Using Dollar Street to Facilitate Conceptual Understanding of Confidence Intervals

GAISE Learning Goals:

3. Students should be able to produce *graphical displays and numerical summaries* and interpret what graphs do and do not reveal.
7. Students should demonstrate an understanding of, and ability to use, basic ideas of *statistical inference,* both hypothesis tests and interval estimation, in a variety of settings.

Performance Objectives:

- Identify categorical and measurement variables.
- Generate appropriate numerical summaries and graphical displays
- Produce approximate confidence intervals from sample data
- Use confidence intervals to make inferences about population
- Discuss limitations of confidence intervals in light of sample size and sampling process

Approximate Time to Complete:

- 75 minutes if done during one class period, though this can be adapted so that students do the activity outside of class, and then discuss the activity in class.

Resources Needed:

- Basic statistical analysis tools (e.g. Excel, R, JASP)
      Note: It's possible to complete this assignment with only a calculator if instructors prefer that students do basic calculations and graphic displays by hand.
- Internet/Computer to access Dollar Street
      (https://www.gapminder.org/dollar-street/matrix)

Recommended Courses:

- Undergraduate Introduction to Statistics course
- High School AP Statistics course

Skills Required:

- Students should be able to calculate means, proportions, estimated standard error of mean, and estimated standard error of proportion
- Students should have basic understanding of the relationship between samples and populations, statistics and parameters.
- Students should be able to create a simple bar graph or histogram.

Instructor Directions

Introduction:

This lesson is designed to introduce students to the concept of confidence intervals for population means and proportions. Students will use Dollar Street, a website that allows individuals to explore income and how people live around the world using both numbers and pictures as data, to identify a variable, select a sample, and generate a confidence interval in order to make inferences about their chosen population. In addition, students will think about the limitations of their inferences in light of their sample size and the sampling process.

Part One: Identification of Variable and Selection of Sample, 15 minutes

- Introduce Dollar Street to students and explain how it works and where the data comes from. Briefly remind students about the difference between random sampling processes and the non-probabilistic sampling processes used on Dollar Street (e.g. convenience and volunteer response sampling).
- Allow students five minutes to explore the website and to click on several pictures of interest.
- Hand out student instructions, and direct students to choose a variable of interest. They should identify whether it is a categorical or measurement variable. At first, students may be stuck and may only be able to identify obvious variables like income or country of origin. You might want to give students some examples of how to use the photos to generate variables.

Example Categorical Variables	Example Measurement Variables
Does their drinking water come from a faucet or not?	What is the income?
Is their comb/brush missing teeth or broken?	How many people are in a family?
Are there pillows on the bed or is the bed on a frame or on the floor?	How many teeth are in view in a smile?
Does the toilet have a flusher?	How many toys are in picture?
What type of pet do they have?	How many books in picture?
Do the toys have faces?	How many windows in house?
Is there a gap in the front teeth?	How many chairs at the table?

Instructor Tip: Use this time to reinforce the importance of operationalization. How exactly will they measure their variable? Could two people operationalize the same variable differently? For example, if I choose the categorical variable presence/absence of jewelry on hand, how am I defining jewelry? Does a tattoo count? Or nail art? How does understanding the way that a researcher operationalizes their variable help those reading the research better make sense of the findings?

- Have students chose a population (or populations) of interest and then confine the data in Dollar Street to show only their chose variable from the sample of interest (e.g. select income for Cambodia, or toys for Europe).

Instructor Tip: Due to the small sample sizes, it might be best for students to choose continents rather than countries. This is a good opportunity to talk about the importance of sample size in statistical research. Aim for a sample size of at least 30.

Part Two: Compute Summary Measures and Generate Graphical Display, 25 minutes

- Have students record data on student instruction sheet.
- Have students create appropriate graphical display for their variable using statistical software or by hand.
    - If categorical, students should create a bar chart.
    - If measurement, students should create a histogram.
- Direct students to calculate appropriate sample summary measures for their data using statistical software or by hand.
    - If categorical, students should calculate the sample proportion and the estimated standard error of proportion.
    - If measurement, students should calculate the sample mean and the estimated standard error of the mean.

Instructor Tip: This is a great opportunity to talk about the purposes of sample statistics, the concept of sampling variability, and how they are used to estimate population proportions. In the context of this activity, students can begin to think about how the data from Dollar Street can allow them to get a better idea about the populations that these data are pulled from (e.g. the sampled data from Asia shows that the proportion of people who wear jewelry is .31; what might this tell me about the population of Asia as a whole in terms of how many people wear jewelry? What if I had a different sample? Would I get a different sample proportion? How can I infer about the population if there is variability amongst my sample statistics?

