ECO240 R- Homework 1
Due Date: April 11, Thursday at 12:45

Objectives

1. Understand Sampling Distribution of Sampling Means

2. Understand Confidence Interval

3. Understand Hypothesis Testing

Refer to: http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/sampling_distributions/sampling_distributions.html (Sampling Distribution)

http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/confidence_intervals/confidence_intervals.html (Confidence Interval)

http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/inf_for_numerical_data/inf_for_numerical_data.html (Hypothesis Testing)

<Sampling Distribution of Sample Means>
Select one variable with large sample size. You can use any database, but you cannot use the one from last year's homework (RHW1 or 2 from 2017_2018). By assuming that the variable itself is the population data, answer the following questions.

(a) Observe the distributional shape of the original data. Is it symmetric or skewed?

(b) By using "for loop" code (refer to the reference page), take 1000 samples from the population with sample size $n = 10$, $n = 50$, $n = 100$, separately.

(c) Draw histograms for each case ($n = 20$, $n = 50$, $n = 100$).

(d) Derive the acceptance intervals for each case.

(e) Comment your findings from (b), (c) and (d).

<Confidence Interval and Hypothesis Testing>

Select one variable with large sample size, which you are going to construct confidence intervals and conduct hypothesis testing.

(a) Explain the reason why you've selected the variable. What is the objective of your research? Why are you interested?

(b) Derive 95% confidence intervals by taking one sample with n = 100. Derive one with population standard deviation, one with sample standard deviation. Explain what you've found.

(c) Repeat (b) for 20 times. Draw the derived confidence intervals. Analyze what proportion of your confidence intervals include the true population mean.

(d) Repeat (b) by setting n = 10. Did you observe any difference? Explain.

(e) Repeat (d) for 20 times. Draw the derived confidence intervals. Analyze what proportion of your confidence intervals include the true population mean. Observe any difference from (c)?

(f) Set up your own hypothesis using your data. Explain why you are interested in testing your hypothesis.

(g) Conduct your hypothesis testing. Don't forget to derive "Decision Rule" and make your "Decision" by taking one sample (n = 100) from the dataset. What did you find?

Enjoy!