ECO240 R- Homework 1

Due Date: April 9, Monday at 16:00

- Submit a hard copy to my office before the due. No late homework will be accepted.
- Submit "*Contribution Paper*" and "*Honor Code*" forms *signed* by all the group members.
- Your group can be up to 4 students.
- Any copied/being copied HW will get ZERO point. No negotiation.
- If suspected, I reserve the right to invite you for the further investigation.
- Please read "Honor Code" document carefully to understand what COPY means.

Objectives

1. Understand Sampling Distribution of Sampling Means

2. Understand Confidence Interval

Refer to: http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/sampling_distributions/sampling_distributions.html (or follow the link on the course web page)

Data Set: https://www.openintro.org/stat/data/?data=ames

<Sampling Distribution of Sample Means>

For the following questions, use SalePrice variable.

1. Examine the distributional shape of SalePrice variable.

2. Calculate mean and standard deviation of SalePrice. (population mean and variance)

3. Take one sample from the population by setting n = 50. Calculate the mean. Repeat this for five times and observe the calculated sample means and its' spread.

4. Repeat 3. by setting n = 10. Report any observation in terms of sample means and its spread by comparing the case in 3.

5. By using the "for loop" code (below), take 3000 random sample from the population, for each sample set the sample size as 100. Examine the distributional shape.

```
sample_means100 <- rep(NA, 3000)

for(i in 1:3000){
   samp <- sample(SalePrice, 100)
   sample_means100[i] <- mean(samp)
   }

hist(sample_means100)
```

6. Repeat the step 5. By setting the sample size as 10. Compare the distributional shapes from 5 and 6. Do not use the same variable name from 5. Change it to *sample_means10*, maybe.

7. What is the expected value of the sample means from the sampling distribution of sample means from 5? Find the variance for the sampling distribution of sample means as well. Compare them with the values found in 1.

8. Repeat the same procedure from 7 by using the result from 6.

9. Take one sample from the population by setting n=100. Calculate the probably that the sample mean exceeds 213500.

<Confidence Interval>

Refer to: http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/confidence_intervals/confidence_intervals.html

10. Derive a 95% confidence interval by taking one sample with n = 100. Consider two cases.
    Case1: use population standard deviation & z-score (qnorm( )).
    Case2: use sample standard deviation & t-score (qt(, df =)).

11. Compare the result from 2. and 10., confirm if the derived interval contains the actual population mean.

12. Repeat 10 and 11 by taking one sample with n = 10. Compare the width of the confidence interval.

13. Derive 50 95% confidence intervals by referring to the codes on the reference webpage. Set n = 100. Using the following function, plot all intervals. What proportion of your confidence intervals include the true population mean? Is this proportion exactly equal to the confidence level? If not, explain why. (You need to first run "plot_ci" function, then run the following code.)

```
plot_ci(lower_vector, upper_vector, mean(SalePrice))
```

Enjoy!