Week 3 – Sampling Distributions Exercise #1

Instructions: In the asynchronous material for this week, we created sampling distributions by building a simulated population from random data and then drawing samples from that population. By examining large collections of samples, we can understand what to expect when we make inferences from samples of real data. In this exercise, we create another simulated population: scores on a test of achievement. Achievement tests like this one are often calibrated so that the population mean is 100 and the population standard deviation is 10. The code develops a simulated population of N=100,000 test takers, each of whom scored somewhere between 60 and 140 on the test. Using the code, answer the following questions about the simulated population:

1. Here’s code to create the population and show a histogram. Add a comment describing the shape of this distribution.

set.seed(1234) # Control randomization

testPop <- rnorm(100000, mean=100, sd=10)# Make simulated pop

hist(testPop)

1. This next line of code marks the 1st, 2nd, and 3rd quartiles. Add a new line of code that shows on the console the specific values associated with each of these quartiles. The second quartile has a special name. What is it?

abline(v=quantile(testPop, probs=c(0.25,0.5,0.75)), col="blue")

1. The 1st and 3rd quartiles in that histogram seem very close to the middle! Can that be correct? What percentage of cases fall below the 1st quartile? Above the 3rd quartile? Add code that calculates the number of observations in each division.
2. By design, this achievement test has a population standard deviation of 10. The following code uses abline() to mark points that are two standard deviations away from the mean. Examine the histogram and report what X value is two standard deviations below the mean and what X value is two standard deviations above the mean:

hist(testPop)

testPopMean <- mean(testPop) # Calculate mean

sumSq <- sum((testPop-testPopMean)^2) # Sum of squares

testPopVar <- sumSq/length(testPop) # Pop Variance

testPopSD <- sqrt(testPopVar) # Pop Std Deviation

abline(v=(testPopMean - 2\*testPopSD), col="red")

abline(v=(testPopMean + 2\*testPopSD), col="red")

1. This next code marks new quantiles, specifically the 0.025 quantile and the 0.975 quantile. Estimate where these quantiles fall on the X-axis.

hist(testPop)

abline(v=quantile(testPop, probs=0.025),col="green") # Lower tail

abline(v=quantile(testPop, probs=0.975),col="green") # Upper tail

1. The area between the green lines is known as the “central region” and the two areas outside of the green lines are the “tails.” Add a comment to tell what percentage of cases occur in the central region. What percentage of cases fall in the lower tail, that is, below the 0.025 quantile in the histogram above? What percentage of cases fall in the upper tail, that is, above the 0.975 quantile in the histogram above?
2. Share your code so far