Lesson 12: Larger Populations

Goals

- Comprehend the terms "population" and "sample" (in spoken and written language) to refer to the whole group and a part of the group under consideration.

- Describe (orally and in writing) a sample for a given population.

- Explain (orally) that a sample may be used when it is unreasonable to gather data about an entire population.

Learning Targets

- I can explain why it may be useful to gather data on a sample of a population.

- When I read or hear a statistical question, I can name the population of interest and give an example of a sample for that population.

Lesson Narrative

This lesson introduces the idea of using data from a *sample* of a *population* when it is impractical or impossible to gather data from every individual in the populations under study. Students consider whether the people in their class would be an adequate sample for several different questions and associated populations (MP3). While all of the answers given in this lesson are samples and may have some benefit to them, most of them are not the best way to select samples. In the next lesson, students learn about what makes some samples more representative of a population than others. In later lessons, students explore the best ways to try to obtain such samples.

Alignments

Addressing

- 7.SP.A.1: Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.

- 7.SP.B: Draw informal comparative inferences about two populations.

Instructional Routines

- MLR2: Collect and Display

- MLR8: Discussion Supports

- Think Pair Share

Required Preparation

Compute the mean and MAD for the length of the preferred names (if students do not go by their first name, use their nickname, middle name, etc.). Do the same for the last names of students in the class prior to the John Jacobjingleheimerschmidt activity.

> **Student Learning Goals**
>
> Let's compare larger groups.

12.1 First Name versus Last Name

Warm Up: 5 minutes
The purpose of this warm-up is for students to begin to see the need for samples of data when the population is too large. In this activity, students are asked to think about a question involving all the students at their school and compare the question to an earlier lesson in which the population was small and it was easy to obtain data for the entire population.

Addressing

- 7.SP.B

Launch

Give students 2 minutes of quiet work time followed by a whole-class discussion.

> **Student Task Statement**
>
> Consider the question: In general, do the students at this school have more letters in their first name or last name? How many more letters?
>
> 1. What are some ways you might get some data to answer the question?
>
> 2. The other day, we compared the heights of people on different teams and the lengths of songs on different albums. What makes this question about first and last names harder to answer than those questions?

Student Response

Answers vary. Sample responses:

1. Get a list of everyone in the school, count how many letters there are in each person's first and last names, and calculate the mean of each data set. Then subtract the means and divide this difference by the mean absolute deviation, to measure how much the data sets overlap. Alternatively, total all the letters for the first names and all the letters for the last names in the school and compare these sums. Another idea would be to survey some of the students in the school to use their data to make an informed guess about the whole school.

2. There are a lot more students in the school than there were people on the teams or songs on the albums that we looked at, so it would be a lot more work to calculate the means and mean absolute deviations for everyone's names.

Activity Synthesis

The purpose of the discussion is to highlight the methods of getting data for the school more than it is the method of computing the answer.

Select some students to share their responses.

Students who have elected to sum all the letters in the first names in the school and all the letters in the last names in the school may note that it is a simple comparison to tell whether there are more in first or last names, since you get one single large number for each group. (Comparing data sets.)

Students who have elected to calculate the mean for each group and use MAD as a method of comparison may note that while the calculations may take more time, they give you more precise information, such as knowing about how long first names and last names are, as well as a way to compare the two sets. (Using the general rule from the previous lesson.)

Students who suggest surveying a small group of students may point out that it would be easier to do the calculation with a smaller group. The information would not be as accurate, but it would take a lot less time and might give a good general idea. It would depend on how accurate you needed your answer to be. (Introduction to sampling.)

12.2 John Jacobjingleheimerschmidt

10 minutes
In this activity, students are asked to compare two groups (length of preferred names and last names) by collecting data from the class. They are asked if the data from the class gives enough information to draw a conclusion about a larger group (MP3). In the following activities, students will be introduced to the idea of sampling. This activity gives students the first chance to experience why sampling might be needed.

