S ST 352

**Lab Assignment 2 SOLUTIONS**

***20 points***

***Due 11:59 PM on Friday, October 11***

**Reminder of the honor code:**

**lab assignments are to be completed individually!**

**Number your answers. Put the questions in order! Once completed, upload your assignment in Gradescope. After uploading your assignment in Gradescope, remember to link each question to the page on which that question appears in your document. (One-point deduction if you do not link questions to a page.)**

# **Part III: The Assignment**

The question of interest is: *Is there a difference in the average prices between the two grocery stores you decided to compare?*

After deciding which inference method to use (the two-sample t-methods or the paired t-methods), refer to Part II above for help with how to enter your data into an Excel spreadsheet correctly to import into R. Then refer to the Lab 2 Notes for how to use R to explore and analyze the data for the appropriate inference method.

1. ***(3 points)*** Which of the following is the most appropriate hypothesis test to use in this problem: paired test or two-sample test? Why?

**+2** Paired Test (although it IS okay to say one-sample t-test ON THE DIFFERENCES in prices between the two stores)

**+1** for explanation: The products at both stores are the same. The selection of a product at one store is *not* independent of the selection of products at the other store. The product selected at one store matches the products selected at the other store.

-2 for saying two-sample t-test

-3 points for saying any other type of test (although it IS okay to say one-sample t-test ON THE DIFFERENCES in prices between the two stores)

Note: we will NOT take of further points in the rest of the problem for using a wrong inference method. As long as a student’s answers for remaining questions in this problem are consistent with the inference method chosen in #1 and are correct for that inference method, no points should be deducted for the remaining questions.

2. ***(3 points)*** State the null and alternative hypotheses in statistical notation. Define any parameters used.

**+ 0.5** for using an correct parameter

**+ 0.5** If student defines the parameter correctly.

**+ 1** for correct null hypothesis

**+ 1** for correct alternative hypothesis

The mean price difference between Store A and Store B for all products carried by both stores (**must define how they are subtracting to get "d**" – if they don’t, -1/2)

If they give two parameters (such as) instead of one, **take off 1.5 points! (The only exception would be if they said they were doing a two-sample test in Question 1. Then, this hypothesis would be consistent with their answer to Question 1, in which case points should not be deducted.)**

3. ***(3 points)*** Do you feel it is appropriate to use the t-methods for this problem? Why or why not? (Part of the support for your answer to this question should refer to a graph. Include the appropriate graph with your answer to question 3.)

Answers may vary.

**+1** for stating if t-method is appropriate or not appropriate

**+1** for reasonable support.

**+1** for including an appropriate graph used to support their answer

t-methods may be appropriate if the shape of the differences is approximately symmetric or normally distributed even though the sample size is small.

t-methods are not be appropriate if the shape of differences is skewed.

4. ***(2 points)*** Regardless of your answer to #3 above, use R to obtain the t-statistic and p-value. Report the t-statistic with degrees of freedom and the p-value here. (See the Lab 2 Notes for help with the R code to use.)

Answers should vary.

+ 0.5 each for t-statistic and degrees of freedom (df = 14 for everyone if they sampled 15 items from each store)

+1 for p-value

5. ***(3 points)*** Based on your p-value, state a conclusion in the context of the problem.

The conclusion should be consistent with the p-value

**+2** Conclusion is written in context of the alternative hypothesis

**+1** proper conclusion based on p-value.

6. ***(3 points)*** Using R, obtain a 95% confidence interval. Report the confidence interval in proper syntax here. Then interpret this confidence interval in the context of the problem. (See the Lab 2 Notes for help with the R code *and* the interpretation.)

**+1** for writing confidence interval in proper syntax: (lower bound, upper bound) with units

**+2** for proper interpretation. Interpretation should include the following:

* **Level of confidence (1/4 point)**
* **The word “mean” or “average” (3/4 point)**
* **The variable (price) (1/2 point)**
* **The populations (i.e. two groups being compared) (1/4 point)**
* **The bounds (1/4 point)**

**A general template for their confidence interval: “We’re 95% confident that the mean price of items at Store A is between xxx and yyy higher (or lower, or combination of higher and lower) than the mean price of items at Store B.**

7. ***(2 points)*** Which of the two stores would you shop at? Why?

**+2** Any reasonable answer is fine, although the idea is to use the results of the analysis to determine which store is "cheaper", on average.

8. ***(1 point)*** Provide a copy of your data. (copy and paste it into your document either here or at the end of your document. If you put it at the end of your document, make a comment here so the grader knows where to look for it.)