Homework 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. Consider Exercise 23 on page 56 of your text (but do not answer any of the questions posed by your textbook).

   (a) To determine a 90% confidence interval of the mean percent change in fatalities just among states that increased the speed limit, which t-tool is appropriate? Choose one of: A. 1-sample, B. 2-sample paired, C. 2-sample unpooled, D. 2-sample pooled. Justify your answer.

   (b) Load in the data for this problem and plot side-by-side boxplots. Cut and paste this into R:

   ```
   library(Sleuth3)
   s=ex0223
   boxplot(s$PctChange ~ s$SpeedLimit,
           names=c("increased","retained"),xlab="speed limit",
           ylab="Percent change in fatalities",
           main="Investigating the effect of increasing the speed limit over 55mph")
   ```

   Do the boxplots suggest that there is a difference between the states that increased the speed limit and states that did not?

   (c) Use R's `t.test()` function to construct a two-sided 90% CI for the true mean percent change in fatalities just among the states that increased the speed limit from 55mph. Use this 90% CI to test whether the percent of fatalities increased just among the states that increased the speed limit from 55mph? That is, just consider the data `s[s$SpeedLimit=="Inc",]`. *This is a different question than what your book asks!* Report on all 6 steps of the hypothesis test although instead of reporting a p-value to make a decision regarding H_0, use the two-sided 90% CI. Put all R-code and R-output, including the boxplot, into the Apendix.

   (d) Write a conclusion in terms of the problem.

   (e) What is the scope of inference for this problem?

2. Consider Exercise 22 on page 55 of your text (but do not answer any of the questions posed by your textbook).

   (a) Load in the data for this problem and plot side-by-side boxplots. Cut and paste this into R:

   ```
   library(Sleuth3)
   d=ex0222
   boxplot(d$Math,d$Word,d$Arith,d$Parag,
           names=c("math","word","arithmetic","paragraph"),ylab="score",
           main="Results from AFQT")
   ```

   Do the boxplots suggest that there is a difference between the mean "word knowledge" scores and the mean "math knowledge" scores?

   (b) To compare the mean "word knowledge" scores to the mean "math knowledge" scores, explain why a t-test is more appropriate than a z-test.

   (c) To compare the mean "word knowledge" scores to the mean "math knowledge" scores, which t-test is appropriate for this problem? Choose one of: A. 1-sample, B. 2-sample paired, C. 2-sample unpooled, D. 2-sample pooled. Justify your answer.

   (d) A researcher wants to test whether military recruits' math knowledge is less than their word knowledge on the average. Use R's `t.test()` function to test this hypothesis. Put the R-code and R-output, including the boxplot, in the Appendix.

(e) Report on all 6 steps of the hypothesis test at $\alpha = 0.01$.

(f) Write a conclusion in terms of the problem.

(g) What is the scope of inference for this problem?