**Your Name: Bastin Bajiyo Job Partner’s Name: Hendi Kushta**

**IST772 Week 3 Class Exercise - Reasoning with the Central Region and Tails of a Sampling Distribution**

**Instructions: Post this document with your code, results, and answers to all questions in Blackboard.**

**Part 1**

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. **Paste that graph below.**

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)}

# Now generate many samples of size n

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

# These 4 lines of code will generate histograms.

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.

Chart, histogram

Description automatically generated

**Part 2**

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.

**Include the histogram of the sampling distribution for each Case Study.**

**Case Study A:**

The population mean for a Statewide standardized test was 100 and the sd was 10. A sample of 100 students from Syracuse, NY took this standardized test.

The sample mean (**new mean**) for this group of Syracuse students was .

1. Construct the sampling distribution with ablines marking the central region. (This is the same sampling distribution as in Part 1.)

* Central region is the inner 95% cases marked by blue lines.

1. Mark the **new mean** with an abline as well.

* New mean is 101, marked by the red line

1. Paste the R code and graph here.

Chart, histogram

Description automatically generated

1. Based on your graph, do you believe that the 100 Syracuse students were drawn from the same population as those that generated the sampling distribution, or do they represent a different population? **Explain.**

* Yes, since the new mean fall inside the central region, we can safely assume that the 100 students were drawn from the same population.

**Case Study B:**

A sample of **n=49** students from New York, NY took this standardized test. The sample mean for this group of students was .

1. Construct the sampling distribution with ablines marking the central region.

* The central regions are marked by blue lines, giving the inner 95% cases.

1. Mark the **new mean** with an abline as well.

* New mean is marked with red line.

1. Paste the R code and graph here.

Chart, histogram

Description automatically generated

1. Based on your graph, do you believe that the 49 students were drawn from the same population as those that generated the sampling distribution, or do they represent a different population? **Explain.**

* **No,** since the new mean falls outside the central region (95%), the 49 students are very less likely to be drawn from the same population.

**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. Construct the sampling distribution with ablines marking the central region.

* The central regions are marked by blue lines, giving the inner 95% cases.

1. Mark the **new mean** with an abline as well.

* New mean is marked with red line.

1. Paste the R code and graph here.

Chart, histogram

Description automatically generated

1. Based on your graph, do you believe that the 500 students were drawn from the same population as those that generated the sampling distribution, or do they represent a different population? **Explain.**

* **No,** since the new mean falls outside the central region (95%), the 500 students are less likely to be drawn from the same population.

**Case Study D:**

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

1. Construct the sampling distribution with ablines marking the central region.

* The central regions are marked by blue lines, giving the inner 95% cases.

1. Mark the **new mean** with an abline as well.

* New mean is marked with red line.

1. Paste the R code and graph here.

Chart, histogram

Description automatically generated

1. Based on your graph, do you believe that the 64 children were drawn from the same population as those that generated the sampling distribution, or do they represent a different population? **Explain.**

* **No,** since the new mean is far outside the central region (95% cases), the 64 children are nearly unlikely to be drawn from the same population.

**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 second and the first 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. Construct the sampling distribution with ablines marking the central region.

* Central regions are marked by blue lines.

1. Mark the **new mean** with an abline as well.

* The sample mean here is 48 +3 = 51, marked by red line

1. Paste the R code and graph here.

Chart, histogram

Description automatically generated

1. Based on your graph, do you believe that the 36 children were drawn from the same population as those that generated the sampling distribution, or do they represent a different population? **Explain.**

**Yes,** since the new mean falls inside the central region (95% cases), we can safely assume the 36 students were drawn from the same population.