Cody\_York\_M11\_lab

Cody York

2022-11-07

## Imports

This is where I am keeping all my R library import statements.

library(readr)  
library(plotrix)

## Read the Dataset

For this step I am using the readr library to read in the csv containing the data and viewing it to make sure it is coming correctly. I will also get the population mean since we will need that in the future.

Soil\_Organic\_Carbon <- read\_csv("Soil Organic Carbon.csv")

## Rows: 261 Columns: 468  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): 10-20cm, 20-30cm  
## dbl (466): Sample ID, 0-5cm, 5-10cm, DEM, Aspect, Slope, MaxH, NDVI, Band 1,...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#View(Soil\_Organic\_Carbon)  
  
pop\_mean <- mean(Soil\_Organic\_Carbon$`0-5cm`)

## Sample 50 rows and for 100 times calulate the 95% Confidence Level

For the first part of this task, I will be using the sampling method we learned from the last R lab.

1. I first set the seed so that the randomness won’t be different every time for testing the code.
2. Next, I create a variable to hold the number of samples I need.
3. Next, I set a variable to the number of runs.
4. Now, I will create a function that will run 100 times to get the data we need.
5. Then, I get the samples from the imported data set above using the sample() command in R.
6. After this I will calculate the z-value using the qnorm() function and the alpha of 95% confidence interval which is .025.
7. I will then get the mean and standard deviation of the sample.
8. Next, I will calculate the margin of error. This is the z-value times the standard deviation divided by the square root of the sample size.
9. After getting those pieces of information, I will calculate the lower and upper bounds of the confidence interval by subtracting and adding the margin of error to the sample mean.
10. I will then run the function 100 times using the replicate function to get the data.
11. Next I do a transpose on the data to get the three observations I need.

set.seed(111)  
  
number\_of\_samples <- 50  
  
runs <- 100  
  
population\_mean <-   
  
one.trial <- function(){  
 Soil\_Organic\_Carbon\_50\_Sample <- Soil\_Organic\_Carbon[  
 sample(nrow(Soil\_Organic\_Carbon),number\_of\_samples),2  
 ]  
 positive\_z <- qnorm(.975)  
 negative\_z <- qnorm(.025)  
 sample\_mean <- mean(Soil\_Organic\_Carbon\_50\_Sample$`0-5cm`)  
 standard\_deviation <- sd(Soil\_Organic\_Carbon\_50\_Sample$`0-5cm`)  
 margin\_of\_error <- positive\_z \* (standard\_deviation/sqrt(number\_of\_samples))  
 lower\_bound <- sample\_mean - margin\_of\_error  
 upper\_bound <- sample\_mean + margin\_of\_error  
 middle\_of\_bounds <- (lower\_bound + upper\_bound) / 2  
 confidence\_intervals <- c(middle\_of\_bounds, upper\_bound, lower\_bound)  
 return(list(middle\_of\_bounds, upper\_bound, lower\_bound))  
}  
  
sample\_50\_confidence\_interval <- replicate(runs, one.trial())  
  
final\_sample\_50\_confidence\_interval <- as.data.frame(  
 t(sample\_50\_confidence\_interval))

## 

## Graph the 50 sample confidence intervals

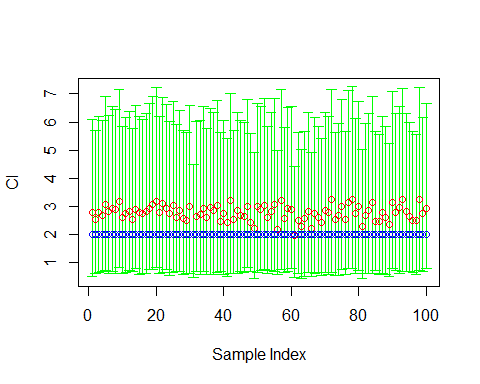
Now I will graph the confidence intervals collected above for the 50 samples.

1. First, I will set the sample index to the number of runs form above.

2. Next, I will use plotCI() to plot the middle, upper, and lower confidence intervals.

3. Then I will add the population mean to the observations in the plotCI().

sample\_index <- c(1:runs)  
  
plotCI(sample\_index,   
 as.numeric(final\_sample\_50\_confidence\_interval$V1),  
 as.numeric(final\_sample\_50\_confidence\_interval$V2),  
 as.numeric(final\_sample\_50\_confidence\_interval$V3),  
 lwd = 0.5,  
 col = "red",  
 scol = "green",  
 ylab = c("CI"),  
 xlab = c("Sample Index")  
 )  
plotCI(sample\_index,   
 rep(as.integer(pop\_mean), runs), 0, 0, col = "blue", add = TRUE)



## Sample 15 rows and for 100 times calulate the 95% Confidence Level

I will now do the same as above but only use 15 rows as a sample instead of 50.

set.seed(111)  
  
number\_of\_samples <- 15  
  
runs <- 100  
  
population\_mean <-   
  
one.trial <- function(){  
 Soil\_Organic\_Carbon\_15\_Sample <- Soil\_Organic\_Carbon[  
 sample(nrow(Soil\_Organic\_Carbon),number\_of\_samples),2  
 ]  
 positive\_z <- qnorm(.975)  
 negative\_z <- qnorm(.025)  
 sample\_mean <- mean(Soil\_Organic\_Carbon\_15\_Sample$`0-5cm`)  
 standard\_deviation <- sd(Soil\_Organic\_Carbon\_15\_Sample$`0-5cm`)  
 margin\_of\_error <- positive\_z \* (standard\_deviation/sqrt(number\_of\_samples))  
 lower\_bound <- sample\_mean - margin\_of\_error  
 upper\_bound <- sample\_mean + margin\_of\_error  
 middle\_of\_bounds <- (lower\_bound + upper\_bound) / 2  
 confidence\_intervals <- c(middle\_of\_bounds, upper\_bound, lower\_bound)  
 return(list(middle\_of\_bounds, upper\_bound, lower\_bound))  
}  
  
sample\_15\_confidence\_interval <- replicate(runs, one.trial())  
  
final\_sample\_15\_confidence\_interval <- as.data.frame(  
 t(sample\_15\_confidence\_interval))

## 

## Graph the 15 sample confidence intervals

I will now graph the 15 samples like I did for the above 50 samples.

sample\_index <- c(1:runs)  
  
plotCI(sample\_index,   
 as.numeric(final\_sample\_15\_confidence\_interval$V1),  
 as.numeric(final\_sample\_15\_confidence\_interval$V2),  
 as.numeric(final\_sample\_15\_confidence\_interval$V3),  
 lwd = 0.5,  
 col = "red",  
 scol = "green",  
 ylab = c("CI"),  
 xlab = c("Sample Index")  
 )  
plotCI(sample\_index,   
 rep(as.integer(pop\_mean), runs), 0, 0, col = "blue", add = TRUE)

