

ACTIVITY 3: Sampling Error, Sampling Variability

OVERVIEW: In this activity, the class will begin with the instructor discussing the difference between a statistic and a parameter. The formula for the sample mean and sample standard deviation, as well as instructions on how to calculate these measures using technology, will be provided. Students will use these instructions to calculate the measures by hand for the small sample and use technology for the larger samples of data they found in Activity 3. They will write their results in a table on the board and on their worksheet for comparison. The instructor will also have the population mean value for the students to compare their sample means to. Students will then answer questions regarding their findings to enhance their critical thinking skills. After the students have answered these questions, the instructor will present the definition of sampling error and sampling variability using the students' data as an illustration. As a class, an applet that introduces the idea of sampling variability will be explored. The class will end by having students reflect on the activity and having them compile additional questions they should ask themselves when reading research online.

MATERIALS NEEDED TO CARRY OUT THE LESSON:

- In-Class Activity Worksheet
- Laptop with Excel

CONCEPTS TO BE LEARNED/APPLIED:

- Students will understand that a statistic is a measure that is calculated from a sample and used as an estimate for the parameter which is the measure for the population.
- Students will understand that the sample mean (\bar{x}) and the sample standard deviation (s) are statistics.
- Students will understand the variability inherent in the sampling process, and how that causes sampling error.
 1. Students will understand that the sample they select will impact their estimate.
 2. Students will understand that a larger sample size will reduce the amount of sampling error and the variability of the statistic.
- Students will understand how the estimations relate to the population parameter they are estimating and how to ensure a good estimate.
 1. Students will understand that a large sample size is important because it reduces the variability of the statistic thus increasing the precision.
 2. Students will understand that the likelihood of obtaining a sample with all countries with low life expectancies is small when the sample size is large enough.
 3. Students will understand that increasing the sample size and reducing bias will make it more likely to get a representative sample.

INSTRUCTIONAL PLAN:

The instructor will begin by explaining the difference between a statistic and a parameter. They will then give two examples of statistics, the sample mean (\bar{x}) and the sample standard deviation(s).

Vocabulary to present:

A **parameter** is a number that describes the **population**. A parameter is a fixed number, but in practice its value is unknown.

A **statistic** is a number that describes a **sample**. The value of a statistic is known when a sample is taken, but statistics vary from sample to sample.

A **statistic**, which varies from sample to sample, is used to estimate a **parameter**.

Two statistics that will be explored in this lesson.

1) The sample mean is the sum of measures taken from a sample divided by the total number of measures. It is used as an estimate for the population mean μ .

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{(x_1 + x_2 + \dots + x_n)}{n}$$

2) The sample standard deviation is the square root of the sum of the squares of the differences between each value and the mean. It is used as an estimate for the population standard deviation σ .

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}} = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

Note: The instructor may suggest the students keep their intermediate answers with a specific number of decimal places. Since the final answer needs to contain one additional decimal place to that of the original data, it is recommended that the intermediate answers contain at least two additional decimal places to prevent excessive roundoff error.

Students will have instructions to calculate both these measures by hand and using Excel. The students will follow the instructions on the worksheet using the data they collected in Activity 2.

Once students have their calculations completed, write the values that they found for the sample mean on the board for comparison. The students should record this data on the table in their Worksheet as they will refer to it when answering the questions. The students should also save the mean their group calculated as it will be used in Activities 4 and 5.

Explain to the students that all the values from the different groups on the board are estimates for the mean of the population calculated from samples of various sizes. Then have them answer the questions on their worksheet concerning the data.

When they have finished answering the questions, introduce the concept of sampling variability and sampling error using their data as an example.

Vocabulary to present:

Sampling Variability represents the variation associated with the value of the statistic that is generated by repeatedly selecting samples of the same size while using the same probability sampling design from the population. For example, suppose I consider the students enrolled at your home institution as my population and randomly select 30 students to calculate an estimate of the average age for all students. If I were to repeat the process multiple times, the estimate of the average age I obtained would differ from sample to sample.

Sampling Error results from random chance and is not bias. When a sample is selected from a population, an estimate is calculated. This estimate will not necessarily equal the value in the population of interest. This does not mean that bias has occurred.

As a class, explore the applet below that introduces the idea of sampling variability.

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

Change the sample size on the applet to show them that the sample means will stay more closely clustered around the population mean with bigger sample sizes.

At the end of the class, students should reflect on the activity and compile additional questions they should ask when reading research online.