Part Three: Confidence Interval Construction and Interpretation, 35 minutes

- Have students create a 95% confidence interval for the population mean/proportion using their sample statistic.

Instructor Tip: Depending on the level of the class, the students can find the correct t-multiplier for a 95% confidence interval, or they can find the approximate 95% confidence interval using 2 as the multiplier. In addition, you can have students change their level of confidence and discuss what happens to interval when they become more or less confident.

- Have students write a statement interpreting their confidence interval (e.g. I'm 95% confident that this interval contains the true population proportion of people in Asia who wear jewelry on their hands).
- Have students share their results with the class. This not only allows students practice in communicating their statistical results with others, but it gives students experience with several variables/data that can be culled from Dollar Street.

Instructor Tip: As part of their presentation, students can share the sample size they used, and they compare the width of their resulting confidence interval. As students begin to see that larger N sizes result in narrower intervals, instructors can discuss why this makes sense theoretically.

- Discuss as a class the usefulness of their confidence intervals as well as their limitations. For example, how does the sampling process used make a difference? Could there be something unique about using a convenient sample versus using a simple random sample? This could also lead to a discussion about how Dollar Street can be used by the general public to increase global knowledge about differences/similarities among countries and continents in the world.

  Instructor Tip: This debriefing process is a good opportunity to remind students that often we do not know the population parameter that we are trying to estimate. As such, we don't know whether our resulting interval is reflective of the 95% that covers the true population parameter or the 5% that doesn't cover the true population parameter. Discuss with students how this knowledge should inform the generalizations we make from data.

Alternate Instructions/Lesson Extensions:

1. Instead of having students choose their own variables, instructors could choose one variable for the whole class. Then, each student could choose a unique sample and create a confidence interval using that sample (e.g. one student looks at jewelry on the hand for the United States, another for Asia, another for France and Cambodia, etc.). Then, each confidence interval can be used to make an inference about the world population. Instructors can display class confidence intervals on the board and look for overlap, using this an opportunity to better understand the concept of 95% confidence as the idea that 95% of intervals created from samples should cover the true population parameter. Again, instructors would want to talk about the limitations/context of confidence intervals related to sample size and sampling process.

2. Students could compare continents on their variable by creating an interval from two different samples for two different populations. For example, students could create a confidence interval for the proportion of people in Asia who wear jewelry on their hands and another one for the proportion of people in Africa who wear jewelry on their hands. This is a great opportunity to practice comparing confidence intervals and looking to see whether there is overlap in the intervals or not, and how this leads us to determine whether we are 95% confident that a greater (or less) proportion of people in Asia wear jewelry on their hands than in Africa, or whether the confidence intervals result in a sort of statistical tie from which we cannot tell for sure which country has a greater proportion of people wearing jewelry.

Using Dollar Street to Facilitate Conceptual Understanding of Confidence Intervals
## Student Instructions

*For this assignment, you will be using the website Dollar Street (*https://www.gapminder.org/dollar-street/matrix*) to practice creating and interpreting confidence intervals. You should familiarize yourself with the website and then refer to the website to complete the following questions.*

1. What is your population of interest?

2. Choose a sample and identify a variable to measure on the sample.

3. Is your variable categorical or measurement? How do you plan to operationalize your variable?

4. Record your data in the space below.

5. What is your sample size?

6. Create a visual display for your data.

7. What is the sample mean/proportion of your data?

8. What is the estimated standard error of the mean/proportion for your data?

9. Using your sample data, construct a 95% confidence interval using 2 as your multiplier.

10. Write a sentence interpreting your confidence interval.

11. Take a moment to think about your confidence interval. How is it useful in learning about your population of interest? Can you think of any limitations/things to be aware of when using your confidence interval to make inferences about the population? Do you have any questions about the activity or the use of confidence intervals that you would like to have answered?

Using Dollar Street to Facilitate Conceptual Understanding of Confidence Intervals
Example Student Assignment

For this assignment, you will be using the website Dollar Street (https://www.gapminder.org/dollar-street/matrix*) to practice creating and interpreting confidence intervals. You should familiarize yourself with the website and then refer to the website to complete the following questions.*

1. What is your population of interest? Asia

2. Choose a sample and identify a variable to measure on the sample.
   Households from Asia that are on Dollar Street will make up my sample, and my variable will be whether or not there is jewelry present on the hands in the picture.

