**Assignment 2: Regression Diagnostics**

The purpose of this assignment is to demonstrate what you have learned regarding regression diagnostics and data transformations.

1. Use your research hypotheses and variables from Assignment 1
   1. Add two predictor variables to the model (1 outcome and 4 predictors total)
2. Create three models from the variables
   1. Model 1 should be a multiple regression with all four predictor variables and the outcome variable
   2. Model 2 should be a multiple regression with at least one of the variables transformed (if your model is not violating any assumptions then select the variable that you think most impacts the model)
   3. Model 3 should also be a multiple regression with at least one unusual observation removed that you think may be influential

***Note***: if a predictor is nominal, you will need to dummy code before fitting a regression

1. Evaluate the three models and pick one to write about. You will need to justify the model selected in the writing.

The submitted assignment should consist of:

1. A four page max (not including table) APA formatted paper that includes:
   1. Brief introduction paragraph with identification of the research hypothesis and justification for the study (explain why you think these variables should be related)
   2. A paragraph on the basic descriptive statistics for the five variables (best measure of central tendency and variability and also the skewness statistic with standard error).
   3. A paragraph identifying the three models tested and the diagnostics and model violations (normality, linearity, equal variance, etc.). This paragraph should include the criteria for determining potential influential observations and explanations for any data transformations and the type of transformation used. This paragraph should also include a statistical test (*ΔR*2) among competing models.
   4. A paragraph on the results of the model that best fits the data. You should indicate why you think it best fits the data and include the relevant inferential statistics. Report *and* interpret the regression coefficients, test statistic between the variables, *p*-value, *R*2, confidence intervals, and the SEE in relation to the variables used in the model.
   5. A brief concluding paragraph. Give a brief restatement of your findings (one sentence; no stats) and then interpret your findings. The interpretation is the “why” part of the assignment. Why do you think there is a relationship or lack of a relationship? Briefly explain.
2. A table with all three of your models. Remember that tables follow the text in APA format and should not be imbedded in the text. Make sure to refer to the table in text.
3. An SPSS datafile with only the five variables and diagnostic variables so I can replicate the findings