What Makes a Good Sample?

Goals

- Calculate the mean or median of various samples, and compare them with the mean or median of the population.
- Comprehend that the term "representative" (in spoken and written language) refers to a sample with a distribution that closely resembles the population's shape, center, and spread.
- Given dot plots, determine whether a sample is representative of the population, and explain (orally and in writing) the reasoning.

Learning Targets

- I can determine whether a sample is representative of a population by considering the shape, center, and spread of each of them.
- I know that some samples may represent the population better than others.
- I remember that when a distribution is not symmetric, the median is a better estimate of a typical value than the mean.

Lesson Narrative

In this lesson, students examine multiple samples from the same population and learn what it means for a sample to be *representative* of the population. A sample is *representative* of a population if its distribution resembles the population's distribution in center, shape, and spread. Students look at the structure of dot plots, attending to center, shape, and spread, to help them compare the samples and the population. Although earlier activities pointed out the usefulness of using samples when working with large populations, the problems in this lesson use smaller populations so that students can compare each sample against the entire population.

An optional activity is included as extra practice for students to understand the relationship between a sample and population.

Student Learning Goal

Let's see what makes a good sample.

Lesson Timeline

| **5** min Warm-up | **15** min Activity 1 | **15** min Activity 2 | **10** min Activity 3 | **10** min Lesson Synthesis | **Assessment** **5** min Cool-down |

Instructional Routines

Math Talk

ilclass.com/r/10694967

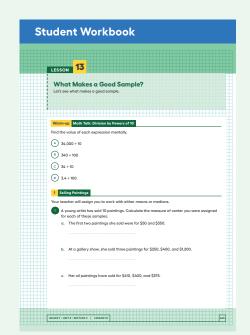Please log in to the site before using the QR code or URL.

Access for Students with Diverse Abilities (Warm-up, Student Task)

Action and Expression: Internalize Executive Functions.
To support working memory, provide students with sticky notes or mini whiteboards.

Supports accessibility for: Memory, Organization

Student Workbook

Warm-up

Math Talk: Division by Powers of 10

5 min

Activity Narrative

This *Math Talk* focuses on division by powers of 10. It encourages students to think about ways to divide and to rely on the patterns when dividing by a power of 10 to mentally solve problems. The strategies elicited here will be helpful later in the lesson when students find means for various samples.

To find different strategies, students need to look for and make use of structure.

Launch

Tell students to close their books or devices (or to keep them closed). Reveal one problem at a time. For each problem:

- Give students quiet think time, and ask them to give a signal when they have an answer and a strategy.
- Invite students to share their strategies, and record and display their responses for all to see.
- Use the questions in the *Activity Synthesis* to involve more students in the conversation before moving to the next problem.

Keep all previous problems and work displayed throughout the talk.

Student Task Statement

Find the value of each expression mentally.

A. $34{,}000 \div 10$

   3,400

   Sample reasoning: There are 100 tens in 1,000 and 34 hundreds in 34,000, so there are 34 hundred tens in 34,000.

B. $340 \div 100$

   3.40

   Sample reasoning: $100 \cdot 3.4 = 340$, so $340 \div 100 = 3.4$.

C. $34 \div 10$

   3.4

   Sample reasoning: There are 3 tens in 30 with 4 left over. Dividing this by 10 gives $\frac{4}{10}$.

D. $3.4 \div 100$

   0.034

   Sample reasoning: The decimal can be moved two places to the left.

Activity Synthesis

To involve more students in the conversation, consider asking:

💬 *"Who can restate _____'s reasoning in a different way?"*

"Did anyone use the same strategy but would explain it differently?"

"Did anyone solve the problem in a different way?"

"Does anyone want to add on to _____'s strategy?"

"Do you agree or disagree? Why?"

"What connections to previous problems do you see?"

Activity 1
Selling Paintings
15 min

Activity Narrative

In this activity, students begin to see numerical evidence that different samples can produce different results and thus different estimates for population characteristics. Students look at a small population and some different collections of samples from this population. Although the data for this population is small enough that it is not necessary to use a sample, it is helpful to get an idea of how data from a sample compares to the population data.

