Nolan Harris - nph2tx
Ethan Zeh - ekz5jf

1. During the previous lab activity, you played the Stroop effect game and recorded your times for both rounds. The file stroop students.csv contains these data. We will assume that current STAT 2120 students are representative of introductory STEM students at UVa.

A. Create a variable measuring the increase in time due to the Stroop effect. Note: The resulting values should be directly comparable to the values measuring the STAT TAs' increase in time that you created in the previous lab activity. In other words, use the same equation on the student data that you used on the TA data.

```
18 #%%
19
20 #create a variable to see the increase
21
22 increase = stroopData['Different'] - stroopData['Same']
23
24 n = 335
25 xbar = np.mean(increase)
26 s = np.std(increase, ddof = 1)
27 t_star = stats.t.ppf(0.95, df = n - 1)
```

B. Are the values measuring the increase in time created in activity A. independent? Explain

The values are not independent because we are measuring the increase in time it takes due to exterior factors. The first column, “Same” has the times using a different stress on the subject than the “Different” column. Each column, or test, relies on a specific challenge to the test taker, making it dependent on which test they are taking as to the resulting time.

2. Recall the STAT TAs data from the Stroop effect game analyzed in the previous lab activity.

A. Are the values measuring the increase in time due to the Stroop effect for STAT TAs created in the previous lab activity independent? Explain.

These are also not independent because like the previous explanation, they are relying on a sort of test - different or same, to determine the time it will take the subject.

B. Are the values measuring the time increase for STAT TAs independent of the the values measuring the time increase for current STAT 2120 students independent? Explain.

These are independent of one another because while it is the same test, there are different groups taking it - TAs, vs. Students.

C. What category of inference procedures would be appropriate for exploring any difference between the average time increase due to the Stroop effect for current STEM students and the average time increase due to the Stroop effect for STEM TAs? Explain.

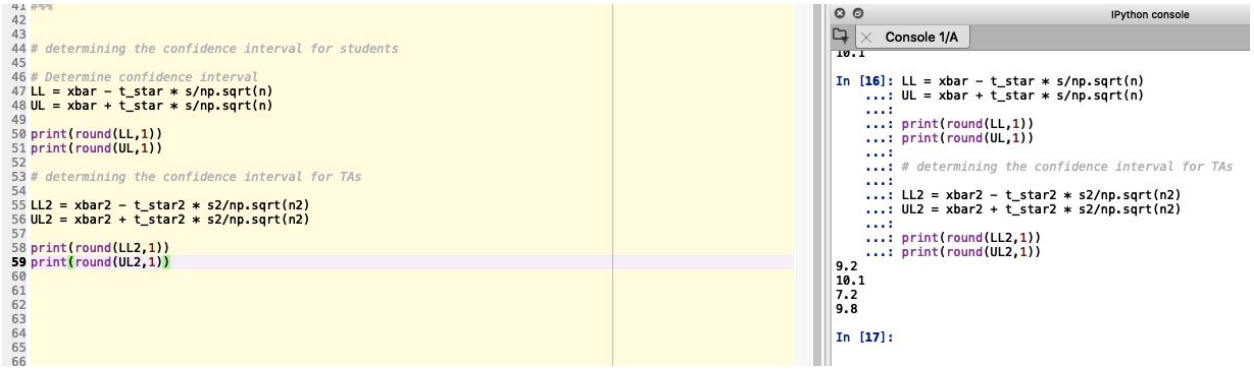
I believe interval estimation would be appropriate because we are not exactly making a hypothesis, and we are using the sample data to calculate an interval of possible (or probable) values of an unknown population parameter, in contrast to point estimation, which is a single number.

D. What are the appropriate hypotheses to test if there is a difference between the average time increase for current STEM students and STEM TAs? Be sure to specifically define all parameters.

Hypotheses:

a. there is no difference between the students and tas
b. The students have a larger difference than the tas
c. The tas have a larger difference than the students

E. Conduct the appropriate test using the conservative approach to determine if there is evidence that there is a difference in the average time increase due to the Stroop effect for current STEM students and STEM TAs.



```
#%%


# determining the confidence interval for students

# Determine confidence interval
LL = xbar - t_star * s/np.sqrt(n)
UL = xbar + t_star * s/np.sqrt(n)

print(round(LL,1))
print(round(UL,1))

# determining the confidence interval for TAs

LL2 = xbar2 - t_star2 * s2/np.sqrt(n2)
UL2 = xbar2 + t_star2 * s2/np.sqrt(n2)

print(round(LL2,1))
print(round(UL2,1))
```

IPython console

Console 1/A

```
10.1

In [16]: LL = xbar - t_star * s/np.sqrt(n)
    ...: UL = xbar + t_star * s/np.sqrt(n)
    ...:
    ...: print(round(LL,1))
    ...: print(round(UL,1))
    ...:
    ...: # determining the confidence interval for TAs
    ...:
    ...: LL2 = xbar2 - t_star2 * s2/np.sqrt(n2)
    ...: UL2 = xbar2 + t_star2 * s2/np.sqrt(n2)
    ...:
    ...: print(round(LL2,1))
    ...: print(round(UL2,1))
9.2
10.1
7.2
9.8

In [17]:
```

F. The graphic below shows boxplots for the time increase for both current STAT students and STAT TAs. Is your conclusion from activity E. reliable? Explain.

Our conclusion is somewhat reliable because while we may not have gotten exactly what the plots show, we do see a similar level of increase in the Tas, vs the increase in the students.

