**Spring 2023** **Lealand Morin**

**University of Central Florida**
**College of Business**

**QMB 6912**
**Capstone Project in Business Analytics**

**Problem Set #7**

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

Revisit the empirical specifications you recommended in Problem Set #6. Now consider some variables as candidates for nonlinear specifications of the form allowed by the Box–Tidwell transformation, which is analogous to a Box–Cox transformation on the dependent variables. Often, continuous variables are candidates for this sort of specification but ordinal variables such as `school_score` and `transit_score` are also worth considering, since these numeric variables also lie within a continuous scale.

First, make sure that you understand the relationships and similarities between these modeling options and introduce this modeling approach to the reader. Compare the method of using the Box–Cox transformation with that of fitting a GLM, and with the Box–Tidwell transformation. Explain the difference between using the Box–Cox transformation and the other two modeling approaches.

Reconsider the continuous or ordinal variables for which you found linear relationships with the dependent variable. Estimate a model with the Box–Tidwell transformation applied to your chosen explanatory variables. The `car` package can be used to estimate a model with a Box–Tidwell transformation. For the ordinal variables, you should also estimate a linear model in which the ordinal variable is included as a categorical variable, or a factor variable as it is called in R, to gain insight into the form of the relationship of the values of these characteristics of homes. Recommend a functional form for a model that may or may not include a nonlinear transformation, depending on the results of your tests based on the Box–Tidwell transformation and the results of your analysis of categorical forms of variables.

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, and in `Figures/` keep any figures you created for your answers. Similarly, keep any LaTeX scripts for tables in the `Tables/` 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 March 2023, at 11:59 PM.**