met. For example, if one assumption is that the observation is a normal distribution, you should use Q-Q plots of the residuals and other scatter plot to access nornality. If they are not met, and you are not able to transform the data in any way so that they are met, make sure to comment on this in your final report (you should still perform the intended analysis, however).

Report

You should write a report loosely formatted as a manuscript and include the following sections:

1. Introduction
2. Data description and exploratory analysis
3. Methods and analysis (*This section can be split into multiple sections, one for each type of analysis performed.)
4. Conclusions

Sections in your report should be clearly defined with a proper paragraphs, grammar, and complete sentences should be used throughout. Tables and figures should be labeled and captioned. **Including computer output by cutting and pasting huge sections of output is not acceptable;** you need to pick from the output the important quantities and include them (NEATLY!!!) in the body of your report. Other than these requirements, there is no other required format (APA, MLA, etc.) for the final report. If you do a bit of research on your own relevant to your project, a list of literature citations should also be included.

A certain length is not required as I am more concerned that your report is complete. However, projects between 7 and 12 pages are probably a good length. If you wish to include (unedited) software code and output, please do so in an Appendix

Due Date

Projects are due on Tuesday, April 20 by 5:00 pm. You must submit only one pdf file via grade-scope.

Grading

:There are 100 points possible for this project. I will assign scores based on the point breakdown below.

- Introduction [10]
- Data Description and explanatory analysis [25]
- Method and Analysis [50]
- Conclusion [15]

Let me know if you have any questions!

The data set housing.txt (and housing.xls) contains information on 2425 single-family detached residential houses sold in a Midwestern city between 2006 and 2010. The variables recorded for each include:

- Order: Observation number
- LotArea: Lot size in square feet
- Alley: Alley access (yes/no)
- LotConfig: Lot configuration (inside lot, corner lot, cul-de-sac, frontage on 2 sides of property, frontage on 3 sides of property)
- OverallQual: Rating of overall material and finish of house (0-10)
- OverallCond: Rating of overall condition of the house (0-10)
- YearBuilt: Original construction date
- YearRemod: Remodel date (this will be same as the construction date if no remodeling or additions)
- Foundation: Type of foundation (brick & tile, cinder block, poured concrete, slab, stone, wood)
- BsmtFin: Square footage of basement that is finished
- BsmtUnf: Square footage of basement that is unfinished
- AC: central air conditioning (yes/no)
- GrLivArea: Above ground living area in square feet
- HalfBath: Number of half bathrooms
- FullBath: Number of full bathrooms
- BedroomAbvGr: Number of bedrooms above ground
- KitchenQual: Kitchen quality (excellent, good, typical/average, fair, poor)
- TotRmsAbvGrd: Total number of rooms above ground (not including bathrooms)
- Fireplaces: Number of fireplaces
- GarageFinish: Interior finish of garage (finished, rough finished, unfinished, NA/no garage)
- GarageCars: car capacity of garage
- GarageArea: size of garage in square feet
- WoodDeckSF: Wood deck area in square feet
- PorchSF: Porch area (included open porches, enclosed porches, three-season porches, and screened porches) in square feet
- YrSold: Year house was sold
- SalePrice: sale price