Addressing

- 7.SP.A.1

- 7.SP.B

Instructional Routines

- MLR8: Discussion Supports

Launch

Compute the mean and MAD for the number of letters in each student's preferred name (if students do not go by their first name, you may use their nickname, middle name, etc.). Do the same for their last names.

Give students 1 minute of quiet work time for the first 2 questions followed by a quick display of information then 5 more minutes of quiet work time and a whole-class discussion.

If a digital solution is available, input the data for the class to find the mean and mean absolute deviation for each data set. If a digital solution is not available, this information should be calculated based on the class roster prior to this activity. After students have had a minute to work on answering the first two questions, provide students with the mean and MAD for the names in the class.

Tell students that if they have a preferred name other than their official first name (nickname, middle name, etc.) they may use this in place of the first name.

Student Task Statement

Continue to consider the question from the warm-up: In general, do the students at this school have more letters in their first name or last name? How many more letters?

1. How many letters are in your first name? In your last name?

2. Do the number of letters in your own first and last names give you enough information to make conclusions about students' names in your entire school? Explain your reasoning.

3. Your teacher will provide you with data from the class. Record the mean number of letters as well as the mean absolute deviation for each data set.

   a. The first names of the students in your class.

   b. The last names of the students in your class.

4. Which mean is larger? By how much? What does this difference tell you about the situation?

5. Do the mean numbers of letters in the first and last names for everyone in your class give you enough information to make conclusions about students' names in your entire school? Explain your reasoning.

Student Response

Answers vary. Sample response:

1. First name: 5. Last name: 8.

2. No, looking at just one person's name does not give enough information to answer about everyone in the school, because some students have longer or shorter names than others.

3.   a. Mean: 6.2. MAD: 2.1

   b. Mean: 7.3. MAD: 2.8

4. The mean number of letters in our last names was larger, by 1.1 letters, which is about 0.4 times the MAD. This means that the last names are longer, but not by a lot. Take the difference between the means, or $7.3 - 6.2 = 1.1$. Since the MAD measures the variability of the data set, this difference divided by the MAD gives a comparison. $1.1 \div 2.8 \approx 0.4$, which is a small

number. This small number means that although the last names had a higher mean number of letters, the two groups were not very distinct. Also, their dot plots would have a lot of overlap.

5. Maybe. There are still a lot of students that we did not count.

Activity Synthesis

The purpose of the discussion is for students to see how the data they have might relate to a larger group. In particular, that a sample might give some estimate of a larger population, but the estimate should not be assumed to be exact.

Consider asking these questions for discussion:

- "Do you expect the mean length of first names for the school to be exactly the same as the mean length for the class?" (Probably not exactly the same. It may be close, though.)

- "Do you expect the mean length of first names for the school to be much larger or smaller or about the same as the mean length for the class? Explain your reasoning." (Unless there are a few outliers in the class, it should be fairly close to the mean from the class.)

12.3 Siblings and Pets

10 minutes

In this activity, students think a little more deeply about the data we would like to know and how that compares to the data we can collect easily and quickly (MP1). They are presented with a statistical question that does not have an obvious answer. Students are then asked to consider ways they might begin gathering data to answer the question, but are asked to realize that the data they could reasonably collect is not everyone addressed by the question. Following the activity, the discussion defines the terms *population* and *sample*.

Addressing

- 7.SP.A.1

Instructional Routines

- MLR2: Collect and Display

Launch

Arrange students in groups of 2.

Set up the context by asking students, "Do people who are the only child have more pets?" then to provide a possible explanation for their answer. For example, maybe only children do have more pets because the family can better afford to take care of an animal with only 1 child. Maybe they do not because smaller families may live in smaller places and not have room for a lot of pets.

Give students 5 minutes of partner work time followed by 5 minutes of whole-class discussion.

> **Student Task Statement**
>
> Consider the question: Do people who are the only child *have more pets*?
>
> 1. Earlier, we used information about the people in your class to answer a question about the entire school. Would surveying only the people in your class give you enough information to answer this new question? Explain your reasoning.
>
> 2. If you had to have an answer to this question by the end of class today, how would you gather data to answer the question?
>
> 3. If you could come back tomorrow with your answer to this question, how would you gather data to answer the question?
>
> 4. If someone else in the class came back tomorrow with an answer that was different than yours, what would that mean? How would you determine which answer was better?

