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 auestion. 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. A **population** includes all the individuals related to a question, and a sample is a part of the population. Students consider how information from students in their class might apply to larger populations, then identify possible samples from populations given a statistical question. They must also make sense of a problem in which the population is too large to collect data from and are led to solve the problem using samples.

Student Learning Goal

Let's compare larger groups.

Access for Students with Diverse Abilities

• Representation (Activity 2)

Access for Multilingual Learners

• MLR2: Collect and Display (Activity 2)

Instructional Routines

• MLR2: Collect and Display

Required Preparation

Lesson:

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 or whatever they prefer to be called.) Do the same for the last names of students in the class.

Lesson Timeline



Warm-up



Activity 1



Activity 2



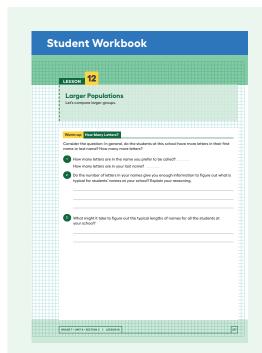
Activity 3



Lesson Synthesis



Cool-down



Warm-up

How Many Letters?



Activity Narrative

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 how information about an individual could be a start to estimating information about a population, but that more information should probably be collected.

Launch

Give students 2 minutes of quiet work time, and follow with a wholeclass 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. How many letters are in the name you prefer to be called? How many letters are in your last name?

Sample response: preferred name: 5; last name: 8

2. Do the number of letters in your names give you enough information to figure out what is typical for students' names at your school? Explain your reasoning.

Sample response: Not really. It gives a place to start for a single possibility, but many people have longer or shorter names, and I'm not sure if the lengths of my names are typical.

3. What might it take to figure out the typical lengths of names for all the students at your school?

Get a list of everyone in the school, count how many letters there are in each person's preferred and last names, and calculate the mean of each data set. 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.

Activity Synthesis

The purpose of the discussion is to get an idea of how samples might give some information about population data.

Invite students to share whether they think the length of their names gives enough information to understand the length of names at the school. If it does not come up, note that it may not be enough to know typical values very well, but it is a start.

Ask whether knowing the lengths of all students' names except one at the school might be enough to get an idea of what is typical. Students should begin to understand that having a group somewhere between all of the students and a single student might be enough to get an idea of typical name lengths.

Ask students whether it would be just as difficult to find typical name lengths for people all over the world. If it does not come up, point out that the larger the group a question applies to, the more difficult it gets to collect the data and compute the measures of center and variability.

Activity 1

John Jacobjingleheimerschmidt

5 min

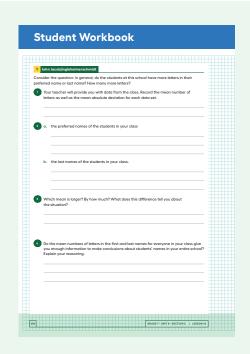
Activity Narrative

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. This activity gives students the first chance to experience why sampling might be needed.

Launch

Display the mean and MAD for the number of letters in each student's preferred name and last name that were calculated earlier.

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



Student Task Statement

Consider the question: In general, do the students at this school have more letters in their preferred name or last name? How many more letters?

- **1.** 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.
 - Answers vary.
- 2. a. the preferred names of the students in your class
 - Sample response: mean: 6.2 letters; MAD: 2.1 letters
 - b. the last names of the students in your class.
 - Sample response: mean: 7.3 letters; MAD: 2.8 letter
- **3.** Which mean is larger? By how much? What does this difference tell you about the situation?

The mean number of letters in our last names was larger by I.I 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: 7.3-6.2=I.I. Since the MAD measures the variability of the data set, this difference divided by the MAD gives a comparison: I.I ÷ $2.8\approx0.4$, 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.

4. 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.

There is enough information to make an estimate for the school, but I don't expect the mean to be exactly the same as for our class. There are still a lot of students that were not included to find this mean.

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:

- O "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.

Stop Signs



Activity Narrative

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. They are presented with a statistical question that does not have an obvious answer. Students then consider ways they might begin gathering data to answer the question, but realize that the data they could reasonably collect is not everyone addressed by the question. Following the activity, the discussion defines the term "population" as a set of people or things that is studied and the term "sample" as a part of the population from which data is actually gathered.

