Winter 2019 – ST 314 Data Analysis 05

Topics:

- One and Two Sample t Procedures
- Describing Visual Displays
- Hypothesis Testing and Confidence Intervals
- Lessons Covered: 26 - 30
- Textbook Chapter (Optional): 8 and 9

Grading:

- Points are listed next to each question and should total 25 points overall.
- Grading will be based on the content of the data analysis as well as the overall appearance of the document.
- Late assignments will not be graded.

Deadlines:

- Final Submission: **Monday, February 11th**. All submissions must be PDF files.

Instructions:

- Clearly label and **type answers** to the questions on the proceeding pages, ***without*** question prompts, in Word, Google Docs, or other word processing software.
- Insert **diagrams or plots as a picture** in an appropriate location.
- Math Formulas need to be typed with Math Type, LaTeX, or clearly using key board symbols such as +, -. *, /, sqrt() and ^
- Submit assignment to the Canvas link as a PDF. Verify the correct document has been uploaded. If not, resubmit. You can submit up to three times.

Allowances:

- You may use any resources listed or posted on the Canvas page for the course.
- You are encouraged to discuss the problems with other students, the instructor and TAs, however, all work must be your own words. Duplicate wording will be considered plagiarism.
- Outside resources need to be cited. Websites such as Chegg, CourseHero, Koofers, etc. are discouraged, but if used need to be cited and used within the boundaries of academic honesty.

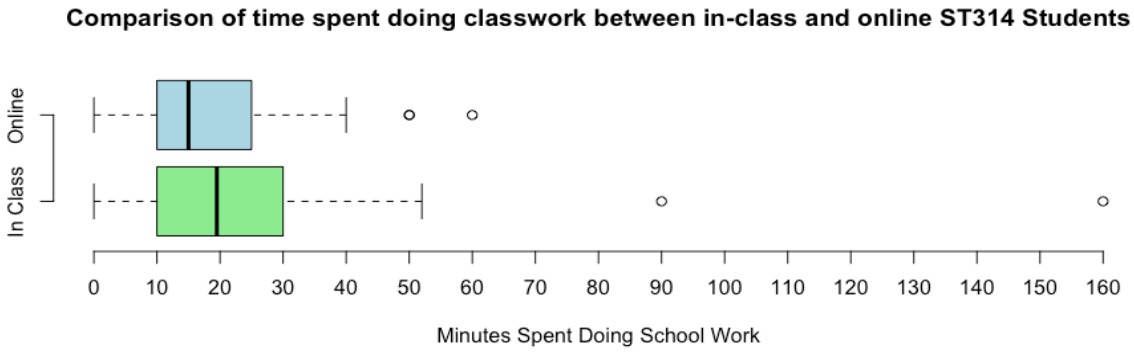
Part 1. (15 points)
The microbeersW19.csv dataset is a representative sample of 1,244 microbrews from around the United States. The variable `abv` represents the percent of alcohol by volume for each craft beer. According to the National Institute of Health, one standard serving of alcohol is 12 ounces of regular beer, which is usually about 5% alcohol by volume (abv).

Does the sample of microbrews provide evidence the average alcohol by volume of all craft beers is different from a standard serving of beer at 5% abv?

Use this dataset and the R script DA5_t_procedures.R to complete the following:

a) (1 point) What is the parameter of interest in this scenario? Provide the symbol and context.
b) (1 point) State the null and alternative hypothesis to answer the question of interest.
c) (2 point) Make a histogram or boxplot to visualize the variable abv. Is there visual evidence the average alcohol by volume is different than 5%?
d) (1 point) Calculate the sample mean and standard deviation using R. State the values.
e) (1 point) Check the conditions for inference. State them and whether they are met.
f) (1 point) Calculate the test statistic by hand. Show work.
g) (1 point) State the p-value. Is it one or two sided?
h) (2 points) Calculate the 95% Confidence Interval by hand. Show work.
i) (1 point) Use the `t.test()` command in R to verify the results of the t test. How do your answers compare?
j) (4 points) From the R output, write a four-part conclusion describing the results. Use $\alpha = 0.05$. Provide a statement in terms of the alternative hypothesis. State whether (or not) to reject the null. Give in context an interpretation of the point and interval estimate. Include any other information you might feel to relevant.

Part 2. (10 points)
The goal of this analysis is to compare the average time spent doing schoolwork during a week for ST314 students who are either in-class students (attend lectures in person) or online students (completing the course online). This data is from the combined ST314 Winter 2019 student information survey. The following software output is an analysis of this data:



	Mean	Std. Dev.	N
In Class	20.31	17.22	150
Online	18.31	12.67	67

```
        Welch Two Sample t-test

data:  st314data$SchoolWorkHours by st314data$Course
t = 0.95644, df = 168.89, p-value = 0.3402
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.127949  6.127750
sample estimates:
mean in group In class   mean in group Online
              20.31333               18.31343
```

Do these data provide evidence of a difference between the average time spent doing schoolwork in a week among in-class and online students? Use a significance level of 0.05 and answer the following questions using the software output**.**

a) (2 points) Describe the side-by-side boxplot. Is there visual evidence time spent doing schoolwork is different among in-class and online students? Explain.
b) (2 point) State the null and alternative hypothesis to answer the question of interest. Is the alternative one or two sided?
c) (2 points) Check conditions for the test. State each condition and whether it is met. If not met, state why. Then continue.
d) (4 points) From the R output, write a four-part conclusion describing the results. Provide a statement in terms of the alternative hypothesis. State whether (or not) to reject the null. Give in context an interpretation of the point and interval estimate. Make sure to provide a *direction* to your interval, for example, one group had a smaller (or larger) mean than the other, include

this relationship in your point and interval estimate. Include any other information you might feel to relevant.