Homework 3: Correlation and Regression

Answer the following questions in a .pdf or .docx, explaining all of your answers and putting any tables and figures in the document as necessary. When data is called for to answer applied questions, I will provide it in bblearn. Turn in your R code that created all of the tables and figures separately, and be sure that it runs from source in such a way that it loads the data and performs all the tests without me fiddling with it. Make sure to document your R source code using # comments if you want partial credit.

W will be using some over-time data from on US productivity and compensation overall and in non-farm sectors. I have provided a .txt codebook that briefly describes the variables in the data.

1. What is the correlation between compensation in the business sector and productivity in the business sector? Subjectively, do you consider this high or low and how can you tell?

2. Run and interpret a bivariate regression predicting all businesses compensation with all businesses productivity. What do you conclude about their relationship? What specific predictions does this model make for an increase in productivity index of 5 points? Of 12 points?

3. Plot the regression line for this regression with the sample's data points. Then plot the residuals of the regression regression against the actual predicted compensation values.

4. Examining the residuals, which five year observations does this productivity model predict compensation data most poorly? (by default the R wrapper Modern Dive uses gives "ID" values that reflect the row number in the data you ran the regression with)

5. What percentage of variance does our productivity model explain in compensation? With respect to the actual variables of productivity and compensation, what fundamental substantive post-WWII USA economic truth is this whole exercise expressing?

6. Now, run a regression predicting non-farm wage compensation with non-farm productivity. What is our estimate of the relationship between the two?

7. Comparing the coefficients between the two models of compensation and productivity indices, do non-farm or all businesses data predict a closer to 1 to 1 relationship between productivity and compensation?