Launch 22

Arrange students in groups of 2.

Set up the context by asking students,

"Do people who ride to school see more stop signs than people who walk or bike?"

Then ask them to provide a possible explanation for their answer. For example, maybe if one rides to school in a car or bus or train, they are probably coming from farther away and might see more stop signs. Maybe they do not because they change their route to avoid stop signs.

Use Collect and Display to create a shared reference that captures students' developing mathematical language. Collect the language students use to describe the population and groups they might use to answer the questions. Display words and phrases such as "part," "whole," "all," or "everyone."

Instructional Routines

MLR2: Collect and Display

ilclass.com/r/10690754





Access for Multilingual Learners (Activity 2)

MLR2: Collect and Display.

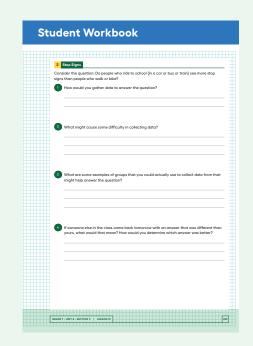
This activity uses the *Collect and Display* math language routine to advance conversing and reading as students clarify, build on, or make connections to mathematical language.

Access for Students with Diverse Abilities (Activity 2, Launch)

Representation: Internalize Comprehension.

Use multiple examples and non-examples to emphasize the definitions of population and sample.

Supports accessibility for: Conceptual Processing, Attention



Student Task Statement

Consider the question: Do people who ride to school (in a car or bus or train) see more stop signs than people who walk or bike?

- 1. How would you gather data to answer the question?

 Sample response: I would ask people who go to school how they get there and how many stop signs they see.
- 2. What might cause some difficulty in collecting data?
 - Sample response: It would be hard to ask all of the people in the world this question because they would all have to be located and have the means to answer it. It would cost a lot of money and time.
- **3.** What are some examples of groups that you could actually use to collect data from that might help answer the question?
 - Sample response: We could start with asking the students in our class, then if we wanted more data we could ask people in our neighborhoods or post an online poll.
- **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?
 - Sample response: 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 guestion and the data we have available.

Define population and sample. A **population** is a set of people or things that are studied. In this usage, it does not have to refer only to groups of people or animals. For example, when finding the average size of a grain of sand, the population would be all the grains of sand that exist. A **sample** is the part of the population from which data is actually collected.

Direct students' attention to the reference created using *Collect and Display*. Ask students to share whether the collected words describe a population, a sample, or both. Invite students to borrow language from the display as needed, and update the reference to include additional phrases as they respond.

Ask students,

"What is the population for the question about stop signs?"
everyone in the world who goes to school

Note that we would need data from everyone, including those who ride as well as those who walk or bike.

- "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
- © "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 the other way around.
- "What would you do with the data collected from everyone to answer the questions?"

I would find the mean and MAD of the data from the 2 sets and compare them like we did in previous lessons.

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 fish.
- Speed. Sometimes a rough estimate is all that is needed, and a sample of data is enough to estimate the population characteristic. Example: Find 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.

Activity 3

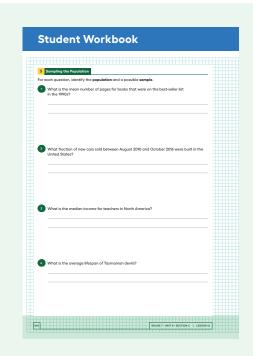
Sampling the Population

10 min

Activity Narrative

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 each of these words has a very specific meaning in the context of statistics that is different from the colloquial use of the word, it is important for students to work with the vocabulary word in specific situations to understand its meaning.

172





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 books that were on the best-seller list in the 1990s?

Sample response: Population: all best seller books from the 1990s; sample: 20 of the books on the list that I could find in our library

2. What fraction of new cars sold between August 2010 and October 2016 were built in the United States?

Sample response: Population: all cars sold between August 2010 and October 2016; sample: 100 cars sold during those dates

3. What is the median income for teachers in North America?

Sample response: Population: all teachers in North America; sample: the teachers at our school

4. What is the average lifespan of Tasmanian devils?

Sample response: 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.