Launch 👥

Arrange students in groups of 2. In each group, one student should be assigned to work with mean as their measure of center, and the other should work with median as their measure of center.

Tell students that, often in this unit, the data sets are small enough that sampling is not necessary, but it will be easier to work with small data sets so that we may compare information from the sample to the same information from the population.

Student Task Statement

Your teacher will assign you to work with either means or medians.

1. A young artist has sold 10 paintings. Calculate the measure of center you were assigned for each of these samples:

   a. The first two paintings she sold were for $50 and $350.

   Mean: $200. Median: $200.

   b. At a gallery show, she sold three paintings for $250, $400, and $1,200.

   Mean: $616.67. Median: $400.

   c. Her oil paintings have sold for $410, $400, and $375.

   Mean: $395. Median: $400.

Access for Multilingual Learners (Warm-up, Synthesis)

MLR8: Discussion Supports.
Display sentence frames to support students when they explain their strategy. For example, "First, I _____ because …" or "I noticed _____, so I …" Some students may benefit from the opportunity to rehearse what they will say with a partner before they share with the whole class.
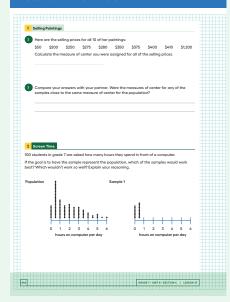
Advances: Speaking, Representing

Access for Students with Diverse Abilities (Activity 1, Student Task)

Action and Expression: Provide Access for Physical Action.
Provide access to tools and assistive technologies such as a graphing calculator or spreadsheet software to calculate means and medians.

Supports accessibility for: Visual-Spatial Processing, Conceptual Processing, Organization

Student Workbook

2. Here are the selling prices for all 10 of her paintings:

$50 $200 $250 $275 $280 $350 $375 $400 $410 $1,200

Calculate the measure of center you were assigned for all of the selling prices.

Mean: $379. Median: $315.

3. Compare your answers with your partner. Were the measures of center for any of the samples close to the same measure of center for the population?

Sample response: The mean oil paintings was close, but not exact. The other means were not very close. The sample medians were not very close for any of the samples.

Activity Synthesis

The purpose of this discussion is to show that different samples can result in different estimates for a population characteristic, and its purpose is to act as a reminder of reasons we might choose one measure of center over another.

Some questions for discussion:

💬 *"What is the population for this situation?"*

all of the paintings that were sold

💬 *"What are the samples used in the calculations?"*

the first two paintings that were sold, those sold at a gallery show, and the oil paintings

💬 *"Why did the different samples have different means?"*

because they used different paintings

💬 *"Why were the means for the first two paintings sold and those sold at the gallery show so far off from the mean of all the paintings?"*

because they contained the cheapest one and most expensive one, respectively, with only a few other numbers to balance it out

💬 *"Based on the numbers in the population, does it make more sense to use median or mean?"*

Median. The $1,200 painting is much greater than the rest of the values, so the measure of center is affected much more by the one painting when using mean.

Activity 2
Screen Time

15 min

Activity Narrative

In this activity, students begin to see that some samples represent the population better than others. Students compare the dot plot of a population of data with the dot plots of several samples and discuss some aspects that would make some samples better than others. In the discussion, the phrase "**representative** sample" is defined.

Launch 👥

Arrange students in groups of 2.

Display the population and sample dot plots from the activity. Ask students,

💬 *"What do you notice? What do you wonder?"*

Give students 1 minute of quiet think time, and then 1 minute to discuss the things they notice and wonder with their partner.

Invite students to share the things they noticed and wondered. Record and display their responses without editing or commentary for all to see. If possible, record the relevant reasoning on or near the dot plots.

Things students may notice:

- The samples each have 20 values.
- One of the samples looks more like the population than others.
- Some of the samples are more spread out than others.

Things students may wonder:

- Which sample might be the best to use to estimate things about the population?
- I wonder how these samples were collected.
- Is the mean for any of the samples close or the same as the population mean?

Use *Collect and Display* to create a shared reference that captures students' developing mathematical language. Collect the language students use to describe what makes some samples better than others. Display words and phrases such as "distribution," "shape," "center," or "similar."

