

Assignment 02

Log-Transformations

This assignment is intended to have you build your understanding of using log-transformations in regression. Turn in a printed document that includes your responses to each of the questions on this assignment. Please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document.
- 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.

In this assignment, you will use the data from the file *fci-2015.csv* (see the [data codebook](#) to explain variation in cost of going to a professional sporting event. Each question is worth one point unless otherwise noted. The entire assignment is worth 15 points.

Model 1: League

1. Create and examine the scatterplot of the relationship between sporting league and FCI. Since league is a categorical predictor, we do not have to worry about linearity. However, homogeneity of variance is still an issue. What does the scatterplot suggest about the tenability of homogeneity of variance? Explain.

Create the variable `lfci` by log-transforming the `fci` variable using the natural logarithm. Also create a dummy variable for each sporting league; four total. Fit the model using league to predict variation in FCI. Make NBA the reference group in this model (Model 1).

2. Report and interpret the NHL coefficient from the fitted model.
3. Back-transform all the fitted model's coefficients. Report and interpret the back-transformed NHL coefficient.

Model 2: Player Salary

4. Create and examine the scatterplot of the relationship between average player salary and log-transformed FCI. Add the loess smoother to this plot. This plot suggests a non-linear relationship. Based on the shape of this relationship, suggest a power transformation (given that you have already log-transformed the outcome) that might help "fix" the non-linearity. Explain. (No need to include the plot here. Instead, describe the non-linearity in your response.)
5. Create and examine the scatterplot of the relationship between transformed average player salary (based on your suggestion in Question 4) and log-transformed FCI. Add the loess smoother to this plot. Did this transformation "fix" the non-linearity? Include the plot as evidence of your response.
6. Fit both the linear and quadratic models using the transformed transformed average player salary (based on your suggestion in Question 4) to predict variation in the log-transformed FCI (do not include sporting league in these models). Use the test for nested models (the ΔF test) to examine whether the quadratic term is necessary. Provide the output of the test, and comment on whether the results suggest the quadratic term is warranted.

Model 3: League and Stadium Capacity (Main Effects)

7. Fit the model using sporting league, and all adopted effects of transformed average player salary (based on result in Question 6) to predict variation in log-transformed FCI. Again, use NBA as the reference group in this model. Report and interpret the back-transformed NHL coefficient from the fitted model.

Models 4 and 5: League and Stadium Capacity (Interaction Effects)

Fit two interaction models using sporting league, and transformed average player salary to predict variation in log-transformed FCI. Again, use NBA as the reference group in this model. The first interaction model will include interaction effects with both the linear and quadratic effects of transformed average player salary. The second interaction model will include interaction effects with only the linear effects of transformed average player salary (but will still include the quadratic main-effect of transformed average player salary). 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 4 will include 11 predictors and Model 5 will include eight predictors.)

8. Use the test for nested models (the ΔF test) to test the necessity of including the different interaction components by comparing three models: the main-effects model (Model 3); the interaction model with linear interaction effects (Model 4); and the interaction model with linear and quadratic interaction effects (Model 5). Provide the output of the test, and explain which of the three models should be adopted based on the results of this test.
9. 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.)

Summarizing and Interpreting the Analytic Results

10. Create a table to display the five fitted models' regression results. This table should be similar in structure to Table 2 from Garcia, D. R., McIlroy, L., & Barber, R. T. (2008). Starting Behind: A Comparative Analysis of the Academic Standing of Students Entering Charter Schools. *Social Science Quarterly*, 89(1), 199–216. Re-create this table as closely as you can (including the caption) using the output from the models you fitted above. **(2pts.)**
11. Write the equation for the adopted fitted model using Equation Editor (or some other program that correctly typesets mathematical expressions).
12. Using the `ggplot()` function, create a plot of the fitted model that allows you to interpret the effects of sporting league and average player salary on cost of attending a game. (Note: Make sure that you back-transform any log-transformed variables when you create this plot.)
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 stadium capacity on FCI.