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?"
 - "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?"

"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.

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?"
 - We can select a few individuals from the group to be part of a sample to collect data from.
- "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?"
 - Samples are more manageable, especially when it is impossible to reach the entire population, or there are concerns about speed and 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 sample could be made of a few dog owners from each of the major cities in the state and a few dog owners from the rural areas.

Responding To Student Thinking

More Chances

Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

"The principal of a school has access to the grades for students at the school. If the principal uses these grades as a sample, what is a population that the data could be applied to?"

It could be applied to any group that the school is a part of, for example, students in the district, state, or country. On the other hand, it might help to consider the kinds of students at the school and whether they are more alike or different than students in the larger groups. If the larger groups are different enough from this school's students, the information from the sample might not apply well.

Lesson Summary

A **population** is a set of people or things that is studied. For example, if we want to know how many apples grow on a typical orchard tree, the population might be all apple trees grown in orchards.

A **sample** is a part of a population. Using the apple tree example, a sample might be 100 randomly selected apple trees grown in orchards in the United States.

When we want to know more about a population but it is not easy to collect data from the entire population, we often collect data from a sample. It might cost too much time, money, or effort to count the number of apples on trees in all of the orchards around the world, but using a sample to count a few trees might give a good idea of what is typical for the entire population.

Cool-down

How Many Games?

5 min

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 it might make sense to use a sample instead of the entire population.

Student Task Statement

Lin wants to know how many games middle school students in the United States have on their phones.

- What is the population for Lin's question?
 all people who are in middle school in the United States who have a phone
- 2. Explain why collecting data for this population would be difficult.

 Sample response: There are too many people to collect data from everyone. It would take too much time, energy, and money to collect the data.
- **3.** Give an example of a sample Lin could use to help answer her question. Sample response: 20 teens at Lin's school.

Practice Problems

5 Problems

Problem 1

Suppose you are interested in learning about how much time seventh-grade students at your school spend outdoors on a typical school day.

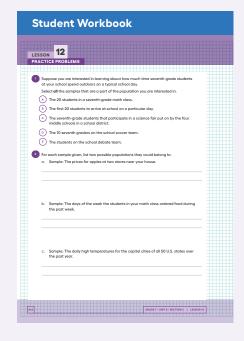
Select **all** the samples that are a part of the population you are interested in.

- **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 that participate 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.

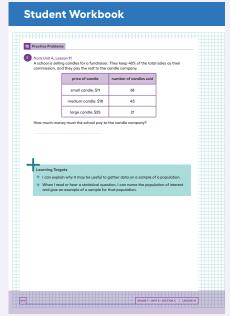
Problem 2

For each sample given, list two possible populations they could belong to.

- **a.** Sample: The prices for apples at two stores near your house.
 - Sample response: Population I: prices for apples at all stores in our state; population 2: prices for all fruit at these two stores
- **b.** Sample: The days of the week the students in your math class ordered food during the past week.
 - Sample response: Population I: 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.** Sample: The daily high temperatures for the capital cities of all 50 U.S. states over the past year.
 - Sample response: Population I: 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 IO years







Problem 3

from Unit 8, Lesson 9

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

63 64

There are 64 outcomes in the sample space $(2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 64)$, and there is only I way to not get any heads (TTTTT), so there are 63 ways that at least I flip lands on heads.

Problem 4

from Unit 8, Lesson 11

A school's art club holds an art sale on Fridays to raise money for art supplies. Here is the amount of money in dollars they made 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 amount of money made in the fall and in the spring.
 - 23 dollars in the fall, 24 dollars in the spring
- b. The MAD for the fall data is 2.8 dollars. The MAD for the spring data is 2.6 dollars. Express the difference in means as a multiple of the larger MAD.
 - 0.36 MADs, since $l \div 2.8 \approx 0.36$
- **c.** Based on this data, do you think that sales are generally higher in the spring than in the fall?

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

Problem 5

from Unit 4, Lesson 11

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?

\$1,249.80

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 $(0.6 \cdot 2,083 = 1,249.80)$.

LESSON 12 • PRACTICE PROBLEMS