Week 3 – Reasoning with the Central Region and Tails of a Sampling Distribution

Instructions: Use the code below to re-create the population and build a sampling distribution of means. The last four lines graphically compare the histogram of the original population and the sampling distribution on the same X scale.

set.seed(1234) # Control randomization

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

# Custom function to pull one sample of size n

sampleTestScores <- function(n){sample(testPop,size=n,replace=TRUE)}

# Write a comment explaining this line

samplingDistribution <- replicate(1000, mean(sampleTestScores(100)))

par(mfrow=c(2,1))

hist(testPop, xlim=c(50,140))

hist(samplingDistribution, xlim=c(50,140))

par(mfrow=c(1,1))

The takeaway is that the sampling distribution of means (lower histogram) converges on the same mean as the population (upper histogram) but the sampling distribution (lower histogram) is much less dispersed. The smaller dispersion results from the corrective influence of having many sampled observations contribute to each sample mean.

In the case studies below, we will only be examining sampling distributions of means, so you will be using the sampleTestScores() function defined above and plotting a histogram of each sampling distribution. **Use abline() to mark each histogram with two vertical lines to designate the 95% central region and the lower (0.025) and upper (0.975) tails. Also mark the specified mean, using a different colored line.** Each of the sampling distributions has a story that goes with it, where researchers worked with a new group of test takers who may or may not be similar to the calibration population. Your job is to answer these five questions about each of the sampling distributions that you create:

1. What is the mean of the sampling distribution?
2. What is the lower bound of the central region – that is, the level of test score that sets off the left-hand tail from the central region?
3. What is the upper bound of the central region – that is, the level of test score that sets off the right-hand tail from the central region?
4. With respect to the new sample mean that is presented as part of the story, does that mean fall in the left-hand tail, the central region, or the right-hand tail?
5. Based on your answer to Question #4 do you believe that the new mean was drawn from the same population as the means in the sampling distribution?

**The case studies begin on the next page.**

**Case Study A:**

A sample of 100 students from Syracuse, NY took a standardized test. The mean of the calibration population was 100 and the sd was 10. The sample mean for this group of students was . Write code to display a histogram of the sampling distribution and add the appropriate markings. Add a comment that answers questions four and five.

1. Mean of the sampling distribution?
2. Lower bound of the central region?
3. Upper bound of the central region?
4. Where does the new mean fall?
5. Was the new sample drawn from the population that created the sampling distribution?

**Case Study B:**

A sample of **n=49** students from New York, NY took a standardized test with a population mean of 100 and sd of 10. The sample mean for this group of students was . Write code to display a histogram of the sampling distribution with the appropriate markings. Add a comment that answers questions four and five.

1. Mean of the sampling distribution?
2. Lower bound of the central region?
3. Upper bound of the central region?
4. Where does the new mean fall?
5. Was the new mean drawn from the population that created the sampling distribution?

**Case Study C:**

In the final stages of an FDA trial, researchers gave dietary supplements to a sample of 500 students to improve their ability to memorize words. The students then took a standardized test (population mean=100, sd=10). The sample mean for these students was .

1. Mean of the sampling distribution?
2. Lower bound of the central region?
3. Upper bound of the central region?
4. Where does the new mean fall?
5. Was the new mean drawn from the population that created the sampling distribution?

**Share your code and comments so far**

**Bonus Case Study D:**

Researchers investigating the effects of natural disasters on human development located n=64 children who had recently survived an earthquake or flood. These children took a standardized test (**population mean = 50, sd = 5**; **Note: You will have to generate a new testPop!)** The sample mean for this group of children was .

1. Mean of the sampling distribution?
2. Lower bound of the central region?
3. Upper bound of the central region?
4. Where does the new mean fall?
5. Was the new mean drawn from the population that created the sampling distribution?

**Super Bonus Case Study E:**

Researchers from Case Study D followed up on the affected children three years after the natural disaster that had affected them. The researchers calculated difference scores between the first and second round of testing. A positive difference meant that a child’s test score had improved. Because some families had moved, the researchers were only able to follow up with a sample of n=36 of the original group. The mean improvement in test scores for this sample of n=36 children was .

1. Mean of the sampling distribution?
2. Lower bound of the central region?
3. Upper bound of the central region?
4. Where does the new mean fall?
5. Was the new mean drawn from the population that created the sampling distribution?

Share your final code and comments