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

Instructions: Recall that at the conclusion of the previous exercise, you ran code that created a pair of histograms that looked like this:



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

In this exercise, we will only be examining sampling distributions of means (like the lower half of the diagram above). Each sampling distribution will be marked with two vertical lines to designate the central region and the lower (0.025) and upper (0.975) tails. Each of the sampling distributions also has a story that goes with it, where researchers have 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 is presented:

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 other means in this sampling distribution?

**Case Study A:**

A sample of 100 students from Syracuse, NY took a standardized test (mean of the calibration population = 100). The sample mean for this group of 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?

**Case Study B:**

A sample of 49 students from New York, NY took a standardized test (population mean = 100). The sample mean for this group of 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?

**Case Study C:**

In the final stages of an FDA trial, researchers gave dietary supplements to a sample of 144 students to improve their ability to memorize vocabulary words. The students then took a standardized test (population mean = 100). 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?

**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 = 100). 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?

**Case Study E:**

Researchers investigating the effects of natural disasters on human development 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 n=36 of the original group. The sample mean improvement in test scores for this group 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?

**Code for Creating These Histograms**

set.seed(1234) # Control randomization

testPop <- rnorm(100000, mean=100, sd=10) # Create simulated population of test scores

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

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))

# Case Study A

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,prob=0.025))

abline(v=quantile(samplingDistribution,prob=0.975))

# Case Study B

samplingDistribution <- replicate(1000, mean(sampleTestScores(49)))

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,prob=0.025))

abline(v=quantile(samplingDistribution,prob=0.975))

# Case Study C

samplingDistribution <- replicate(1000, mean(sampleTestScores(144)))

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,prob=0.025))

abline(v=quantile(samplingDistribution,prob=0.975))

# Case Study D

samplingDistribution <- replicate(1000, mean(sampleTestScores(144)))

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,prob=0.025))

abline(v=quantile(samplingDistribution,prob=0.975))

# Case Study E

samplingDistribution <- replicate(1000, (mean(sampleTestScores(36)-mean(sampleTestScores(36)))))

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,prob=0.025))

abline(v=quantile(samplingDistribution,prob=0.975))