Math 338 # Lab Assignment #21 Fall 2019

In this lab, we will explore the theory of *t* confidence intervals for a population mean. The formula for the CI using a t critical value is given by:

$$\acute{X} \pm E$$

where

$$E = t^{¿} ¿ * ¿ \frac{s}{\sqrt{n}}$$

(Recall that it is standard to just write t*; however, I am using t** to reinforce that this is the same value as used for a N-P test with a two-sided critical region with $\alpha = 1 - C$)

Recreate the simulated squirrels dataset from Lab 15:

```
> n_squirrels <- 4

> n_samples <- 1000

> normal_matrix <- matrix(0, nrow = n_samples, ncol = n_squirrels) # initialize a matrix containing the actual random draws


> set.seed(338)  # reset seed

> for(i in 1:n_samples){  # a for loop - do this n_samples times

+  normal_matrix[i,] <- rnorm(n = n_squirrels, mean = 1.7, sd = 0.5)

+ }  # there is actually an easier way to do this in R, but we code it this way to illustrate conceptually what is actually being done


> mean_vector <- apply(normal_matrix, 1, mean)

> sd_vector <- apply(normal_matrix, 1, sd)

> squirrel_df <- data.frame(mean = mean_vector, sd = sd_vector)
```

Recall that **squirrel_df** contains the sample mean and sample standard deviation of 1000 simulated samples of size 4 from a N(1.7, 0.5) distribution. Although we do know in this simulation the true population standard deviation, we will pretend that we don't and so have to estimate it from the sample.

We know the sample mean, sample standard deviation, and sample size for each sample. Therefore, once we find the appropriate t** value, we can find the confidence interval for each sample. Remember that the t** value changes depending on the confidence level. For this lab, let's assume a 95% confidence level.

Question #1 For our confidence intervals, the t** value comes from a t-distribution with how many degrees of freedom?

3 degrees of freedom because n-1 => 4-1 = 3.

Now let's find the t** value. We want to find a critical value corresponding to the middle 95% of the distribution, so our proportion in the lower tail is 95% + half of the remaining 5%:

```
> C <- 0.95 # confidence level as a percentage

> alpha <- 1 - C

> t.star <- qt(alpha/2, df = degrees_of_freedom, lower.tail = FALSE) #
where degrees_of_freedom is your answer to #1
```

Question #2 What is the t** value for a 95% CI for population mean using a sample of size 4?

3.182446

Now that we have all the information we need, we can create our 1000 confidence intervals.

```
> library(dplyr)

> squirrel_CI <- squirrel_df %>% mutate(se = sd/sqrt(n_squirrels), CI.low =
mean – t.star*se, CI.high = mean + t.star*se)
```

Finally, let's look at the proportion of confidence intervals that contain the true population mean ($\mu=1.7$¿ .

```
> squirrel_CI <- squirrel_CI %>% mutate(hit = CI.low <= 1.7 & CI.high >=
1.7)

> prop.hit <- sum(squirrel_CI$hit)/n_samples
```

The variable `hit` will be TRUE if the interval contains the true population mean and FALSE if the interval does not. The variable `prop.hit`, is the proportion of intervals for which `hit` is TRUE.

Question #3 Report the confidence interval computed from the first sample below. Write a sentence interpreting the confidence interval.

(1.106229899, 3.042801). This is the possible range of values that mu can be. Since we have a given mu value of 1.7 and clearly is in this range, we can accept our null hypothesis.

Question #4 What proportion of values in the first sample are within that 95% confidence interval from **Question #3**? Is this proportion anywhere close to 95%?

Since the first four values of the matrix returns values inside the weight interval, we don't have 95% confidence.

Question #5 Use the `pnorm` command to find find the proportion of values (from a N(1.7, 0.5) distribution) that we expect to be within that 95% confidence interval from **Question #3** (look at Lab 13 for example code). Is this proportion anywhere close to 95%?

The proportion is 0.8824927 and is not 95%.

Question #6 What proportion of population means are within that 95% confidence interval from **Question #3**? Is this proportion anywhere close to 95%?

Since the value of mu is in the range of values, the probability will be either 0% or 100%, and will never be 95%

Question #7 What proportion of your 95% CIs contain the true population mean? Is this anywhere close to 95%?

Our proportion of values is 0.949 which is nearly the 95% we proposed earlier.

Now let's try to estimate the population mean petal length of *setosa* irises.

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

> t.test(setosa.petal.length$Petal.Length, conf.level = 0.95) # we use the
t.test command to do a CI for mean but the only argument we need is the
confidence level
```

Question #8 Paste the output from R below. Write a sentence interpreting the interval.

One Sample t-test

data: setosa.petal.length$Petal.Length
t = 59.528, df = 49, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1.412645 1.511355
sample estimates:
mean of x
   1.462

The interval in this given sample is (1.412645, 1.511355), where the value of mu should be.

Question #9 If we changed the confidence level to 99%, would you expect the confidence interval to get wider or narrower? Explain your reasoning.

The confidence interval would widen because there is less alpha area on the graph.

Now change the confidence level to 0.99 and re-run the `t.test` command.

Question #10 Paste the output from R below. Did the confidence interval get wider/narrower as you expected?

One Sample t-test

data: setosa.petal.length$Petal.Length

t = 59.528, df = 49, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

1.412645 1.511355

sample estimates:

mean of x

1.462

It did get wider as expected