Lab 4

Statistics 411: Spring 2018

Due: In class February 5

Turn in your solutions in a typeset report. You do not need to report your answers to this homework in the format according to the Syllabus and *Writing a Statistical Report*. YOU MUST WORK WITH AT LEAST ONE OTHER CLASSMATE on this homework and TURN IN A SINGLE SOLUTION. The MAXIMUM NUMBER OF STUDENTS WHO CAN WORK TOGETHER IS THREE. List all students on the front page of your solution set.

1. In problem #2 of HW4, the boxplot shows that the "word knowledge" scores are left skewed with some small outliers. Does this suggest that the t-test that you applied is not appropriate? Explain. *Hint: this question is asking about the robustness of the t-test.*

2. In an attempt to transform the "word knowledge" scores to a more symmetric distribution, try log-transforming the data as your book suggests and plot it:

   ```
   boxplot(log(d$Math),log(d$Word),log(d$Arith),log(d$Parag),
           names=c("math","word","arithmetic","paragraph"),xlab = "AFQT categories",
           ylab="log(score)",main="Results from AFQT")
   ```

   You will get some warnings from R because there are some zero scores. Do the log-transformed scores for "word knowledge" look more normal than the original scores? Justify your answer. Use Display 3.8 on page 71 to explain why or why not the transformation worked.

3. Find the Box Cox transform for the "word knowledge" scores by implementing the following code:

   ```
   library(MASS)
   boxcox(d$Word[d$Word>0]~1,lambda=seq(1,3,.1))
   ```

   Which transform does Box Cox suggest? Include your R-code and R-output in an Appendix including the Box Cox plot.

4. Transform the data as suggested by Box Cox and generate a boxplot. Do the Box Cox transformed scores for "word knowledge" look more normal than the original scores? Justify your answer.