ACTIVITY 4: Sampling Variability Revisited through Graphics and the Sampling Distribution

OVERVIEW: In this activity, the class will begin by discussing the difference between the population distribution and the sampling distribution for the mean. Students will use Excel to create a side-by-side boxplot using data they calculated in Activity 3 along with additional data provided by the instructor. Students will reflect on the results through their responses to questions that encourage them to think critically about the information their graphic displays. In an all-class discussion, the instructor will go over the worksheet with the students making sure they understand the difference between the population distribution and the sampling distributions for the various sample size groups. At the end of class, the instructor will present a brief lesson on the goal of sampling to tie together what they have learned thus far.

MATERIALS NEEDED TO CARRY OUT THE LESSON:

- In-Class Activity Worksheet
- Laptop with Excel
- Excel Spreadsheet with additional data

CONCEPTS TO BE LEARNED/APPLIED:

- Students will understand the difference between a population distribution and a sampling distribution.
- Students will understand that the variability of a sampling distribution decreases as the sample size increases.
- Students will understand that increasing the sample size will reduce the variation from one sample mean to the next, which results in a good estimate of the population mean regardless of which sample was selected.

PREREQUISITE CONTENT KNOWLEDGE:

- Students need to know how to interpret boxplots.

INSTRUCTIONAL PLAN:

The instructor will begin by explaining the difference between a population distribution and a sampling distribution.

Vocabulary to present:

The **population distribution** of a variable represents the distribution of its values across all members of a population.

The **sampling distribution**, on the other hand, pertains to the distribution of a statistic. For instance, if the statistic is the sample mean calculated from a simple random sample of 5 countries, the sampling distribution is derived by computing the sample mean for all

possible simple random samples of five countries from the total of 195 countries in the population, and then graphing these means. Technically, this involves 223,1243,664 possible sample means.
Once the instructor goes over the vocabulary, the students are ready to complete the activity. When they are finished with the activity, have an all-class discussion to make sure they answered the questions appropriately.

After the questions have been discussed, revisit the applet from Activity 3.

https://www.geogebra.org/m/MfrKDWyV

Explain to them that the top distribution provided in the applet is the population distribution. If one chooses a sample size of *n*, the applet repeatedly selects SRS of size *n* from the population distribution, calculates the mean for that SRS, and plots it on a given picture. This creates a sampling distribution. Once again, play around with the sample size so the students can see that the sampling distribution has less and less variability as *n* increases.

At the end of the activity, discuss the takeaways from the first four activities as a class. Make sure the students include the "Goals for Sampling" below.

Goals for Sampling:
In the first four activities, we looked at the concepts of a representative sample and sampling variability. When statisticians want to find an estimate for a population parameter, they take a sample and calculate a statistic to use as an estimate. Their estimate must have low variability and low bias for it to be accurate. **When you are reading statistical studies online, remember that if you cannot determine how they selected the units or individuals in the study to ascertain if a random sample was taken; are not told how the questions are phrased or the data is collected; or are not given a sample size, you should be weary of the accuracy of the data.**