Assignment 09

Transformations

Each season, Team Marketing Report (TMR) computes the cost of taking a family of four to a professional sports contest for each of the major sporting leagues. Costs are determined by telephone calls with representatives of the teams, venues and concessionaires. Identical questions were asked in all interviews. Prices for Canadian teams were converted to US dollars and comparison prices were converted using a recent exchange rate.

In this assignment, you are going to focus on whether or not sporting league and stadium capacity predicts variation in the cost of going to a game (FCI). Use the data from the file *FCI-2014.csv* to answer the questions on this assignment. Each question is worth one point unless otherwise noted. The entire assignment is worth 18 points.

The file *FCI-2014.csv* contains these data from the 2014 (or 2014/2015) season for 122 professional sports teams across the United States along with data on the stadium these teams play in. The variables are:

- `team`: Name of professional sports team
- `fci`: Fan Cost Index (FCI). The FCI is a summary of what it costs to take a family of four to a game. It comprises the prices of four (4) adult average-price tickets, two (2) small draft beers, four (4) small soft drinks, four (4) regular-size hot dogs, parking for one (1) car, two (2) game programs and two (2) least expensive, adult-size adjustable caps.
- `league`: Major sporting league the team plays in (MLB = Major Lague Baseball; NBA = National Basketball Association; NFL = National Football League; NHL = National Hockey League)
- `stadium`: Team's home stadium
- `capacity`: Stadium's seating capacity

Data Preparation

To begin the analysis, create a predictor called `cap` that indicates the capacity of each stadium in thousands. To do this, divide the `capacity` variable by 1000. This type of transformation will keep scientific notation out of any regression coefficients we obtain, but will not change any inferences about the predictor (*p*-values will stay the same). This variable (not `capacity`) should be used in all analyses in this section referring to the stadium capacity.

Also create a dummy variable for each sporting league; four total.

Initial Exploration

1. Create a density plot of the distribution of FCI. Does this plot suggest a transformation may be needed for this variable? Explain.

Relationship with League

2. 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.

3. Create the variable `lfci` based on the natural log of `fci`. Fit the model using league to predict variation in FCI. Interpret each of the coefficients from the fitted model. **(2pts.)**

Relationship with Stadium Capacity

4. Create and examine the scatterplot of the relationship between stadium capacity and the log-transformed FCI. Add the loess smoother to this plot. What does the scatterplot and loess line suggest about the relationship between stadium capacity and cost of attending a game? Explain.

5. Fit both the linear and quadratic models using stadium capacity to predict variation in the log-transformed FCI. 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.

6. Based on the model adopted from the ΔF test, describe the relationship between stadium capacity and cost of attending a game.

ANCOVA Model

7. Fit both the model using sporting league, and all adopted effects of stadium capacity to predict variation in log-transformed FCI. Examine the residuals from the fitted model. Are the assumptions satisfied? Explain. **(2pts.)**

Summarizing and Interpreting the Model Coefficients

11. Write the equation for the 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 stadium capacity on cost of attending a game. (Note: Make sure that you back-transform any log-transformed variables when you create this plot.) **(2pts.)**

13. Create a table (suitable for publication) that presents each of the possible pairwise contrasts examining differences between the four sporting leagues. Include the unadjusted p-values, and the Benjamani–Hochberg adjusted p-values for the controlled differences. (Note: To obtain all of these, you may need to fit additional models.) **(2pts.)**

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

15. Use the plot to help describe/interpret the effect of stadium capacity on FCI.