Student Response

Answers vary. Sample responses:

1. There is not enough variation in just the class to figure out the answers for all people who are the only child compared with all people who have siblings. For example, everyone in the class is about 12 or 13 years old, so we can't tell anything about adults or people from other generations or other parts of the world.

2. I could ask my classmates if they are an only child and gather data about their pets.

3. I could also ask my neighbors if they are an only child and gather data about their pets.

4. We might get different answers because we collected data from different people. Whoever asked more people or a wider range of people would have stronger evidence.

Activity Synthesis

The purpose of the discussion is to show the difference between the data we would like to have to answer the question and the data we have available.

Some questions for discussion:

- "If we had all the time and money in the world and wanted to answer this question, who would we need to collect data from?" (Everyone in the world.)

- "What would you do with the data collected from everyone to answer the questions?" (Find the mean and MAD of the data from the two sets and compare them like we did in previous lessons.)

- "Why is it unreasonable to actually collect all the necessary data to answer the question?" (There are too many people to collect data from. There is not enough time to get to everyone in the world, and I cannot travel everywhere.)

- "Since it may be difficult to guess an answer without doing *any* research, but we cannot get all of the data we want, what data could you get that would help estimate an answer?" (It would be good to ask a few people in different parts of the world and try to get different groups represented.)

Define population and sample. A **population** is the entire pool from which data is taken. Examples include (depending on the question) "all humans in the world," "all 7th graders at our school," or "oak trees in North America." In this usage, it does not have to refer only to groups of people or animals. A **sample** is the part of the population from which data is actually collected. Examples (related to the population examples) include "5 people from each country," "the first 30 seventh graders to arrive at our school," or "8 oak trees from the forest near our school."

Ask students, "What is the population for the question about only children and their pets?" (Everyone in the world.) Note that we would need data from everyone, including those who don't have pets or do have siblings.

Ask students, "What might be a sample we could use to answer the question?" (The students in our class, my neighbors, a few people from different countries.) After getting several responses, ask, "What might be the benefits and drawbacks of each of these samples?" (Some may be more convenient, but would not represent the population as well or vice-versa.)

Explain: While it is best to have data for the entire population, there are many reasons to use a sample.

- More manageable. With very large populations, the amount of data can be hard to collect and work with, so a smaller subset may still be informative and easier to work with. Example: Find the average size of a grain of sand.

- Necessary. Sometimes it is impossible to reach the entire population, so a sample is all that is available. Example: Find the average lifespan of tuna.

- Speed. Sometimes a rough estimate is all that is needed and a sample of data is enough to estimate the population characteristic. Example: Out of curiosity, what is the median number of apps on smartphones.

- Cost. Sometimes it is very costly to obtain the data needed, so a sample can reduce the cost. Example: Find the average amount of hydrogen in moon rocks.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Invite students to suggest language or diagrams to include that will support their understanding of: populations and sample.
Supports accessibility for: Conceptual processing; Language

Access for English Language Learners

Representing, Speaking: MLR2 Collect and Display. To help students make sense of the terms "sample" and "population", draw a diagram of a few circles inside a larger circle on a visual display. Label the large outer circle "population" and the small inner circles "sample." As students respond to the question "What is the population for the question about only children and their pets?", write the population on the visual display. As students respond to the question "What might be a sample we could use to answer the question?", write the samples in different inner circles on the visual display. Listen for and amplify words and phrases that define these terms, such as "part of" or "entire." This will help students visualize a sample as part of a population and understand that there are multiple samples inside a population.
Design Principle(s): Support sense-making (for representation); Maximize meta-awareness

12.4 Sampling the Population

10 minutes
This activity gives students the opportunity to practice the new vocabulary of population and sample by identifying the population from a set of questions and describing a possible sample that could be used to get some information to begin answering the question. Since these words have a very specific meaning in the context of statistics that is different from the colloquial use of the words, it is important for students to work with the vocabulary in specific situations to understand their meaning (MP6).

