Lesson 13: 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, centre, 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, centre, 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 symmetrical, the median is a better estimate of a typical value than the mean.

Lesson Narrative

In this lesson, students examine multiple samples of the same population and learn what it means for a sample to be **representative** of the population. Students look at the structure of dot plots, attending to centre, shape, and spread, to help them compare the samples and the population. Although the previous lesson 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.

Building On

- Explain patterns in the number of zeros of the product when multiplying a number by powers of 10, and explain patterns in the placement of the decimal point when a decimal is multiplied or divided by a power of 10. Use whole-number exponents to denote powers of 10.

Addressing

- Use random sampling to draw inferences about a population.
- Understand that statistics can be used to gain information about a population by examining a sample of the population; generalisations 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.

- Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.

Instructional Routines

- Collect and Display
- Compare and Connect
- Discussion Supports
- Notice and Wonder
- Number Talk
- Think Pair Share

Required Materials

Four-function calculators

Student Learning Goals

Let's see what makes a good sample.

13.1 Number Talk: Division by Powers of 10

Warm Up: 5 minutes

The purpose of this number talk is to gather strategies and understandings students have for dividing by powers of 10. These understandings help students develop fluency and will be helpful later in this lesson when students will need to be able to find the mean for various samples.

While four problems are given, it may not be possible to share every strategy. Consider gathering only two or three different strategies per problem, saving most of the time for the final question.

Instructional Routines

- Discussion Supports
- Number Talk

Launch

Reveal one problem at a time. Give students 30 seconds of quiet think time for each problem and ask them to give a signal when they have an answer and a strategy. Keep all previous problems displayed throughout the talk. Follow with a whole-class discussion.

Representation: Internalise Comprehension. To support working memory, provide students with sticky notes or mini whiteboards.
Supports accessibility for: Memory; Organisation

Student Task Statement

Find the value of each quotient mentally.

$34\,000 \div 10$

$340 \div 100$

$34 \div 10$

$3.4 \div 100$

Student Response

- $3\,400$ Possible Strategies: Rewriting as a fraction and reducing.
- 3.40 Possible Strategies: Long division.
- 3.4 Possible Strategies: Regrouping.
- 0.034 Possible Strategies: Consider how many times you need to divide by 10.

Activity Synthesis

Ask students to share their strategies for each problem. Record and display their responses for all to see. To involve more students in the conversation, consider asking:

- "Who can restate ___'s reasoning in a different way?"
- "Did anyone have 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?"

Speaking: 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.
Design Principle(s): Optimise output (for explanation)

13.2 Selling Paintings

15 minutes

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.

Instructional Routines

- Discussion Supports

Launch

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

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.

Representation: Internalise Comprehension. Activate or supply background knowledge of calculating measures of centre. Allow students to use calculators to ensure inclusive participation in the activity.
Supports accessibility for: Memory; Conceptual processing Speaking: Discussion Supports. Use this routine to support whole-class discussion. For each response to the discussion questions, ask students to restate and/or revoice what they heard using precise mathematical language. Ask the original speaker whether their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped clarify the original statement. This will provide more students with an opportunity to produce language as they interpret the reasoning of others.
Design Principle(s): Support sense-making

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 centre you were assigned for each of these samples:
    a. The first two paintings she sold were for £50 and £350.
    b. At a gallery show, she sold three paintings for £250, £400, and £1 200.
    c. Her oil paintings have sold for £410, £400, and £375.
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 centre you were assigned for all of the selling prices.

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

Student Response

1.

   a. Mean: £200. Median: £200.

   b. Mean: £616.67. Median: £400.

   c. Mean: £395. Median: £400.

2. Mean: £379. Median: £315.

3. Answers vary. Sample response: The mean oil paintings were 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 as well as a reminder of reasons we might choose one measure of centre over another.

Some questions for discussion:

- "What is the population for this situation?" (All of the paintings sold.)
- "What are the samples used in the calculations?" (The first two paintings 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 since the £1 200 painting is much greater than the rest of the values, so the measure of centre is affected much more by the one painting when using mean.)

13.3 Sampling the Fish Market

15 minutes (there is a digital version of this activity)

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.

