**Unit 5 HW**

1. **Simply Answer Question 25 on pg. 147** **from the Statistical Sleuth (read it!):**

*Plot the raw data, and also plot the data after a log transform. After a log transform, do the data satisfy the assumptions better?* The data is in ex0525.csv or ex0525.xlsx. Perform this analysis in SAS. [Depending on where you find the data set, you may see the value **<<12**. Note that **<<12 = 12**.]

Regardless of whether the assumptions of the original data or log transformed data are met, please include a **complete analysis** on the **log transformed** data.

1. State the Problem.
2. Address the assumptions. Comment on each assumption. (Use the visual test, as the Brown-Forsythe test will be overpowered due to the large sample size. This simply means that it is able to detect very small effect sizes—here, differences in standard deviations—which may not be big enough to practically affect the test.) Comment on your thoughts of the assumptions, but, in the end, assume there is not enough visual evidence to suggest the standard deviations of the log transformed data are different.
3. Conduct the Test. (An example is in the UNIT 5 PowerPoint.)
4. Write a conclusion. (An example is in the UNIT 5 PowerPoint.)
5. State the Scope. (Can we generalize to the entire population or just the sample that was taken? Is there a causal relationship present?)

ADDITIONAL THINGS TO INCLUDE (for the logged data):

1. Please also identify R2
2. Also specify the mean square error and how many degrees of freedom were used to estimate it.
3. Provide the code to perform the ANOVA in R and a screen shot of the output.

*Looking to the future! This is not an additional problem. Just FYI: The next step will be to look at these pairwise if we reject the Ho to discover WHICH pairs have evidence of different means / medians.*

2.Use an extra sum of squares F-test (BYOA: Build Your Own ANOVA!) to use all the data (to increase the degrees of freedom and thus the power of the test!) to compare only the bachelor’s degree group (16) income to the more than bachelor’s degree group (>16) income. Show your final ANOVA table and your 6-step complete analysis. You will need to assume that the standard deviations of the log-transformed data are again equal to proceed here. A two-sample t-test between these two groups (assuming equal standard deviations on logged data) yields a p-value of **.1648** (try it!), but it only uses 778 degrees of freedom (from a pooled t-test). Make note again of how many degrees of freedom were used to estimate the pooled standard deviation in your extra sum of squares test. You may use SAS or R.

3. Now, suppose that you cannot assume the standard deviations are the same (for both the original or log transformed data). Conduct another complete analysis of the question in Chapter 5, problem 25 in Statistical Sleuth. Answer the question, “How strong is the evidence that at least one of the five population distributions (corresponding to the different years of education) is different from the others?” This question should be answered in at least 1 or 2 sentences after providing a **complete analysis** without the assumption of equal standard deviations for the logged data (or for the original data). Perform the test in SAS or R.