Instructional Routines

Notice and Wonder

ilclass.com/r/10694948

Please log in to the site before using the QR code or URL.

Instructional Routines

MLR2: Collect and Display

ilclass.com/r/10690754

Please log in to the site before using the QR code or URL.
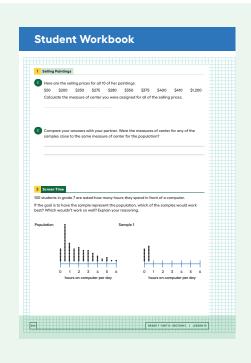

Access for Multilingual Learners (Activity 2)

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.
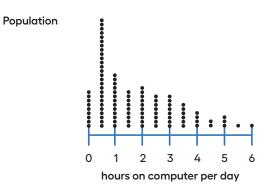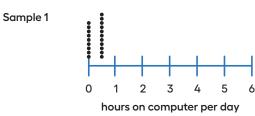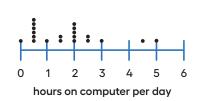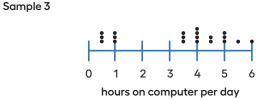
Student Task Statement

100 students in grade 7 are asked how many hours they spend in front of a computer.

If the goal is to have the sample represent the population, which of the samples would work best? Which wouldn't work so well? Explain your reasoning.

Population

hours on computer per day

Sample 1

hours on computer per day

Sample 2

hours on computer per day

Sample 3

hours on computer per day

Sample response: Sample 2 represents the population fairly well because it is spread out about the same amount, and the locations of the dots are more similar to the population than the other samples are. Sample 1 seems to be only the smallest 20 values in the population. Sample 3 does not represent the population very well because most of the dots are on the right side of the graph, while the population seems to have most of the dots on the left.

Are You Ready for More?

When doing a statistical study, it is important to keep the goal of the study in mind. Representative samples give us the best information about the distribution of the population as a whole, but sometimes a representative sample won't work for the goal of a study!

For example, suppose you want to study how discrimination affects people in your town. Surveying a representative sample of people in your town would give information about how the population generally feels, but might miss some smaller groups. Describe a way you might choose a sample of people to address this question.

Sample response: Get a list of the different groups in the town who might experience discrimination. Select multiple people from each group to be a part of the sample.

Activity Synthesis

Direct students' attention to the reference created using *Collect and Display.* Ask students to share how they determined which samples fit with the population better. Invite students to borrow language from the display as needed, and update the reference to include additional phrases as they respond.

Consider asking these discussion questions:

💬 *"What are some aspects that make for a good sample? a bad onead?"*

A sample is "good" if it has a similar distribution to the population data. A sample is "bad" if the data does not have a similar distribution to the population data. For example, Sample 3 is bad because it is not centered in the same place.

💬 *"If you were to find a measure of center to represent a typical value for the population, would you use mean or median?"*

I'd use the median because the data is not approximately symmetric.

💬 *"The population in this example has a mean of 1.76 hours and a median of 1.5 hours. Sample 1 has a mean of 0.33 hours and median of 0.5 hours. Sample 2 has a mean of 1.73 hours and a median of 1.75 hours. Sample 3 has a mean of 3.33 hours and a median of 4 hours. Based on this information, which seems to represent the population the best?"*

Sample 2

💬 *"Why might someone purposefully use Sample 1?"*

They may want to promote the idea that these kids do not sit in front of a computer very much every day and selectively choose the smallest numbers from the population.

💬 *"One of these samples was selected by randomly picking 20 students from the group. Which one do you think it was? Explain your reasoning."*

Sample 2. The other samples seem almost deliberately chosen to use mostly low or high values, so Sample 2 seems the most likely to be selected randomly.

Define **representative** sample. A representative sample is a sample that has a distribution that closely resembles the population distribution in center, shape, and spread.

Explain that a sample with the same mean as the population is not necessarily representative because it may miss other important aspects of the population.

- Example 1: If the population for a question is all of the humans in the world, and we use one person from each country as our sample, it may not actually be representative of the population. A larger country such as China is under-represented because there are actually many people living there, but only one is included in our sample. Similarly, a smaller country like Cuba might be over-represented because it has fewer people living there, but it is represented in the sample exactly the same way as all of the larger countries.

