Math 338  **Lab Assignment #22**  Fall 2019

Recall from lecture that confidence intervals come out of Neyman-Pearson ideas. However, instead of making a choice between two possible parameter values, we estimate an entire interval of values within which we suspect the true parameter value lies. Thus, we can use our confidence interval estimates to do two-sided hypothesis tests in the NHST framework.

First, let’s re-create the setosa.petal.length dataset:

**> library(dplyr)**

**> setosa.petal.length <- iris %>% filter(Species == "setosa") %>% select(Petal.Length)**

Recall that we used this dataset to test the null hypothesis that the population mean petal length was 1.3 cm. Now we will test this null hypothesis using a confidence interval instead.

**> t.test(setosa.petal.length$Petal.Length, conf.level = 0.95) # technically the confidence level is 0.95 by default**

Question #1 Paste the output from R. Write a sentence interpreting the 95% confidence interval for the population mean petal length of *setosa* irises.

Question #2 Based on our confidence interval, does the population mean petal length of *setosa* irises appear to be greater than 1.3 cm or less than 1.3 cm? Or can you not tell? Explain your answer.

Question #3 Based on our confidence interval, can we reject H0: μ = 1.3 in favor of the two-sided NHST alternative Ha: μ ≠ 1.3 at the 5% significance level? Why or why not?

Now, let’s look at the difference of population mean sepal lengths between *setosa* and *versicolor*. We’ll need the original *iris* dataset for this one. Since our *setosa* and *versicolor* samples are unrelated, this is a two-sample *t* confidence interval.

**> iris\_sv <- iris %>% filter(Species %in% c("setosa","versicolor"))**

**> t.test(Sepal.Length ~ Species, data = iris\_sv, conf.level = 0.95)**

Question #4 Paste the output table from R. Write a sentence interpreting the 95% confidence interval for the difference of population mean sepal lengths between setosa and versicolor.

Question #5 Based on our confidence interval, which species has longer petals? Or can you not tell? Explain your answer.

Question #6 Based on our confidence interval, can we reject H0: in favor of the two-sided NHST alternative Ha: at the 5% significance level? Why or why not?

Finally, we will do a *t* confidence interval for the population mean of paired differences (matched pairs confidence interval). We used the straight\_jeans2 dataset for this one and investigated the mean of the paired differences in maximum back pocket width (*maxWidthBackMens* and *maxWidthBackWomens*).

There’s actually a way to do a paired test without getting the difference first; we just almost always look at a histogram of the differences first. Just to get the practice, we’ll use a 99% CI.

**> t.test(x = straight\_jeans2$maxWidthBackMens, y = straight\_jeans2$maxWidthBackWomens, paired = TRUE, conf.level = 0.99)**

or (if you like to do the subtraction first and then just run the one-sample test) you could copy the code from Lab 19 and include the confidence level argument.

**> jeans\_diff <- straight\_jeans2 %>% mutate(maxWidthBackDiff = maxWidthBackMens - maxWidthBackWomens)**

**> t.test(jeans\_diff$maxWidthBackDiff, conf.level = 0.99)**

Question #7 Paste the output from R. Write a sentence interpreting the 99% confidence interval for the population mean of the paired differences in maximum back pocket width.

Question #8 Based on our confidence interval, whose jeans have wider back pockets – men or women? Or can you not tell? Explain your answer.

Question #9 Based on our confidence interval, can we reject H0: μd = 0 in favor of the two-sided NHST alternative Ha: μd ≠ 0 at the 1% significance level? Why or why not?

Question #10 Based on your answers in this lab, what information can you get from a confidence interval that you cannot get from a hypothesis test?