

# Assignment 01

## *R Markdown*

Create an R Markdown document to respond to each of the questions below. Turn in a printed version of the knitted file. This assignment is worth 10 points. 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. Learn how to do this in a code chunk using **knitr syntax**.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Markdown's display equations. See [here](#) for some examples of how mathematics can be typeset in R 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.

For each question, specify the question number using a level-2 (or smaller) header. Submit a printed version of your output in class the day it is due.

---

For this assignment, you will be using data from the file *fci-2015.csv*. These data include information from the 2015 (or 2015/2016) season for 122 professional sports teams across the United States and the stadiums 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 differences in sporting league.

- 
- **Question 1:** Read in the data and show the results using the `head()` function. All syntax for these commands should be displayed, as should the output.
  - **Question 2:** Create dummy variables for each of the leagues represented in the data. Display the four dummy variables for the first six cases in the data. Do not display any syntax.
  - **Question 3:** Use the `lm()` function to fit the linear model regressing FCI on your league dummies. In this model, use MLB as the reference group. Use the `tidy()` function from the *broom* package to display the model's coefficient-level output. (Reminder: Do not display any syntax, only the output.)
  - **Question 4:** Use an unordered list to provide an interpretation of each regression coefficient (including the intercept), one per list item.

- **Question 5:** Create a well-formatted table of the model's coefficient `tidy()` output using either the `kable()` function (in the *knitr* package) or the `xtable()` function (in the *xtable* package). In the final outputted table, the five column names should be "Predictor", "B", "SE", "t", and "p", respectively. All numerical output should be rounded to two decimal places, except the *p*-values, which should be rounded to three decimal places. Also add an appropriate caption (the caption does not have to follow APA formatting). One way to create such a table is to use *dplyr* functionality to mutate (e.g., re-name, round) the output from `tidy()`, and then use `kable()` or `xtable()` to produce the table in your Markdown document. Another way is to use the arguments in `kable()` or `xtable()` to do the rounding and renaming.
- **Question 6:** Create a scatterplot of the FCI (outcome) versus league. In this plot, color the observations by league, and also display each of the four horizontal lines that are defined from the fitted regression. These lines should also be colored so they correspond to the appropriate league. Display the plot, but not the syntax. The plot should be centered on the slide and should have an appropriate caption. Feel free to use the captioning options in the code chunk; the caption does not have to follow APA formatting. Adjust the aspect ratio of the plot so that it looks good, and change the values of the output width/height to reduce the size of the plot from the default values.
- **Question 7:** Use a display equation to write the equation for the underlying regression model (including error) using greek letters, subscripts, and variable names. Also write the equation for the fitted least squares regression equation based on the output from `lm()`. Type these two equations in the same display equation, each on a separate line of text in your slide, and align the equals signs. (Hint: Google "aligning math latex".)
- **Question 8:** Write the following sentence: "The estimated coefficient ( $\beta_{\text{NHL}}$ ) is  $x$ . In this sentence, use an inline code chunk to replace  $x$  with the value for the fitted coefficient from the fitted equation. In this code chunk, do not just write in the value for the coefficient, but use syntax to pull it from the `tidy()` output.
- **Question 9:** Compute the sum of squared residuals for the fitted regression. Although you can use the `anova()` function to check your work, compute this value by actually using R syntax to carry out the computation,  $\sum (y_i - \hat{y}_i)^2$ . Show the syntax you used in your document.
- **Question 10:** Write a sentence that includes two references in an inline citation. This should also generate the actual references when you knit your document. One of the two references should be the Fox textbook. The other should be a journal article of your choice. You can choose the topic of the sentence and how the two references are used in the citation. (Note the references do not actually have to pertain to what is written in the sentence. This is just an exercise in using the bibliography tools in Markdown.) The two references can be included in the document's YAML (be careful of the indentation) *or* drawn from an external BIB file. Specify an appropriate CSL file so that the references and citations are in APA format. (If you want to use a citation style that is different from APA, say for a specific journal, use the appropriate CSL file, and indicate that on the assignment.)