Instructional Routines

- Collect and Display
- Notice and Wonder
- Think Pair Share

Launch

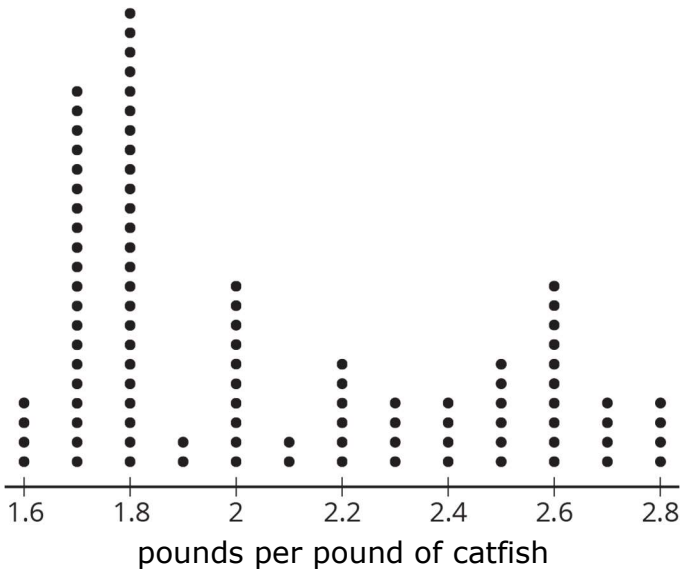
Arrange students in groups of 2. Allow students 3 minutes of quiet work time followed by a partner discussion and whole-class discussion.

Student Task Statement

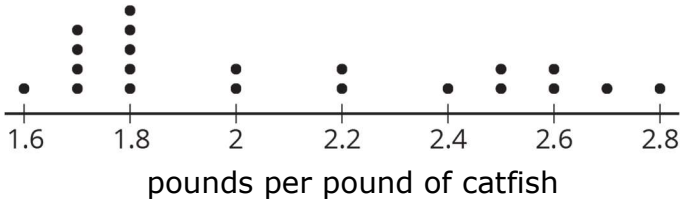
The price per pound of catfish at a fish market was recorded for 100 weeks.

1. Here are dot plots showing the population and three different samples from that population. What do you notice? What do you wonder?
2. 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.

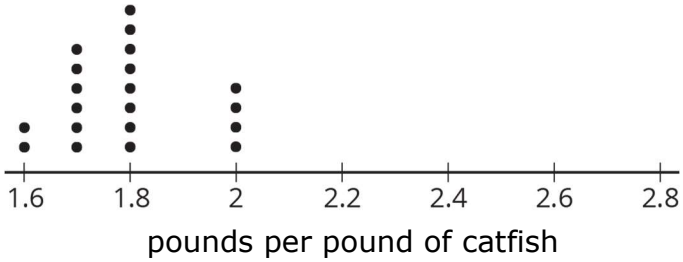
Population



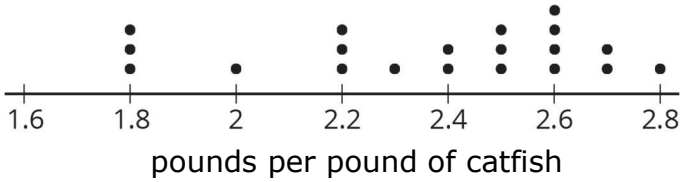
Sample 1



Sample 2



Sample 3



Student Response

1. Answers vary. Sample response: I notice that the samples each have 20 values. I notice that some of the samples look more like the population than the others. I wonder how they got these samples. I wonder if the mean for sample 1 is close to the mean for the population.

2. Answers vary. Sample response: Sample 1 represents the population fairly well since it is spread out about the same amount and has more dots in similar places as the population. Sample 3 does not represent the population very well, since 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.

Student Response

Answers vary. 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

Ask several groups to share things they noticed and wondered about the dot plots. Record responses for all to see. If possible, display the dot plots to refer to while students share.

Consider asking these discussion questions:

- "What are some aspects that make for a good sample? Bad?" (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 2 is bad because it is not centred in the same place.)
- "If you were to find a measure of centre to represent a typical value for the population, would you use mean or median?" (Median since the data is not approximately symmetrical.)
- "The population in this example has a mean of £2.06 and a median of £1.95. Sample 1 has a mean of £2.09 and median of £2. Sample 2 has a mean of £1.79 and a median of £1.80. Sample 3 has a mean of £2.36 and a median of £2.45. Based on this information, which seems to represent the population the best?" (Sample 1.)

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

Explain that a sample with the same mean as the population is not necessarily representative, since 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 you use one person from each country as your sample, it may not actually be representative of the population. Larger countries, such as China are under-represented since there are actually many Chinese people, but only 1 is included in our sample. Similarly, a smaller country like Cuba might be over-represented since it has fewer people living there, but is represented in the sample exactly the same as all of the other larger countries.
- Example 2: The average height of men in the world is approximately 70 inches. You might find two men, one who is 95 inches (7 feet 11 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 men.

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. In future lessons, we will explore methods of collecting samples that are more likely to produce representative samples (although they are still not guaranteed).

Representing, Speaking, Listening: Collect and Display. Create a table with column headings "good sample" and "bad sample". As students share aspects that make for a good or bad sample, write down the words and phrases students use in the appropriate column. Listen for and amplify words that compare features of the samples such as "similar or different centre," "shape," or "spread." Use the words and phrases that describe a good sample to define "representative sample." This will help students use and connect mathematical language that makes a sample representative of the population.
Design Principle(s): Support sense-making; Maximise meta-awareness

13.4 Auditing Sales

Optional: 10 minutes

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 recognise 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.