3. Is your variable categorical or measurement? How do you plan to operationalize your variable?
   My variable is categorical. I will select the "hands" variable on Dollar Street, and I will count any rings/bracelets/wristwatches as jewelry. I will not count painted nails or tattoos as jewelry.

4. Record your data in the space below.

   46 of the pictures show hands with jewelry.

5. What is your sample size?

   148 (there were 148 total pictures in my sample)

6. Create a visual display for your data.

Presence or Absence of Jewerly in Sample from Asia
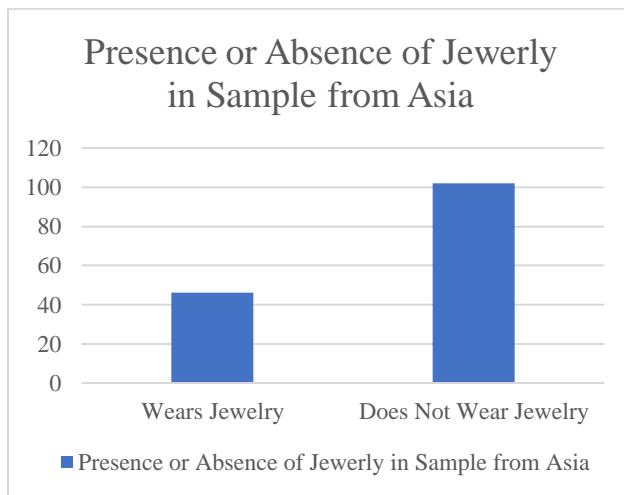
7. What is the sample mean/proportion of your data?
   My sample proportion is .31.

8. What is the estimated standard error of the mean/proportion for your data?

   The standard error of proportion is .038.

9. Using your sample data, construct a 95% confidence interval using 2 as your multiplier.

   .31 +/- 2 (.038) = .234 to .386

10. Write a sentence interpreting your confidence interval.

    I am 95% confident that the interval .234 to .386 contains the true population proportion of people in Asia who wear jewelry.

11. Take a moment to think about your confidence interval. How is it useful in learning about your population of interest? Can you think of any limitations/things to be aware of when using your confidence interval to make inferences about the population? Do you have any questions about the activity or the use of confidence intervals that you would like to have answered?

    Using confidence intervals helps me to use my sample statistic to get a better idea of the population parameter. I don't understand why I used 2 as a multiplier.

Extension Number 2 Example:

Below are my 95% confidence intervals for the proportion of people that wear jewelry on their hands for 4 different continents. To construct each interval, I used a sample from each continent.

Asia:
Sample size: 148
Confidence interval: .234 to .386
Europe:
Sample size: 51
Confidence interval: .198 to .462
Americas:
Sample size: 54
Confidence interval: .068 to .272
Africa:
Sample size: 72
Confidence interval: -.041 to .222

Conclusions: There is a lot of overlap between the intervals, so I can't always tell which country has a higher proportion of people who wear jewelry on their hands. However, because Africa's confidence interval is completely below Asia's confidence interval, I can be 95% certain that Asia has a higher proportion of people who wear jewelry on their hands than Africa.

General Grading Rubric:

Rating	Description
Excellent	Student completes entire assignment with no, or very minor, errors. Student correctly interprets confidence interval and communicates the appropriate inference both in writing and when sharing to with the class. Student has a solid understanding of the usefulness and limitations of confidence intervals.
Satisfactory	Student completes entire assignment with some small errors. Student correctly interprets confidence interval and communicates it appropriately in writing, but makes some minor mistakes when communicating with the class. Student may not understand all of the limitations of confidence intervals, but identifies at least one and clearly identifies their usefulness.
Questionable	Student completes majority of the assignment with several errors. Student can construct confidence interval, but misinterprets confidence interval both in writing and in communication with class. Student doesn't seem to fully understand usefulness of confidence intervals and how to use them to estimate population parameter and no understanding of the limitations of confidence intervals.
Unsatisfactory	Student does not complete the assignment as given and/or has striking errors in completion. Student cannot construct a confidence interval from the data nor can interpret one in either writing or in oral communication. Student has no understanding of the usefulness or the limitations of using confidence intervals to make inferences about the population.