G. Explain the implication of conducting a conservative test in context.

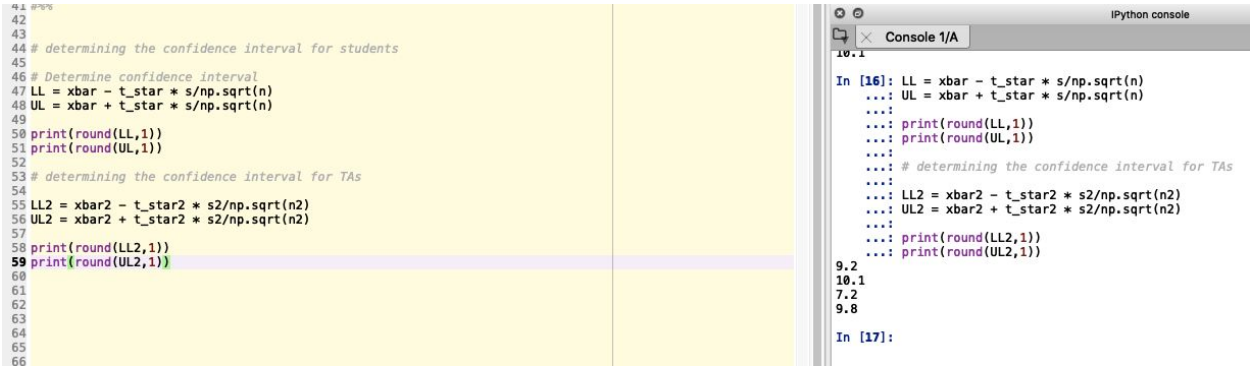
A conservative test keeps the probability of rejecting the null hypothesis below the significance level. In this case, our null hypothesis is that there is no difference between the students and the TAs. We want to be conservative so we don't make a type 1 error without proper, reliable data, in this case this would be rejecting there is no difference.

3. Recall from the previous lab activity that it is reasonable to assume that the time required for adults to read out loud (complete the first round of the game) is Normally distributed.

A. Construct a conservative 95% confidence interval for the difference in average time to complete the first round of the game between current STEM students and STEM TAs.

```
#%%

#create a variable to see the increase

increase = stroopData['Different'] - stroopData['Same']

n = 335
xbar = np.mean(increase)
s = np.std(increase, ddof = 1)
t_star = stats.t.ppf(0.95, df = n - 1)

#%%

url2 = 'https://raw.githubusercontent.com/UVAnog/statprojects/master/stroop_TAs.csv'
taData = pd.read_csv(url2)

taIncrease = taData['Different'] - taData['Same']

n2 = 38
xbar2 = np.mean(taIncrease)
s2 = np.std(taIncrease, ddof = 1)
t_star2 = stats.t.ppf(0.95, df = n - 1)

#%%


# determining the confidence interval for students

# Determine confidence interval
LL = xbar - t_star * s/np.sqrt(n)
UL = xbar + t_star * s/np.sqrt(n)

print(round(LL,1))
print(round(UL,1))

# determining the confidence interval for TAs

LL2 = xbar2 - t_star2 * s2/np.sqrt(n2)
UL2 = xbar2 + t_star2 * s2/np.sqrt(n2)

print(round(LL2,1))
print(round(UL2,1))

```

B. What observations can you draw from the interval constructed in activity A.?



```
#%%


# determining the confidence interval for students

# Determine confidence interval
LL = xbar - t_star * s/np.sqrt(n)
UL = xbar + t_star * s/np.sqrt(n)

print(round(LL,1))
print(round(UL,1))

# determining the confidence interval for TAs

LL2 = xbar2 - t_star2 * s2/np.sqrt(n2)
UL2 = xbar2 + t_star2 * s2/np.sqrt(n2)

print(round(LL2,1))
print(round(UL2,1))
```

IPython console

Console 1/A

```
In [16]: LL = xbar - t_star * s/np.sqrt(n)
    ...: UL = xbar + t_star * s/np.sqrt(n)
    ...:
    ...: print(round(LL,1))
    ...: print(round(UL,1))
    ...:
    ...: # determining the confidence interval for TAs
    ...:
    ...: LL2 = xbar2 - t_star2 * s2/np.sqrt(n2)
    ...: UL2 = xbar2 + t_star2 * s2/np.sqrt(n2)
    ...:
    ...: print(round(LL2,1))
    ...: print(round(UL2,1))
9.2
10.1
7.2
9.8

In [17]:
```

We conclude we are 95% confident the student time(s) will fall between 9.2 and 10.1 seconds, and the TA time(s) will fall between 7.2 and 9.8 seconds.

C. Explain the implication of constructing a conservative confidence interval in context.

When it comes to results, a conservative confidence interval would actually have a higher likelihood of containing the results. In our case we have a higher likelihood of placing our confidence in where a student, or a TA time will be.

4. Which t procedure is most appropriate in the following scenarios? Explain.
One-sample t procedure: one quantitative variable.
Two-sample t procedure: one quantitative variable and one categorical (binary) variable.

A. A lab safety officer plans to compare the durability of nitrile and latex gloves for chemical experiments. A random sample of 30 students in university labs are selected for the experiement. Each student will perform the same organic synthesis using the same procedure twice, once wearing each type of glove.

Matched pairs

B. If a new process for copper mining is to be adopted, it must produce more than 50 tons of ore per day, which is larger than the current process produces.

One sample

C. Students who want to learn French are divided into groups. One of the groups is flown to France where they live for one month. The other group is enrolled in an intensive, month-long French course at the university. At the end of the month, all students are given a standard French language exam

Matched pairs