UNIT 4 HW

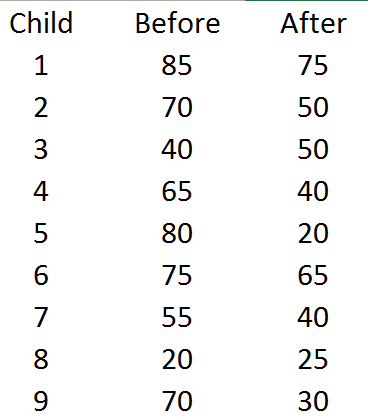
This class allows you to practice preparing professional looking reports. Make sure all reports are typed and all graphs (unless otherwise noted) are computer generated and copied and pasted into your report. If you would like help with Word or Excel please don’t hesitate to ask.

1. Read Chapter 4 from Statistical Sleuth and answer the conceptual problems at the end of the chapter. Note: You do not need to type these up and turn them in. The answers are at the very end of the chapter.
2. When wildfires ravage forests, the timber industry argues that logging the burned trees enhances forest recovery; the EPA argues the opposite. The 2002 Biscuit Fire in southwest Oregon provided a test case. Researchers selected 16 fire-affected plots in 2004, before any logging was done and counted tree seedlings along a randomly located transect pattern in each plot. They returned in 2005, after nine of the plots had been logged, and counted the tree seedlings along the same transects. The percent of seedlings lost from 2004 to 2005 is recorded in the table below for logged (L) and unlogged (U) plots:

Test the EPA’s assertion (and thus the opposite of the logging industries assertion) that logging actually increases the percentage of seedlings lost from 2004 to 2005.

* 1. Perform a complete analysis using a rank sum test in SAS. (Logging data).
  2. Verify the p-value and confidence interval by running the rank sum test in R (using R function Wilcox.test). (You do not need to repeat the complete analysis … simply cut and paste a screen shot of your code and the output.) You may use: <https://www.r-bloggers.com/wilcoxon-mann-whitney-rank-sum-test-or-test-u/> for reference.

1. Conduct a Welch’s two-sample t-test on the Education Data from HW 3 (untransformed). Perform a complete analysis using SAS and R to test the claim that the mean income of college educated people (16 years of education) is greater than the mean of those with a high school education only (12 years of education).
   1. State the problem, address the assumptions. Be sure to support with your knowledge of theory (CLT) as well as with histograms, box plots, q-q plots, etc.
   2. Show all 6 steps, including a thoughtful, thorough, yet non-technical conclusion. Include a confidence interval.
   3. Include a scope of inference at the end. (You may copy and paste this from a previous HW if you like.)
   4. Verify the Welch’s t statistic and p-value with R (using R function t.test). Simply cut and paste your R code and output. You may use: <http://rcompanion.org/rcompanion/d_02.html> for reference.
   5. Would you prefer to run the log transformed analysis you ran in HW3, or do you feel this analysis is more appropriate? Why or Why not? (Make mention of the assumptions as well as the parameters that each test provides inference on. As you know, they are different.)
2. Complete the questions below
   1. Chapter 4, Problem 20 from the text. Show all work. “By hand” here means actually by hand. Simply take a picture of your work and include it in your pdf/doc file or submit an Excel file. Include your sorted, labeled, and ranked data; your calculations of the mean and standard deviation of the assumed distribution of the rank sum statistic under Ho; your calculation of the Z statistic with a continuity correction; your p-value, and conclusion. (No confidence interval necessary here.)
   2. Problem 21 from the text. Take a screen capture of the SAS output in addition to your response.
   3. Write up a complete analysis using the information you have gained from A and B to test the claim that the distributions are different.
      1. State the problem.
      2. State the assumptions you are making and why you are making them. Justify your decisions. Print out any histograms, q-q plots, box plots, etc. that you use in your justification.
      3. Show all 6 steps of the hypothesis test for the rank sum test of the trauma data. Use the critical values, test statistics, p-values, etc. obtained above. Add a confidence interval from the Hodges-Lehmann procedure (from SAS).
      4. Also include a scope of inference statement.
3. A study was performed to test a new treatment for autism in children. In order to test the new method, parents of children with autism were asked to volunteer for the study in which 9 parents volunteered their children for the study. The children were each asked to complete a 20 piece puzzle. The time it took to complete the task was recorded in seconds. The children then received a treatment (20 minutes of yoga) and were asked to complete a similar but different puzzle. The data from the study is below:



a. Calculate the statistic S for a signed rank test by hand (see Section 4.10 from the Asynchronous lecture) showing the final table with the absolute differences, the signs, and the ranks. Also, show your calculation of the z-statistic (standardized S statistic). Note well this is paired data.

b. Verify your calculation in both SAS and R. Simply cut and paste your code and relevant output.

The code to use is below for SAS and R:

**For SAS**

\*To create the variable that is a difference between the two measurements;

data autism;

set autism;

diff = before - after;

run;

\*To get a signed rank test;

proc univariate data = autism;

var diff;

run;

**For R**

wilcox.test(autism$Before, autism$After, pair=T)

c. Conduct the six step hypothesis test using your calculations from above to test the claim that the yoga treatment was effective in reducing the time to finish the puzzle.

d. Use SAS or R to conduct a six step hypothesis test using a paired t-test to test the claim that the yoga treatment was effective in reducing the time to finish the puzzle.

f. Use your data from above to construct a “complete analysis” of the test that you feel is most appropriate to test the claim that the yoga treatment was effective in reducing the time to finish the puzzle. This is simply formatting your results. You should be able to cut and paste most of the work from above.