3818 R Homework 3

*** Student Name ***

8/20/2020

Question 1

You sample 100 people's showering habits and find the average shower time is 12.2 minutes. Given that the population variance shower time is 15, use the qnorm() command to construct an 86% confidence interval for a sampled mean.

Make sure you are using qnorm() correctly. Note: qnorm$(1 - \alpha)$ returns corresponding Z_α corresponding to $P(Z > Z_\alpha) = 1 - \alpha$. For a $100\gamma\%$ confidence interval, we have that $\alpha = \frac{1-\gamma}{2}$. See help(qnorm) for more details.

```
# Code for Question 1 goes here
```

Answer:

Question 2

Load the housing data we've been using: `housing_df <- read.csv("https://mattbutner.github.io/data/housing_df.cs` - Use a combination of the `mean()`, `sd()`, `length()`, `qnorm()`, and `sqrt()` functions to construct a 90% confidence interval for the CRIM variables. - Interpret the confidence interval from CRIM.

```
# Code for Question 2 goes here
```

Answer:

Question 3

Below is a block of code that simulates a bunch of random samples of the same size, constructs the mean and a confidence interval for each sample, and reports the percent of the confidence intervals that capture the true population mean. To begin, make sure you have `sample_size <- 100 num_samples <- 50 ci_level <- 0.95` set up in the beginning of the document. You will need to install the user written package ggplot2. To do this, type `install.packages("ggplot2")` into the console before you run the R script. You will need to be connected to the internet. For more information, see https://ggplot2.tidyverse.org/. No need to provide the R code for these questions.

