ASSIGNMENT #5

EPsy 8252

This assignment introduces you to using R Markdown. You can read about R Markdown at http://rmarkdown.rstudio.com/. You can read more about the different options for customizing R code chunks at http://yihui.name/knitr/options.

Use the data *Sex-Discrimination.csv* to complete this assignment. All R syntax should be included in a code chunk. You will submit your responses to each of the questions below in a slide format of your choice (ioslides, Slidy, or Beamer). Each question should be a separate slide. When you are finished, you can send your RMD document and your slidedeck to the instructor via email.

There are 18 points possible for the assignment. Each question is worth two points.

Using R Code Chunks

- **Slide 1:** Read in the data and create two effects-coded vectors for the sex variable (sexM, and sexF) using the coefficients $[-1, 1]$. Show the results using the head() function. All syntax for these commands should be displayed, as should the output.

- **Slide 2:** Using the lm() function, fit the linear model: salary ~ 1 + sexM + sexF. Display the syntax for this command, but not the output. On the same slide, identify which coefficients would be estimated and which would not. Explain why some of the coefficients would not be estimated.

- **Slide 3:** Using the lm() function, fit the linear model: salary ~ 1 + sexF. Display the output for this command, but not the syntax. On the same slide, using bullet points, interpret the intercept and slope coefficients.

- **Slide 4:** Create a table using Markdown syntax of the estimated regression coefficients, standard errors, and p-values for the model fitted in the previous slide. Do not just write in the values for these estimates in the table cells, but pull them from the regression summary() output and put them in the table cells using Markdown syntax to write inline R code.

- **Slide 5 and 6:** Use ggplot() to create a plot of salary (outcome) versus education. In this plot, color the observations by sex, and also show the regression lines for both sexes. On Slide 5 display the syntax for this command, but not the output. This syntax should be indented similar to other ggplot() examples in the script files for the course. On Slide 6, display the output for this command, but not the syntax. The plot should be centered on the slide and should have an appropriate caption. The figure should be 1000 pixels wide x 800 pixels tall on the slide.

Equations

- **Slide 7:** Use a display equation to write the full regression equation (including error) using greek letters, subscripts, and generic variable names (X and Y) for the model associated with fitting a quadratic polynomial model. See this page, http://en.wikibooks.org/wiki/LaTeX/Mathematics for the syntax needed for different mathematical symbols.

- **Slide 8:** Use an inline equation to write the following sentence, "The estimated coefficient ($\hat{\beta}_{\text{sexF}}$) is x. In this sentence, replace x with the value for the fitted coeffcient from the regression `summary()` output from Slide 3. This value should be written using Markdown syntax to pull it from the regression `summary()` output.

References and Bibliography

- **Slide 9:** Add two references to your document. One should be an article of your choice, and the other should be the Gelman and Hill textbook. These references can be added inline using YAML (be careful of the indentation) or using an external BIB file. Write a sentence that includes both references in a citation. 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.) Specify an appropriate CSL file so that the references and citations are in APA format.