

Assignment 11

Put It All Together

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 League 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 predicting variation in the cost of going to a game (FCI) using several predictors. Use these data to answer each of the following questions.

Since this assignment includes material from throughout the course, it provides some insight into whether you have synthesized and can use the material from the course. Because of this, rather than assigning points for each question, I am going to judge the assignment holistically. If my judgment is that you have been able to mostly do this, I will replace your lowest assignment score with a perfect score for that assignment. If my evaluation is that there are major pieces of knowledge missing, I will not replace your lowest score. There will be no partial credit given.

Is there an effect of age of the Stadium?

To begin the analysis, create a predictor called **age** that indicates the current age (in years) of each stadium. Mimicing the analysis you carried out in Assignment 10, examine the effect of age of stadium on cost of attending a game. As a reminder here is how we did that for stadium capacity.

- Examine marginal distribution of the predictor
- Examine the relationship between the predictor and the outcome (un-transformed)
- Fit models to examine any polynomial effects (un-transformed)
- Examine residuals from the “best” fitting polynomial model (un-transformed)
- Make any necessary transformations
- Examine the relationship between the predictor and the outcome (transformed)
- Fit models to examine any polynomial effects (transformed)
- Examine residuals from the “best” fitting polynomial model (transformed)
- Adopt a model

1. Write the fitted regression equation based on the `summary()` output for the model you adopt. Write the regression equation using Equation Editor (or some other program that correctly types mathematical expressions).
2. Is there an effect of the age of stadium on cost of attending a game? Explain. If there is an effect, interpret it. Include any plots that you create to help you interpret these effects. (Be sure to back-transform any transformed variables to ease reader interpretation.)

Is there an effect of league?

Fit a model that uses league to predict variation in FCI.

3. Write the fitted regression equation based on the `summary()` output for the model you adopt. Write the regression equation using Equation Editor (or some other program that correctly types mathematical expressions).
4. Is there an effect of the age of stadium on cost of attending a game? Explain. If there is an effect, interpret it. Include any plots that you create to help you interpret these effects. (Be sure to back-transform any transformed variables to ease reader interpretation.)

Multiple Predictors (Main Effects)

You now have fitted and adopted structures for each of the predictors (1) stadium capacity, (2) age of stadium, and (3) league. Some of these include quadratic effects, and some may be log-transformed as well. Fit a model that includes all of these pertinent effects to predict variation in FCI.

5. Write the fitted regression equation based on the `summary()` output for this main-effects model. Write the regression equation using Equation Editor (or some other program that correctly types mathematical expressions).
6. Are there any effects that can be dropped from this model to make it more parsimonious? Explain.

Drop any effects you identified in the previous question (one at a time) and re-fit the model. Continue this until you identify a model that you are happy with. Adopt this model as your “final” main-effects model.

7. Examine the residuals from this model. Comment on whether each of the model’s assumptions are satisfied.
8. Using the `ggplot()` function, create a plot of the “final” main-effects model that you adopted. Be sure to appropriately differentiate between lines that you include in the plot (e.g., do not differentiate by color unless you plan to print in color). Include the plot (with an appropriate caption) in your word-processed document.

Prediction

9. Compute the predicted FCI for the 2016 (or 2016/2017) seasons for the Minnesota Twins, the Minnesota Vikings, the Minnesota Timberwolves, and the Minnesota Wild. Also compute the prediction intervals for each of these predictions. Include all of this in a table.

Adjusted League Differences

10. Create a table (suitable for publication) that presents each of the possible pairwise contrasts (hypothesis) to examine potential league differences in FCI. This table should include each of (1) the contrasts, (2) the adjusted mean differences, and (3) the Benjamani–Hochberg adjusted p -values for the adjusted differences. (Note: To obtain all of these, you may need to fit additional models.)

Table of Regression Results

11. Create a table to display your 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 following models.

- (1) Adopted model for the effect of stadium capacity;
- (2) Adopted model for the effect of stadium age;
- (3) Adopted model for the effect of league;
- (4) Initial main-effects model; and
- (5) Adopted “final” main-effects model