Instructions for R Homework 3*

Question 1

You sample 100 people's showering habits and find the average shower time is 12.2 minutes. Given that the population variance in 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(x)` returns the Z_x such that $P(Z > Z_x) = x$, where $Z \sim N(0, 1)$. For a $100 - \gamma$% confidence interval, we have that $\alpha = \frac{1-\gamma}{2}$. See `help(qnorm)` for more details.

Question 2

Load the housing data we've been using:

```
df <- read.csv("https://mattbutner.github.io/data/housing_df.csv")
```

- 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.

* For all HW assignments, I need to see all the code used

Question 3

In the assignment `HW3.Rmd` there 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 this chunk. 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. No need to provide the R code for these questions (you can set `include = FALSE`).

Run the chunk with the initial settings. This will be the comparison.

a. Increase the sample size from 100 to 200, by changing $samp \leq_s ize$ 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?

b. 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?

c. 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?

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 using the `read.csv()` command.
- Create a histogram of the time_until_bus variable using `hist()`. 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?