ACTIVITY 2: Understanding the Simple Random Sample and its Limitations

OVERVIEW: In this activity, the class will begin with a discussion of the definitions of a unit or individual, a statistical variable, a simple random sample, and how such a sample would be obtained. Students will also be asked if they think that a simple random sample is possible in all scenarios. After that, students will be broken into groups of two and their group will be given a population list of all the countries in the world. Each group must take a simple random sample of 5 countries; a separate simple random sample of 15 countries; and another simple random sample of 30 countries. Once they have selected their countries, they will be asked to look up the life expectancies for these countries from two different websites given by the instructor. Once the students have collected their data, students are asked to answer questions that will lead to a discussion on the three types of bias: undercoverage, non-response, and response bias. Lastly, the students will reflect on the activity to help them generate questions to ask themselves when reading statistics online.

MATERIALS NEEDED TO CARRY OUT THE LESSON:
- In-Class Activity Worksheet
- Population list (list of all countries)
- Laptop or iPad
- Access to the internet

CONCEPTS TO BE LEARNED/APPLIED:
- Students will understand to obtain a simple random sample, they will:
  - need a list of the population with each unit or individual numbered
  - use a random number generator to generate numbers
  - consider the units or individuals that correspond to the random numbers selected in the sample.
- Students will understand the limitations of a simple random sample. Firstly, the simple random sample can only be used when the population can be numbered (reasonably counted). Secondly, even with a simple random sample, it is possible to obtain biased data.
- Students will understand how to identify various types of bias that can occur even when the sampling process is random: undercoverage, non-response, and response bias.
- Students will understand that undercoverage bias can be reduced by making sure you have a complete population list.
- Students will understand that non-response bias can be reduced by offering incentives, calling back, etc.
- Students will understand the importance of accuracy and wording to reduce response bias.

INSTRUCTIONAL PLAN:
The instructor will begin by defining the variable, unit, or individual, and a simple random sample. They will then explain the process of selecting a simple random sample.

Vocabulary to present:

The **variable** is a characteristic, a number, or a quantity that can be measured or counted, and whose value can change. It is what is recorded for each unit.

Note: In algebra, a variable is a letter or symbol that denotes a quantity that may change within a problem. In statistics, a variable is a measure or characteristic that changes from one unit to the next. Notice that the variable in statistics is not necessarily a number.

The **unit/individual** represents a person, animal, or object upon which the response variable or variable of interest is measured. (The unit is always what you are randomly selecting).

A **simple random sample (SRS)** is a sample that is selected from the population in such a way that every SRS of n units has an equal chance of being selected.

How do we select a simple random sample (SRS)?

1. Obtain a list of all the units/individuals in the population.
2. Give each unit in the population a number.
3. Use a random number generator to generate random numbers that correspond to units in the population. For each number generated, add the unit assigned to that number to your sample. If the same number is generated, skip it, and generate another number until there are n unique numbers where n is the sample size.

After presenting the above, ask students the following questions:

Are there populations for which you could not take a simple random sample? If the students are unsure, ask them some of the following questions.

- Do you think you could take a simple random sample of water samples from Lake Arcadia? Why or why not?

    Answer: You would have to take samples using all the water from the lake so that you could number each sample and hence the lake would have to be drained.

- Do you think you could take a simple random sample from all individuals in the US? Why or why not?

    Answer: No, the population is too large.

Explain to them that other types of samples would work better for these scenarios and that statisticians create new sampling techniques to gather good representative samples for situations like these all the time. In class, however, the focus will be on the SRS.

After the discussion, hand out the list of all the countries and the worksheet.
Demonstrate how to use the following random number generator.

https://www.random.org/

The list they were given contains 195 countries, so they should put the following into the generator:
Min: 1
Max: 195

They should then keep clicking "Generate" till they have 5 unique numbers. The five countries that correspond to the 5 numbers they generated will be the countries for their first sample of 5.

They should repeat the process to generate another sample of 15 countries and another new sample of 30 countries.

Note: If the instructor has limited time, they may only ask their students to generate samples of 5 and 30 countries.

After they have obtained their three samples, they will look up the life expectancies on the following two websites and fill in the tables on the worksheet. They will be recording the life expectancy for all individuals for 2022 from the following two websites.

https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy

Note: To help students find the countries in their sample, they may alphabetize the data by clicking on the arrow in the 1st column labeled "Countries & Territories."

https://www.cia.gov/the-world-factbook/about/archives/2022/field/life-expectancy-at-birth/country-comparison

Note: To help students find the countries in their sample, suggest that they can do a search by using "Control-F."

After collecting the data, they can answer the questions that follow and decide whether they encountered undercoverage, non-response, or response bias during the data collection process. End the class with the students coming up with more questions they should ask themselves when reading statistical studies.

Important Note: Students will need to keep the data that they generated for future Activities in the CoRD.