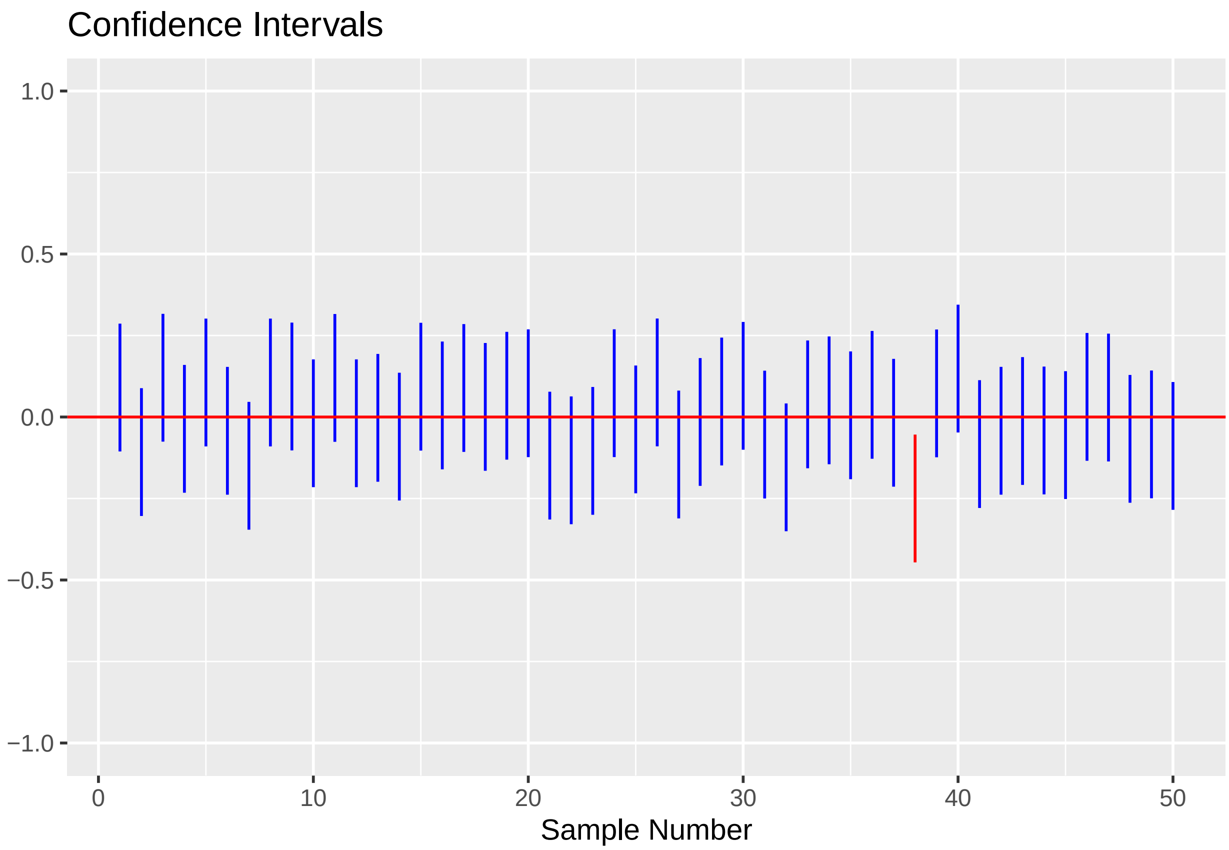
```
# This makes sure we get the same random numbers every time
set.seed(123)

# Setting the parameters of the simulation
sample_size <- 100
num_samples <- 50
ci_level <- 0.95

# theoretical pop mean / standard deviation (sd) for standard normal
pop_mean <- 0
pop_sd <- 1
```

```r
# Uses `for` loop, draws `sample_size` from random normal and makes mean

# What we will store sample means into
sample_means <- c()

for(i in 1:num_samples){
    # Draw normal random variable
    sample <- rnorm(n= sample_size, mean= pop_mean, sd= pop_sd)

    # Mean
    sample_mean <- mean(sample)

    # Store result in position `i` from loop.
    sample_means[i] <- sample_mean
}

# finding the margin of error
moe <- qnorm(1 - (1-ci_level)/2) * pop_sd / sqrt(sample_size)

# binding x\bar-moe and x\bar+moe as two columns
CI <- data.frame(CI_lower= sample_means - moe, CI_upper= sample_means + moe)

# to make the nice picture, you need the user written package ggplot2
## you might need to type: install.packages("ggplot2") into the console if you get an error
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
## Warning: replacing previous import 'vctrs::data_frame' by 'tibble::data_frame'
## when loading 'dplyr'
```

```r
# a logical vector, TRUE if the confidence interval captures the population mean
CI$is_mu_in_CI <- ((pop_mean > CI$CI_lower) & (pop_mean < CI$CI_upper))

## making the plot
ggplot() +
  geom_linerange(
    aes(x = 1:num_samples,
        ymin = CI$CI_lower,
        ymax = CI$CI_upper,
        col = CI$is_mu_in_CI,
        )
    ) +
  scale_color_manual(values=c("TRUE"="blue","FALSE"="red")) +
  guides(col= FALSE) +
  labs(title= "Confidence Intervals", x= "Sample Number") +
  geom_hline(yintercept= pop_mean, col= "red") +
  coord_cartesian(ylim= c(pop_mean-pop_sd, pop_mean+pop_sd))
```




```r
# What percent of the confidence intervals capture the population mean?
mean(CI$is_mu_in_CI)
```

```
## [1] 0.98
```

- Increase the sample size from 100 to 200, by changing `sample_size` and rerunning the code.
  - What happens to the width of the confidence intervals?
  - Does the true population mean fall inside more of the confidence intervals?

Answer:

- Return the sample size back to 100. Now change the number of samples from 50 to 100.
  - How does this change the percentage of the confidence intervals that capture the population mean?
  - As you increase the number of samples, towards infinity, what percentage of the confidence intervals will capture the true population mean?

Answer:

- Return the number of samples back to 50. Now change the confidence level to 0.8.
  - How does the width of the confidence intervals change?
  - Does the percentage of confidence intervals that capture the population mean increase or decrease?

Answer:

Question 4

You started taking the bus to work. The local transit authority says that a bus should arrive at your bus stop every five minutes. After a while, you notice you spend a lot more than five minutes waiting for the bus, so you start to keep a record.

You spend the next two months recording how long it takes for the bus to arrive to the bus stop. This give a total of sixty observations that denote the number of minutes it took for the bus to arrive (rounded to the nearest minute). These observations are hosted at https://mattbutner.github.io/data/bus_stop_time.csv

- Load these data into R as a data frame titled bus_stop_time using the read.csv() command.
- Create a histogram of the time_until_bus variable. Would you say that five minutes is a reasonable guess for the average arrival time based on this picture alone?
- Create 95% confidence interval for the bus arrival times using the z-distribution. Does 5 minutes fall within the 95% confidence interval?
- How would you communicate your finding to the local transit authority?

```
# Code for Questin 4 goes here
```

Answer: