Assignment 02

R Markdown

In this assignment you will use R Markdown to produce a slide deck that includes your responses to each of the questions on this assignment. Choose one of the following formats:

- ioslides
- Slidy
- Reveal.js

Note that the `revealjs_presentation` format is not contained within the rmarkdown package but rather in a separate **revealjs** package. See the Rmarkdown documentation for more details about selecting the reveal.js format When you finish the assignment, you will send (1) your RMD document, and (2) your slidedeck to the instructor and TA via email.

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 included should be included in an R Markdown code chunk and be appropriately commented. Follow the Data Camp Style Guide (http://docs.datacamp.com/teach/style-guide.html as close as you can.

This assignment is worth 18 points. Each question is worth 2 points unless otherwise noted.

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 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 predicting variation in the cost of going to a game (FCI) using differences in sporting league.

R Code Chunks

- **Slide 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.

- **Slide 2:** Create dummy variables for each of the leagues represented in the data. Then, use the `lm()` function to fit the linear model regressing FCI on your league dummies. Display the syntax for the commands for creating the dummy variables, and fitting the model, but not the output. Be sure to also comment your syntax. On the same slide, identify the reference category.

- **Slide 3:** Fit the exact same model you fitted in Slide 2, but this time display the output, but not the syntax.

List

- **Slide 4:** Use an unordered list to give the interpretations of each regression coefficient, one per list item.

Table

- **Slide 5:** Create a table using Markdown syntax that includes the values for the (1) estimated regression coefficients, standard errors, and p-values from the fitted model. Note the table does not have to be formatted to correspond to APA, but it should have an appropriate caption.

Plot

- **Slide 6:** Use `ggplot()` to create a scatterplot of the FCI (outcome) versus league. In this plot, color the observations by league, and also show each of the four horixontal 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. The figure should be 500 pixels wide x 400 pixels tall on the slide.

Display Equation

- **Slide 7:** Use a display equation to write the underlying regression model (including error) using greek letters, subscripts, and variable names. Also write the fitted 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".)

Inline Equation and Inline Code Chunk

- **Slide 8:** Use an inline equation to write the following sentence: "The estimated coefficient (β_{NHL}) is x. In this sentence, replace x with the value for the fitted coeffcient from the fitted equation. Do not just write in the value for the coefficient in the sentence, but pull it from the `coef()` or `summary()`output, and add it using an inline code chunk.

Citations and References

- **Slide 9:** Write a sentence that includes two references in an inline citation. This should also generate the actual references when you knit your slides. 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 include in the document's YAML (be careful of the indentation) *or* drawn from an external BIB file. (If you use a BIB file, also email that when you turn in your assignment.) 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 slide.)