

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

In this section you will be exploring the predictor of sporting league and its relationship to FCI.

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 `lfcf` 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

In this section you will be exploring the predictor of player salary and its relationship to FCI.

4. Create and examine the scatterplot of log-transformed FCI versus average player salary. Add the loess smoother to this plot. This plot suggests a non-linear relationship. As such, we will need to transform player salary. Based on the trend you see in the loess smoother, explain why using a power transformation of player salary (e.g., quadratic, cubic) is more appropriate than using a log-transformation. (No need to include the plot here. Instead, describe the non-linearity in your response.)

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 player salary; (2) a model that posits a quadratic effect of player salary; and (3) a model that posits a cubic effect of player salary. (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 adopted in Question 5). Add the loess smoother to this plot. Did this transformation “fix” the non-linearity? Include the plot as evidence of your response.

Model 3: League and Player Salary (Main Effects)

7. Fit the model using sporting league, and all adopted effects of transformed average player salary (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.

Models 4, 5, and 6: League and Player Salary (Interaction Effects)

Fit a series of interaction models using the effect of sporting league and the effects of average player salary to predict variation in log-transformed FCI.

- The first interaction model (Model 4) will include interaction effects between sporting league and only the linear effect of average player salary (but will still include the quadratic and cubic main-effect of transformed average player salary).
- The second interaction model (Model 5) will include interaction effects between sporting league and only the linear and quadratic effects of average player salary (but will still include the cubic main-effect of transformed average player salary).
- The third interaction model (Model 6) will include interaction effects between sporting league and the linear, quadratic, and cubic effects of 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 6 will include 16 predictors.)

8. Use the test for nested models (the ΔF test) to test the necessity of including the different interaction components by comparing four models: the main-effects model fitted in Question 7 (Model 3); the interaction model with linear interaction effects (Model 4); the interaction model with linear and quadratic interaction effects (Model 5); and the interaction model with linear, quadratic, and cubic interaction effects (Model 6). Provide the output of the test, and explain which of the four 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 regression results from a series of five fitted models: Models 1, 2, 3, 4, and the adopted interaction model (either Model 4, 5, or 6). 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 fitted equation for the “final” adopted model using Equation Editor (or some other program that correctly typesets mathematical expressions).
12. Using the `ggplot()` function, create a plot of the fitted equation for the “final” adopted 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 average player salary on FCI.