- Example 2: The average height of people in the world is approximately 66 inches. Two people may be found: one who is 87 inches (7 feet 3 inches) tall and one who is 45 inches (3 feet 9 inches) tall. Their mean height may be the same as the world's, but these two certainly do not represent the heights of most people.

Explain that a representative sample is the ideal type of sample we would like to collect, but if we do not know the data for the population, it will be hard to know if a sample we collect is representative or not. If we do know the population data, then a sample is probably unnecessary. There are ways to select samples that are more likely to produce representative samples that will be explored later.

Activity 3: Optional
Auditing Sales

10 min

Activity Narrative

This activity is additional practice for students to understand the relationship between a sample and population. It may take additional time, and so is included as an optional activity.

In this activity, students attempt to recreate the data from the population data using three given samples. It is important for students to recognize that this is difficult to do and that some samples are more representative than others. Without knowing the population data, though, it can be difficult to know which samples will be representative. Methods for selecting samples in an unbiased way are explored in future lessons.

Launch 👥

Keep students in groups of 2.

Remind students of the activity from a previous lesson where they selected papers (labeled A through O) from the bag and guessed the sample space. That was an example of trying to interpret information about the population given a sample of information.

Read the first sentence of the task statement:

💬 *"An online shopping company tracks how many items they sell in different categories during each month for a year."*

And then ask the students,

💬 *"How many dots should be represented in the population data for one year?"*
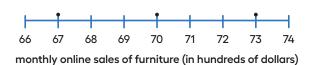
    12, one for each month of the year

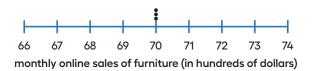Allow students 5 minutes of partner work time, and follow with a whole-class discussion.

Student Task Statement

An online shopping company tracks how many items they sell in different categories during each month for a year. Three different auditors each take samples from that data. Use the samples to draw dot plots of what the population data might look like for the furniture and electronics categories.
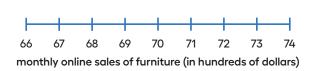
Auditor 1's sample

monthly online sales of furniture (in hundreds of dollars)

Auditor 2's sample

monthly online sales of furniture (in hundreds of dollars)

Auditor 3's sample

monthly online sales of furniture (in hundreds of dollars)

Population

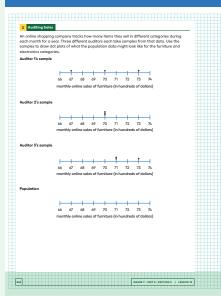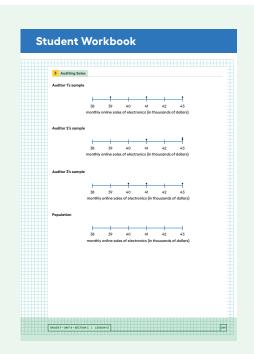monthly online sales of furniture (in hundreds of dollars)

Building on Student Thinking

Students may consider that each of the auditors' samples should be added together to create one larger sample rather than considering that the auditors may have chosen the same data point in their separate samples.

Therefore, each auditor having a data point at $41,000 may mean that there is only one data point there, and each auditor included it in the sample, or it may mean that there are actually three data points there, and each auditor included a different point from the population.
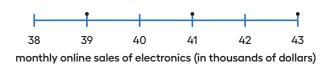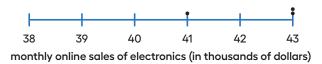
Student Workbook

Student Workbook

1. The dot plot for the population will have 12 dots, and each of the auditors' samples should be able to come from it, so there are some things the population must have: at least one dot at $6,700, at least 3 dots at $7,000, at least 2 dots at $7,100, and at least 1 dot at $7,300. The other 5 data points could be anywhere on the plot.
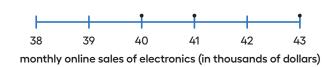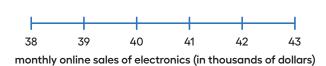
Auditor 1's sample

monthly online sales of electronics (in thousands of dollars)

