

University of Central Florida
College of Business

QMB 6912
Capstone Project in Business Analytics

Problem Set #5

Due Date: Sunday, 19 February 2023, at 11:59 PM.

Go to the course Webpage on Webcourses and find the file `Rosen1974.pdf` in the **Files** tab; download and read this document; then write a one-page summary of this paper, following the guidance provided in `ReadPaper.pdf`. As with Lancaster's characteristic theory, this paper can guide your thoughts about how the characteristics of homes might be valued by different types of prospective buyers. Explain how Rosen's hedonic price theory can be used to put theoretical structure on products like used homes.

Analyze the data in `HomeSales.dat` in graphical form. The aim is to visualize some of the numerical findings in Problem Set #4 and to gain further insight into potential modeling approaches in future problem sets. Ideally, you can do this in a way that would communicate to a manager how the data should be analyzed, without getting into the statistical details.

Revisit the covariance of pairs of variables. Create scatterplots for the variables the numeric variables `log_price`, `age`, `floor_space`, `lot_size`, `transit_score`, and `school_score`, and save them to files for further processing in L^AT_EX. Again, you may consider any differences between the categorical variables, if you find that useful. The R command `sploM()` from the package `lattice` is a useful tool for plotting scatterplots for combinations of pairs of variables. The scatterplots can also be depicted together in a matrix with the `pairs()` function in R, as well as many other functions shown in `demo_05` of the course repository.

Continue your analysis by exploring other plots of the data. The objective is to investigate the relationships between the variables to determine the specification for a regression model. Consider plotting the average sale price across levels of categorical variables for each type of buyer. Do the relationships appear to differ between the properties sold to each type of buyer? Suggest any implications for the potential differences in models to predict prices of each type of sale.

Then continue analyzing the data according to other categorizations that you might find relevant for the analysis of this dataset. Present these statistics in graphical form as well. Use your findings in the previous problem sets to guide your analysis. Creativity is rewarded both in terms of grading of this problem set and in the insight gained to guide future analysis.

Interpret the results and conclude with a summary of your findings and your ideas for your regression model.

Prepare and compile your work in \LaTeX and include scripts for any of the calculations in R. In particular, create the following directory structure, separate from your existing work:

- Code/
- Data/
- Figures/
- Tables/
- Text/
- Paper/
- Misc/

In a file called `README.md`, which should also live in the directory containing the above folders, provide the instructions concerning how to run the executable shell script `DoWork.sh` (in the same directory) that will execute the code that produced all of the answers collected and documented in your report, which will live in the subdirectory `Paper/`. In the subdirectory `Code/`, keep the R code; in `Data/` keep the raw data file you downloaded, so that `DoWork.sh` can load it into R for your answers. Keep any \LaTeX scripts for tables in the `Tables/` folder. You may not need to create any figures for this exercise but, if you do, keep any figures you created in the `Figures/` folder. You may put any written text in the `Text/` folder, if not already included in a `tex` file in your `Paper/` folder. Put anything else in the subdirectory `Misc/`. I should then be able to replicate all of your work simply by typing

- `$./DoWork.sh`

on the command line of a terminal window.

To provide you a template, which makes preparation easier for you and grading easier for me, I have placed sample \LaTeX and R code in the GitHub repository for the course: `QMB6912S23`, under my GitHub username `LeeMorinUCF`; pull this repository and use these files a framework within which to create the answers for this problem set. Push the files to a folder on your GitHub repository and I will pull your submissions to my computer for grading.

Be sure to support your calculations with descriptions of what you were trying to do (for example, in comments in your R code as well as in the \LaTeX explanations) because partial credit will be given.

Due Date: Sunday, 19 February 2023, at 11:59 PM.