Lab

# Lab 4 – Sampling: Decision-Making Thresholds

A key focus of Chapter 10 is how to make inferences about populations based on samples. The essential logic lies in comparing a single instance of a statistic, such as a sample mean, to a distribution of such values. The comparison can lead to one of two conclusions – the sample statistic is either extreme or not extreme. But what are the thresholds for making this kind of judgment call (i.e., whether a value is extreme or not)? This activity explores that question.

The problem is this: You receive a sample containing the ages of 30 students. You are wondering whether this sample is a group of undergraduates (mean age = 20 years) or graduates (mean age = 25 years). To answer this question, you must compare the mean of the sample you receive to a distribution of means from the population. The following fragment of R code begins the solution:

set.seed(2)

sampleSize <- 30

studentPop <- rnorm(20000,mean=20,sd=3)

undergrads <- sample(studentPop,size=sampleSize,replace=TRUE)

grads <- rnorm(sampleSize,mean=25,sd=3)

if (runif(1)>0.5) { testSample <- grads } else { testSample <- undergrads }

mean(testSample)

After you run this code, the variable “testSample” will contain either a sample of undergrads or a sample of grads. The line before last “flips a coin” by generating one value from a uniform distribution (by default the distribution covers 0 to 1) and comparing it to 0.5. The question you must answer with additional code is: Which is it, grad or undergrad?

Here are the steps that will help you finish the job:

1. Annotate the code above with line-by-line commentary. To get full credit on this assignment, you must demonstrate a clear understanding of what the six lines of code actually do! You will have to look up the meaning of some commands.

2. The next line of code should generate a list of sample means from the population called “studentPop.” Very similar code to accomplish this appears right in Chapter 7. How many sample means should you generate? You can create any number that you want – hundreds, thousands, whatever – but I suggest that you generate just 100 means for ease of inspection. That is a pretty small number, but it makes it easy to think about percentiles and ranks.

3. Once you have your list of sample means generated from studentPop, the trick is to compare mean (testSample) to that list of sample means and see where it falls. Is it in the middle of the pack? Far out toward one end? Here is one hint that will help you: In Chapter 7, the quantile() command is used to generate percentiles based on thresholds of 2.5% and 97.5%. Those are the thresholds we want, and the quantile() command will help you create them.

4. Your code should end with a print() statement that could say either, “Sample mean is extreme,” or “Sample mean is not extreme.”

5. Please submit both the output of your runs and the R code.