Auditor 2's sample

monthly online sales of electronics (in thousands of dollars)

Auditor 3's sample

monthly online sales of electronics (in thousands of dollars)

Population

monthly online sales of electronics (in thousands of dollars)

2. The dot plot for the population will have 12 dots, and each of the auditors' samples should be able to come from it, so there are some things the population must have: at least one dot at $39,000, at least 1 dot at $40,000, at least 1 dot at $41,000, and at least 2 dots at $43,000. The remaining 7 dots may be placed anywhere.
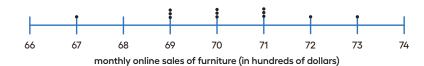
Activity Synthesis

The purpose of the discussion is for students to understand that getting an understanding of the population data from a sample can be very difficult, especially when it is not known whether samples are representative of the population or not.
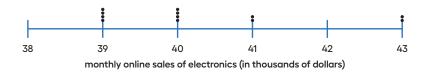
Display the population dot plots for all to see.

For furniture sales, the samples came from data represented in this dot plot.

monthly online sales of furniture (in hundreds of dollars)

For electronics sales, the samples came from data represented in this dot plot.

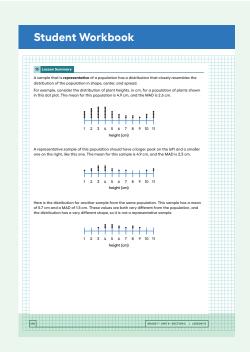monthly online sales of electronics (in thousands of dollars)

Ask students:

💬 *"How close was your estimate to the actual dot plot? Consider the shape, center, and spread of the data in your answer."*

"Are any samples better at mimicking the population than others?"

"What could the auditors have done to make their samples more representative of the population data without knowing what the population would be?"

They could have included more information in their samples. They should have also thought about how the samples were selected. For example, if the auditors only came on months when there were large sales happening, they may have missed important data.

Lesson Synthesis

Consider asking these discussion questions:

💬 *"What does it mean for a sample to be representative of the population?"*

The sample has a similar center, shape, and spread as the population data.

💬 *"Why might it be important to get a representative sample rather than a more convenient sample?"*

If we are going to answer questions about the entire population, it is useful if the sample looks similar to the population data. If not, we may miss some important information.
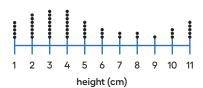
💬 *"Usually, a sample is used because we can't get data for the entire population. How do we know if the sample is representative of the population if we don't know the population?"*

It is OK for students to struggle with this answer at this point. In the next lesson, students explore ways to make their best attempt at getting a representative sample, so this question does not need to be answered right now.
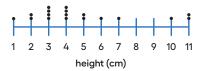
Lesson Summary

A sample that is **representative** of a population has a distribution that closely resembles the distribution of the population in shape, center, and spread.
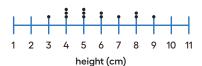
For example, consider the distribution of plant heights, in cm, for a population of plants shown in this dot plot. The mean for this population is 4.9 cm, and the MAD is 2.6 cm.

height (cm)

A representative sample of this population should have a larger peak on the left and a smaller one on the right, like this one. The mean for this sample is 4.9 cm, and the MAD is 2.3 cm.

height (cm)

Here is the distribution for another sample from the same population. This sample has a mean of 5.7 cm and a MAD of 1.5 cm. These values are both very different from the population, and the distribution has a very different shape, so it is not a representative sample.
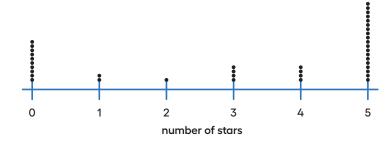

height (cm)

Cool-down

Reviews for School Lunches

5 min

Student Task Statement

Andre is designing a website that will display reviews of school lunches. Each item on the menu is rated from 0 to 5 stars. The main display can only show 6 reviews, so Andre needs to decide how to choose which reviews to show at the top.

This is a dot plot of all 40 reviews for the lasagna.

number of stars

This is a dot plot of the stars shown on the first page of results.

number of stars

1. If each rating also has a sentence or two explaining the rating, what are some good reasons to keep this sample displayed first? What are some good reasons to change the sample that is displayed first?

