Assignment 10

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.

The file *FCI-2015.csv* contains these data from the 2015 (or 2015/2016) 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
- `yearOpened`: Year the stadium was opened
- `capacity`: Stadium's seating capacity

In this assignment, you are going to focus on whether or not the capacity of a stadium predicts variation in the cost of going to a game (FCI). Use these data to answer each of the following questions. Each question is worth one point unless otherwise noted. The entire assignment is worth 15 points.

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.

Initial Exploration

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

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

3. Create and examine the scatterplot of the relationship between stadium capacity and FCI. Add the loess smoother to this plot.

4. 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 FCI. What do these models suggest about the relationship between stadium capacity and cost of attending a game? Explain.

6. Examine the residuals from the quadratic model. Do these plots suggest any further transformations are needed? Explain.

Log Transformation of the Predictor

Using the natural logarithm, create a variable, `lcap` that contains the values of the log transformed `cap` predictor. This variable (not `cap`) should be used in all analyses in this section referring to the stadium capacity.

7. Create and examine the scatterplot of the relationship between stadium capacity and FCI. Add the loess smoother to this plot.

8. What does the scatterplot and loess line suggest about the relationship between the stadium capacity and cost of attending a game? Explain.

9. Fit both the linear and quadratic models using stadium capacity to predict variation in FCI. What do these models suggest about the relationship between stadium capacity and cost of attending a game? Explain.

10. Examine the residuals from the quadratic model. Do these plots suggest any further transformations are needed? Explain.

Summarizing and Interpreting the Model Coefficients

11. Write the fitted regression equation based on the `summary()` output for the linear model from the previous section. Write the regression equation using Equation Editor (or some other program that correctly types mathematical expressions).

12. Interpret the effect of stadium capacity from the linear model from the previous section.

13. Write the fitted regression equation based on the `summary()` output for the quadratic model from the previous section. Write the regression equation using Equation Editor (or some other program that correctly types mathematical expressions).

14. Using the `ggplot()` function, create a plot of the fitted model that allows you to interpret the quadratic effect between stadium capacity and cost of attending a game. Make sure that you also back-transform any log-transformed variables when you create this plot. **(2pts.)**

15. Using the plot you just created, interpret the quadratic effect between stadium capacity and cost of attending a game.