Addressing

- 7.SP.A.1

Instructional Routines

- Think Pair Share

Launch

Arrange students in groups of 2. Allow students 3 minutes of quiet work time followed by 3 minutes of partner discussion then a whole-class discussion.

While in partner discussion, suggest students compare their answers and discuss any advantages or disadvantages for the samples they proposed.

> ### Student Task Statement
>
> For each question, identify the **population** and a possible **sample**.
>
> 1. What is the mean number of pages for novels that were on the best seller list in the 1990s?
>
> 2. What fraction of new cars sold between August 2010 and October 2016 were built in the United States?
>
> 3. What is the median income for teachers in North America?
>
> 4. What is the average lifespan of Tasmanian devils?

Student Response

Answers for the sample vary. Sample responses:

1. Population: All New York Times Best Seller books from the 1990s. Sample: 20 of the books on the list that I could find in our library.

2. Population: All cars sold between August 2010 and October 2016. Sample: 100 cars sold during those dates.

3. Population: All teachers in North America. Sample: The teachers at our school.

4. Population: All Tasmanian devils that ever lived. Sample: The Tasmanian devils kept at our local zoo.

> ### Are You Ready for More?
>
> Political parties often use samples to poll people about important issues. One common method is to call people and ask their opinions. In most places, though, they are not allowed to call cell phones. Explain how this restriction might lead to inaccurate samples of the population.

Student Response

Answers vary. Sample response: Some people, especially younger people, may only have cell phones, so they will not be included in the sample. This may lead to more information being

gathered from older people than younger people, and the information may not accurately represent everyone.

Activity Synthesis

The purpose of the discussion is to further solidify the meaning of the terms population and sample for students.

Consider asking these questions for discussion:

- "For each question, could there be another population than the one you gave?" (No. The population refers to *all* of the individuals that pertain to the question.)

- "For each question, could there be another sample than the one you gave?" (Yes. A sample refers to a few of the individuals from whom data will be collected and does not specify the number or how the individuals are selected.)

- "What are some of the advantages and disadvantages you determined for the samples you chose?" (Some are easy to work with, but might miss large sections of the population.)

- "What is a question you could ask for which the *population* would be all of the books in your house?" (For example, "What is the average number of pages in books in my house?")

- "What is a question you could ask for which the *sample* could be all of the books in your house?" (For example, "What is the average number of pages in all the books ever written?")

Explain that a well-phrased question should only have 1 population (a question that is not well-phrased should be reconsidered so that the purpose of the question is clear), but there are usually many ways to find samples within that population. In future lessons, we will explore some important aspects to consider while selecting a sample.

Lesson Synthesis

Consider asking these questions to reinforce the ideas from this lesson:

- "When the groups become too large, how can we obtain some data to begin answering a question about the group?"

- "What are some drawbacks of using samples instead of the entire population?" (The value for the measure of center will not be exact and some variability may be lost. Some groups may not have been included in the sample, so their input is lost.)

- "What are some reasons samples are necessary?" (More manageable, impossible to reach the entire population, speed, cost.)

- "Someone wants to know what breed of dog is most popular as a pet in the state. What is a sample that could be used?" (A few dog owners from each of the major cities in the state and a few dog owners from the rural areas.)

- "The principal of a school has access to the grades for students at the school. If we use these grades as a sample, what is a population that the data could be applied to?" (The entire school district, the state, the United States, or all students around the world.)

12.5 How Many Games?

Cool Down: 5 minutes
The cool-down checks whether students understand the meanings of the terms population and sample as well as their use in context. Additionally, students are asked for at least one reason why a sample might make sense to use rather than the entire population.

Addressing

- 7.SP.A.1

> **Student Task Statement**
>
> Lin wants to know how many games teenagers in the United States have on their phones.
>
> 1. What is the population for Lin's question?
>
> 2. Explain why collecting data for this population would be difficult.
>
> 3. Give an example of a sample Lin could use to help answer her question.