   Sample response: It might be good to keep it so that students can see the wide range of reviews possible for the lasagna. It might be good to change it because there are a lot more 0 and 5 star ratings than ones in the middle, so maybe there should be more of those ratings shown.

2. Is the sample **representative** of the population?

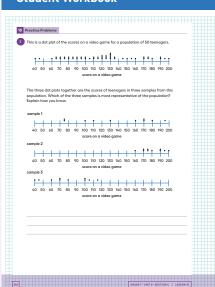   *It is not representative since the shape of the distributions are not similar.*

Practice Problems 5 Problems

Problem 1

Suppose 45% of all the students at Andre's school brought in a can of food to contribute to a canned food drive. Andre tries to pick a representative sample of 25 students from the school. He will calculate the percentage of the sampled students who brought in canned food.
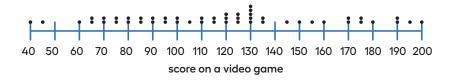
He expects the percentage for this sample will be 45%. Do you agree? Explain your reasoning.

No

Sample reasoning: The percentage of students cannot be exactly 45% because 45% of 25 is 11.25. The percentage in the sample may be close to 45%, but it cannot equal the population percentage for this sample. Even if it were possible to hit 45% exactly, it is likely for there to be some variation in samples.
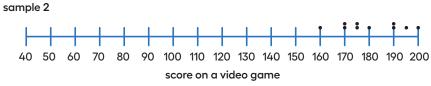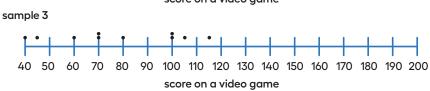
Problem 2

This is a dot plot of the scores on a video game for a population of 50 teenagers.

40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
score on a video game

The three dot plots together are the scores of teenagers in three samples from this population. Which of the three samples is most representative of the population? Explain how you know.
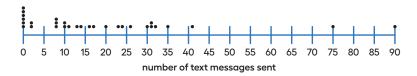
sample 1

40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
score on a video game

sample 2

40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
score on a video game

sample 3

40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
score on a video game

Sample reasoning: Sample 1. It is the only sample that has roughly the same center and spread of the population. Sample 2 has a very high center and very low spread. Sample 3 has a lower center and lower spread than the population.

Problem 3

This is a dot plot of the number of text messages sent one day for a sample of the students at a local high school. The sample consists of 30 students and is selected to be representative of the population.

number of text messages sent

a. What do the six values of 0 in the dot plot represent?

Six students in this sample didn't send any text messages.

b. Since this sample is representative of the population, describe what you think a dot plot for the entire population might look like.

Sample response: The population dot plot should have a lot of values at 0 representing students who didn't text that day. Most of the other values would be less than 50, but there would be a few dots representing students who send a lot more text messages than the typical student.

Problem 4

from Unit 8, Lesson 12

A doctor suspects you might have a certain strain of flu and wants to test your blood for the presence of markers for this strain of virus. Why would it be good for the doctor to take a sample of your blood rather than use the population?

Sample response: To use the population, the doctor would have to test all the blood in my entire body, and that is probably not possible while keeping me alive, so a smaller sample would be better.

Problem 5

from Unit 8, Lesson 8

How many different outcomes are in each sample space? Explain your reasoning. You do not need to write out the actual options, just provide the number and your reasoning.

a. A letter of the English alphabet is followed by a digit from 0 to 9.

260 outcomes

Sample reasoning: There are 26 letters and 10 digits, and $26 \cdot 10 = 260$.

b. A baseball team's cap is selected from 3 different colors, 2 different clasps, and 4 different locations for the team logo. A decision is also made to include or not to include reflective piping.

48 outcomes

Sample reasoning: There are 3 colors, 2 kinds of clasps, 4 positions for the team logo, and 2 piping options. $3 \cdot 2 \cdot 4 \cdot 2 = 48$

c. A locker combination like 7-23-11 uses three numbers, each from 1 to 40. Numbers can be used more than once, like 7-23-7.

64,000 outcomes

Sample reasoning: $40 \cdot 40 \cdot 40 = 64,000$