**R Simulation Exercise I Lecture 5 10/13**

In this simulation exercise, you will get an opportunity to conduct simulation analyses for simple regression model using an interactive R shiny app.

First, open app.R in R studio, then click on “Run App”

Graphical user interface, text

Description automatically generated

An R studio pop-up shiny window will appear. To view the proper format, click on “Open in Browser”. Then you are ready to go!

Graphical user interface, text, application, email

Description automatically generated

There are three modules in this shiny app.

1. “Simulate” allows you to simulate/randomly draw one data set with your choices of population intercept, slope and sample size. A regression based on the sample will be estimated. You can compare the estimated sample intercept and slope with the population intercept and slope you specify, and observe how you expect the results might change given different inputs.
2. “Sampling Distribution” allows you to *repeatedly draw samples* using the same set up (sample size, intercept and slope etc). The regression result based on each sample data will be shown. You can observe the extent of variations of these sample estimates as compared with the population intercept and slope that you specify, and get to “see” what Standard Error really quantifies, and see Central Limit Theorem in action.
3. “Hypothesis Testing” allows you to look at the HT problem from a perspective that knowing whether the H0 or the Ha is true. The app allows you conduct the same test many times, you can then observe how many times you get the results wrong by rejecting when H0 is actually true (the type I error), and how many times you will be able to successfully reject the null when Ha is actually true (the power of the test).

Graphical user interface

Description automatically generated

**Some guiding questions**

We will split into breakout sessions and work on these problems in groups. (There will be a video guiding the asynchronous students through these exercises)

1. “Simulate!” tab
   1. What happens to estimated regression line as compared to the true/population regression line when you increase the sample size? What about SE? p-value? And the t statistics? How about other regression results such as R2?
   2. What happens when you draw a different sample given the same settings?
   3. What happens when you alter the ratio between slop and model variance (signal-noise ratio)? When signal-noise ratio increases, given the same sample size, do you see tend to see more or less significant results?
2. “Sampling Distribution” tab
   1. Draw many samples from the same settings (imagine you can draw multiple samples for the same study), appreciate the variations of estimated regression lines and how they vary around the true/population regression line
   2. Examine the distribution (histogram) of many estimated slopes, what does this distribution tell you? What shape does it look like, what’s the center/mean of this distribution? Compare it with the true population slope value. What does the SD and variance of this distribution represent?
   3. Compare the SD of this distribution and the SE of estimated slope using one of the datasets (samples), what do you notice? And what happens if you increase the sample size?
3. “Hypothesis Testing” tab
   1. The parameter settings are given in the “simulate!” tab, and the dataset will be drawn according to the settings.
   2. If you set population beta1=1, and input 0 for H0: beta1=0, what’s the ideal test result? Reject H0 or accept H0?
      1. In this situation, if you end up rejecting H0, is it a mistake? How about if you fail to reject H0? Is it a mistake?
      2. Given the same dataset, let’s change the null hypothesis to be H0: beta1=1. Now what’s the ideal test result? Reject H0 or accept H0?
         1. Given the same dataset, if you change your hypothesized value of beta, the test result will be different!
   3. Examine the results of HT over many sample results (i.e. many tests, each is based on a separate sample data of equal sample size).
      1. Follow the definition of type I error, and examine type I error rate (% of rejections when H0 is true). As long as H0 is true, does type I error rate change much when you change sample size or signal noise ratio?
      2. Follow the definition of power, and examine the power when H0 is not true, and how power varies by sample size and signal-noise ratio.