Assignment 03

Logarithmic Transformations

This assignment is intended to have you build your understanding of using log-transformations in regression. Submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and also should have an appropriate caption. If you are using Markdown, all syntax should be hidden (i.e., not displayed) unless specifically asked for. Any messages or warnings produced from loading packages should also be hidden. This assignment is worth 17 points. (Each question is worth 1 point unless otherwise noted.)

In this assignment, you will use the data from the file *fci-2019.csv* (see the data codebook to explain variation in cost of attending a game.

Model 1: Effect of Sporting League on FCI

1. Create and examine the scatterplot of the relationship between sporting league and FCI. What does the scatterplot suggest about the tenability of homogeneity of variance? Explain.
2. Regress the log-transformed FCI variable (using the natural logarithm) on sporting league. Use NBA as the reference group in this model (Model 1). Report and interpret the NHL coefficient from the fitted model.
3. Back-transform all the estimated coefficients from Model 1. Report and interpret the back-transformed NHL coefficient.

Model 2: Effect of Payroll on Log-Transformed FCI

4. Create and examine the scatterplot of log-transformed FCI versus payroll. Include the loess smoother in this plot. Does this plot suggest any nonlinearity in the relationship between payroll and log-transformed FCI that we need to address?
5. Fit a series of models that allow you to examine the non-linearity in the relationship. This series of models should include: (1) a model that posits a linear effect of payroll; (2) a model that posits a quadratic effect of payroll; and (3) a model that posits a cubic effect of payroll. (Do not include sporting league in these models.) Use the test for nested models (the ΔF test) to examine whether the any higher-order terms are necessary. Provide the output of the test, and comment on whether the results suggest whether the higher-order terms tested are warranted. (Henceforth the model you adopt from this analysis will be referred to as Model 2.)
6. Create and examine the scatterplot of the standardized residuals versus the fitted values (based on the Model 2). Add the loess smoother to this plot. Did this transformation "fix" the non-linearity? Include the plot as evidence of your response.

Model 3 and 4: Effects of Sporting League and Player Salary (Main Effects)

7. Fit the model using sporting league, and all adopted effects of payroll (based on result in Question 5) to predict variation in log-transformed FCI (Model 3). Again, use NBA as the reference group in this model. Report and interpret the back-transformed NHL coefficient from the fitted model.
8. Based on the inferntial results from the coefficient-level output, can we drop any terms to make this model more parsimonious? Explain.
9. Fit the more parsimonious model you identified in Question 8. Henceforth this will be referred to as Model 4.

Model 5: Effects of Sporting League and Player Salary (Interaction Effect)

8. Fit the model using sporting league, and any adopted effects of payroll (based on result in Question 9) to predict variation in log-transformed FCI. Also include the interaction between sporting league and payroll in this model (Model 5). Again, use NBA as the reference group in this model. Remember that to include interaction effects when you have a categorical variable represented by more than one dummy variable, you need to create and include ALL product terms. (Hint: Model 5 will include seven predictors.) Use the test for nested models (the ΔF test) to test the necessity of including the interaction components by comparing the main-effects model (Model 4) and the interaction model (Model 5). Provide the output of the test, and explain which model should be adopted based on the results of this test.

Adopting a Candidate Model

9. Examine the structure and formatting of Table 9 at http://zief0002.github.io/epsy-8251/misc/creating-tables/creating-tables.html. Mimic the format and structure of this table to create a table to present the numerical information from the five models you fitted in this assignment. Make sure the table you create also has an appropriate caption. If the table is too wide, change the page orientation in your word processing program to "Landscape", rather than changing the size of the font. (Note: Only this table should be presented in landscape orientation…not your entire assignment!) **(2pts.)**

10. Based on all of the analyses you have done, which of the five candidate models will you adopt as your "final" model? Write the fitted equation for the adopted candidate model using Equation Editor (or some other program that correctly typesets mathematical expressions).

11. Examine the residuals from the adopted model from Question 8. Are the assumptions for the model satisfied? Explain. (Include any plots that help make your case.) **(2pts.)**

12. Create a publication quality plot that displays the fitted curves from your adopted candidate model. Display four separate lines to show the effect of sporting league. The four lines should be displayed using different linetypes or colors (or both), or in different panels, so that they can be easily differentiated in the plot. (Note: Make sure that you back-transform any log-transformed variables when you create this plot.) **(2pts.)**

13. Use the plot to help describe/interpret the effect of sporting league on FCI.

14. Use the plot to help describe/interpret the effect of payroll on FCI.