This Methods and Data Analysis assignment involves linear regression with one predictor. The data can be found in the Assignments folder on Sakai. Useful R commands are in the R scripts on Sakai and at the end of this document. Please turn in work using R Markdown or another word processor.

Due Date: Thursday, September 6.

1 Methods and Data Analysis

1. **Old Faithful**. In your answer, make sure you include the output from the regression model including the estimated intercept, slope, residual standard error, and R^2. Also include the 95% confidence interval for the slope, and explain what the interval reveals about the relationship between duration and waiting time. Describe in a few sentences whether or not you think the regression assumptions are plausible based on residual plots (you don't need to include the plots). For the part of the question about prediction bands, construct a 95% prediction interval for the waiting time until the next eruption if the duration of the previous one was 4 minutes. This is the only prediction interval you need to report.

2. **Respiratory Rates for Children**. In addition to the plot, include the output of the regression that predicts (possibly transformed) respiratory rates from (possibly transformed) age, as well as evidence that the model fits the assumptions reasonably well. Demonstrate the usefulness of the model by providing 95% prediction intervals for the rate for three individual children: a 1 month old, an 18 months old, and a 29 months old.

3. **The Dramatic U.S. Presidential Election of 2000**. Include the output from the final regression model that you used, as well as evidence that the model fits the assumptions reasonably well. Also include the 95% prediction interval based on your final model.

4. This is not an assignment that you need to turn in. In the R Resources folder on our Sakai site, you will see a link to "R for data science: the tidyverse". Over the next 10 days, work through sections 1, 2, and 3. The third section is the most important one. It covers "ggplot" which is a set of extremely useful graphical display commands. We will use ggplot once we get into multiple linear regression. Note that you should install the "tidyverse" package to use ggplot. You are welcome to use ggplot for this assignment, but it is not required.

2 Useful R commands

You might want to make a graph with multiple y values for each x value, e.g., predicted y values with upper and lower limits for a particular x. Here is one quick way to do it.

Suppose that you want to plot two sets of points, which we will call "upper" and "lower", versus an x on the same graph. First, make a new variable (call it something like "allypoints") that combines the y values for all the points you want to plot. For example, you can type

allypoints = c(upper, lower)

Now do the same for all the x values, which essentially makes two copies of the x values.

allxpoints = c(x, x)

Now, just use the plot command:

plot(y = allypoints, x = allxpoints)

If you want to add one or more lines to this plot, use abline(a, b) where you enter the intercept for a and the slope for b.