MATH 62.2 M **DUE: 28 Mar 2023**
Problem Set #3 **J. Briones**

General Instructions: *Answer the following exercises using* `R`*. Show all pertinent and relevant solutions. Answers do not magically appear in papers. So do not give them without solutions or explanations. Submit your problem set (as .pdf file), codes (as .R file), and workspace (as .pdf file) via* `Canvas` *on or before Mar 28, 2023, class time. GOODLUCK!*

1. Consider the data given in `PS3.csv`. Here, the dependent variable $\boldsymbol{y}$ is in the column labelled `y`, while the independent variables $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{12}]$, are in the columns labelled `x01`, `x02`, ..., `x12`. **[20pts]**

   (a) Fit a linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_o + \boldsymbol{\varepsilon}_o$ to the given data.

   (b) Check the fit of the model, and identify which coefficients are significant.

   (c) Perform model diagnostics by identifying which variables should be included, analyzing the resulting residuals, and testing for multicollinearity.

2. In some cases, data transformations are necessary to improve the fit of the linear model. Consider the transformation $\boldsymbol{z} = \ln \boldsymbol{y}$ (that is, $z_i = \ln y_i$) for the data given in `PS3.csv`. **[10pts]**

   (a) Fit a linear model $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta}_l + \boldsymbol{\varepsilon}_l$ to the given data.

   (b) Check the fit of the model, and identify which coefficients are significant.

   (c) Compare this model with the previous linear model. Which of the two models is a better fit for the given data? Justify your answer.