STAR 511 HW #4 32 points total

1. A study is being planned to compare a new vs standard fertilizer for tomatoes. They plan to test for a difference in the mean yield (lb/plant) comparing between two groups of tomatoes (grown using new or standard fertilizer). The experimental units will be individual tomato plants of the same variety and age (randomly assigned to either new or standard fertilizer). Based on a previous study, they conjecture that the pooled standard deviation is 1.2. They want to be able to detect a meaningful difference (between means) of 1.5 (lb/plant). What sample size is required (per group) to achieve 90% power using alpha = 0.05. (**4 pts**)

Questions 2 through 10 (Potency data): The data from the O&L textbook problem 6.59 (ex6-59.txt) concerns drug potency. The values labeled "Sample1" correspond to measured potency for a random sample ($n_1 = 10$) of bottles drawn from current production. The values labeled "Sample2" correspond to measured potency for a random sample ($n_2 = 10$) bottles previously drawn and stored for a year. The goal of the study is to compare mean potency for current vs stored bottles (Sample 1 vs Sample 2).

Reminders about textbook datasets:

- You already downloaded the textbook datasets for a previous assignment.
- We will use the CSV files ("**ASCII-comma**"). The file extension is .TXT even though the files are actually CSV, meaning that the values are "comma separated".
- Since the column names in the textbook datasets are (single) quoted, adding the argument below to the read.csv statement will be handy. Don't copy/paste this code from this .pdf file; type it out! Copy/pasted characters from .pdfs are sometimes not equivalent to the similar looking characters R recognizes; quotation marks are notorious in this regard.

```
read.csv( , quote = " ' ")
```

2. Construct a side-by-side boxplot (should be a single graph), using meaningful labels. (**2 pts**)
3. Give the sample means and standard deviations for each "sample" (Sample1 and Sample2). (**2 pts**)
4. Using the summary statistics and plot from the previous questions, is the pooled two-sample t-test or Welch-Satterthwaite t-test preferred here? You may justify your response using visual inspection, or the "rule of thumb" from lectureNotes06.1 notes. You might also take into account the role of sample sizes in deciding which test to use. (**2 pts**)
5. Considering your response to the previous question, run an appropriate two-sample t-test. State the hypotheses, find the test statistic value and then use the p-value to make a (brief) conclusion, using $\alpha = 0.05$. (**4 pts**)

6. The p-value from question 5. is a probability statement about its corresponding test statistic. Write a precise interpretation of what this p-value tells you, in a way that incorporates the value of the test statistic. (**3 pts**)
7. Find the 95% confidence interval for the difference between the (population) means. Do we have evidence of a difference between the (population) means? Briefly justify your response using the confidence interval. (**4 pts**)

Note for Q2 – Q7: The data are in "wide" format. All questions can be answered using the current format. An alternative is to "transpose" the data to "long format". This is NOT required, but may be handy. The following example code assumes (1) the original data is called Potency after importing and (2) column names are Sample1, Sample2.

```
library(tidyverse)
PotencyTr <- Potency %>%
pivot_longer(Sample1:Sample2, names_to = "Sample", values_to = "Y")
      str(PotencyTr)
```

Questions 8 through 10 (Salt Sensitivity): The data from the textbook problem 6.28 (ex6-28.txt) concerns salt sensitivity after treatment for high blood pressure. Salt sensitivity is measured before and after treatment for n = 10 subjects.

8. Provide side-by-side boxplots of After, Before and Difference (should be a single graph). **Note:** Start by calculating the differences (Diff = After – Before) for each subject.
9. Use an appropriate t-test to test the null hypothesis that there is no population level difference in salt sensitivity before vs. after treatment.

    A. State the statistical hypotheses. (**2 pts**)
    B. Show your output for the test. Identify the values of the test statistic and p-value. (**2 pts**)
    C. Report an appropriate 95% confidence interval. (**1 pt**)

10. Use the Wilcoxon Signed Rank test to analyze these data, and show your output. (**3 pts**)
11. Based on your analyses, write a brief summary of what these data suggest regarding salt sensitivity before and after treatment for high blood pressure, using specific statistical results where you feel they are appropriate. (**3 pts**)