Student Response

1. All people who are 13 to 19 years old in the United States who have a phone.

2. There are too many people to collect data from everyone. It would take too much time, energy, and money to collect the data.

3. Answers vary. Sample response: Ask 20 teens at Lin's school how many games they have on their phones.

> # Student Lesson Summary
>
> A **population** is a set of people or things that we want to study. Here are some examples of populations:
>
> - All people in the world
>
> - All seventh graders at a school
>
> - All apples grown in the U.S.
>
> A **sample** is a subset of a population. Here are some examples of samples from the listed populations:
>
> - The leaders of each country
>
> - The seventh graders who are in band
>
> - The apples in the school cafeteria
>
> When we want to know more about a population but it is not feasible to collect data from everyone in the population, we often collect data from a sample. In the lessons that follow, we will learn more about how to pick a sample that can help answer questions about the entire population.

Lesson 12 Practice Problems

Problem 1

> ### Statement
>
> Suppose you are interested in learning about how much time seventh grade students at your school spend outdoors on a typical school day.
>
> Select **all** the samples that are a part of the population you are interested in.
>
> A. The 20 students in a seventh grade math class.
>
> B. The first 20 students to arrive at school on a particular day.
>
> C. The seventh grade students participating in a science fair put on by the four middle schools in a school district.
>
> D. The 10 seventh graders on the school soccer team.
>
> E. The students on the school debate team.

Solution

["A", "D"]

Problem 2

> ### Statement
>
> For each sample given, list two possible populations they could belong to.
>
> a. Sample: The prices for apples at two stores near your house.
>
> b. Sample: The days of the week the students in your math class ordered food during the past week.
>
> c. Sample: The daily high temperatures for the capital cities of all 50 U.S. states over the past year.

Solution

Answers vary. Sample responses:

a. Population 1: Prices for apples at all stores in our state. Population 2: Prices for all fruit at these two stores.

b. Population 1: The days of the week the students in your math class ordered food all year. Population 2: The days of the week everyone in our city ordered food during the past week.

c. Population 1: The daily high temperatures for the world over the past year. Population 2: The daily high temperatures for the capital cities of all 50 U.S. states over the past 10 years.

Problem 3

Statement

If 6 coins are flipped, find the probability that there is at least 1 heads.

Solution

$\frac{63}{64}$, since there are 64 outcomes in the sample space ($2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 64$) and there is only 1 way to not get any heads (TTTTTT), so there are 63 ways that at least 1 heads shows up.

(From Unit 8, Lesson 9.)

Problem 4

Statement

A school's art club holds a bake sale on Fridays to raise money for art supplies. Here are the number of cookies they sold each week in the fall and in the spring:

fall	20	26	25	24	29	20	19	19	24	24
spring	19	27	29	21	25	22	26	21	25	25

a. Find the mean number of cookies sold in the fall and in the spring.

b. The MAD for the fall data is 2.8 cookies. The MAD for the spring data is 2.6 cookies. Express the difference in means as a multiple of the larger MAD.

c. Based on this data, do you think that sales were generally higher in the spring than in the fall?

Solution

a. 23 in the fall, 24 in the spring

b. 0.36 MADs since $1 \div 2.8 \approx 0.36$

c. The mean of sales is higher for the spring, but the difference in means is not very big considering the variability in the data.

(From Unit 8, Lesson 11.)

Problem 5

Statement

A school is selling candles for a fundraiser. They keep 40% of the total sales as their commission, and they pay the rest to the candle company.

price of candle	number of candles sold
small candle: $11	68
medium candle: $18	45
large candle: $25	21

How much money must the school pay to the candle company?

Solution

$1,249.80, because the school sold $2,083 worth of candles $(2{,}083 = 68 \cdot 11 + 45 \cdot 18 + 21 \cdot 25)$ and 60% is paid to the company $(0.6) \cdot 2{,}083 = 1{,}249.80$.

(From Unit 4, Lesson 11.)