Assignment 1

3 Possible Points

This assignment's due date is Friday, June 16th by 8:00 PM. The grading policy for late assignments can be found in the syllabus. You are strongly encouraged to work in groups of up to 3 students, but each of you must submit his/her own version of the assignment. Please put the names of your group members in a comment in the program.

Write a single text file that contains all of your work for this assignment. Name it ps1_[yourlastname].txt - for example, mine would be ps1_hoff.txt The text file should submit all of the commands necessary to answer the questions. Please put the answers to the **bold questions** in a comment line in your text file, even if you think the answer is readily apparent.

This assignment is designed to introduce you to statistical analysis and R. You will input data into RStudio as well as randomly generate data.

NBA Data

exper = years of NBA experience
wage = salary (in millions of dollars)

Data Basics

- Download the "nbasalary"" data from Springboard into a directory of your choosing.

- Using the `read.csv()` function, load the data into RStudio's workspace. Now use the `View()` function to make sure that the data has loaded correctly. You should have 294 observations with 22 variables.

- Try finding the mean and standard deviation of the variable age, using the functions `mean()` and `sd()`, respectively.

- **Why does RStudio return "NA" as the mean and standard deviation of the variable *age*?**

- Now use the option `na.rm=TRUE` to complete the next step.

- **What is the mean and standard deviation of the variable *age*?**

Plotting the Data

- **Make a scatter plot with *exper* on the x-axis and *wage* on the y-axis. Make the graph look good (i.e. at the very least - titles).**

- **Describe what you see with respect to the relationship between experience and wages.**

Statistical Analysis

- Use the `cor()` function to calculate the Pearson correlation coefficient between *exper* and *wage*. Hint: set `use="complete.obs"` and `method="pearson"`.
- **Interpret the result of the correlation.**

- Estimate the following equation using OLS regression analysis:

$$wage = \beta_0 + \beta_1 exper_i + \epsilon_i$$

  This can be done using R's `lm()` function. Use the `summary()` function to create the regression output.

- **Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$.** Note: be careful in using causal language.
  - **Graduate Students Only: Why should you be cautious in using causal language in this scenario?**

- **Are your statistical analysis results in line with your visual analysis?**

Randomly Generated Data

Creating the Data

- Set the randomizing algorithm in RStudio using: `set.seed(1234)`. This allows you to generate the same random numbers each time.

- Create a random variable X with 100 observations that is uniformly distributed from -2 to 2 (use the `runif()` function).

- Create a random error term, U, that is Normally distibruted with mean zero and standard deviation of one (use the `rnorm()` function).

- Create a random variable $Y = 2 - (1/2)X^2 + U$

Statistical Analysis

- Use the `cor()` function to find the correlation between X and Y.

- **What do the results of the correlation coefficient suggest? Is the correlation coefficient a good measure of the relationship between X and Y? Provide evidence.**

- Use regression analysis to better measure the relationship between X (including any transformation of X) and Y.

- **Is the regression analysis better or worse than the simple correlation? Why or why not?**