Instructional Routines

- Compare and Connect

Launch

Keep students in groups of 2.

Remind students of the activity from a previous lesson where students selected papers (labelled A through O) from the bag and guessed at 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 company tracks the number of pieces of furniture they sell 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 followed by a whole-class discussion.

Anticipated Misconceptions

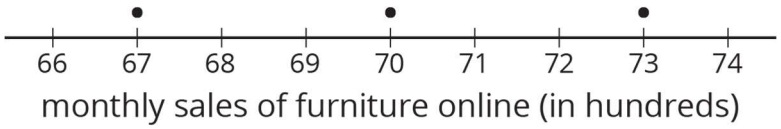
Students may consider that each of the auditors' samples should be added together to create one larger sample, rather than considering that the auditors may have chosen the same data point in their separate samples.

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.

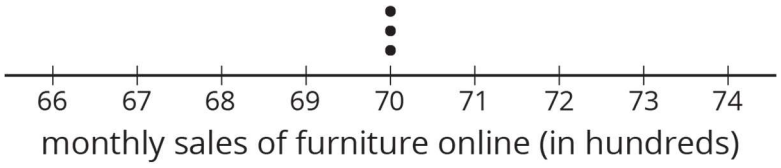
Student Task Statement

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.

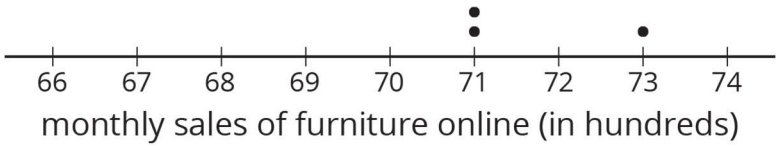
Auditor 1's sample



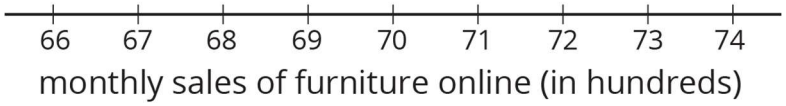
Auditor 2's sample



Auditor 3's sample



Population



Auditor 1's sample

Auditor 2's sample

Auditor 3's sample

Population

Student Response

1. Answers vary. The data 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.

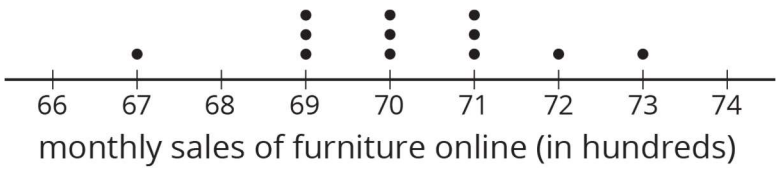
2. Answers vary. The data plot for the population will have 12 dots, and each of the auditors' samples should be able to come from it, so there are some things the population must have: at least one dot at 39 000, at least 1 dot at 40 000, at least 1 dot at 41 000, and at least 2 dots at 43 000. The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

Activity Synthesis

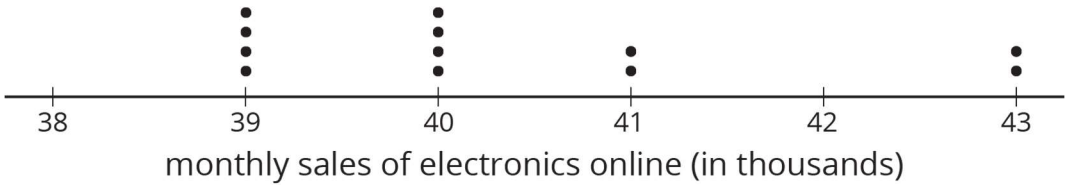
The purpose of the discussion is for students to understand that getting an understanding of the population data from a sample can be very difficult, especially when it is not known whether samples are representative of the population or not.

Display the population dot plots for all to see.

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



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



Ask students

- "How close was your estimate to the actual dot plot? Consider the shape, centre, and spread of the data in your answer."
- "Were 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 include more information in their samples. They should also think about how the samples were selected. For example, if the auditors only came on months when there were large sales happening, they may be missing important data.)

We will explore how to be careful about selecting appropriate samples in future lessons.

Representing, Conversing: Compare and Connect. Invite students to prepare a visual display that shows their dot plots for what the population data might look like for the furniture and electronics categories. Students should consider how to display their work so another student can interpret what is shown. Some students may wish to add notes or details to their drawings to help communicate their thinking. Invite students to investigate each other's work and to compare representations. Listen for and amplify the language students use to describe the data plot for the population, and how they explain why it is difficult to create an accurate population dot plot, given three small samples. This will foster students' meta-awareness and support constructive conversations as they relate sample and population data.
Design Principle(s): Optimise output (for comparison); Cultivate conversation

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 centre, 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, we'll explore ways to make our best attempt at getting a representative sample.)

13.5 Reviews for School Lunches

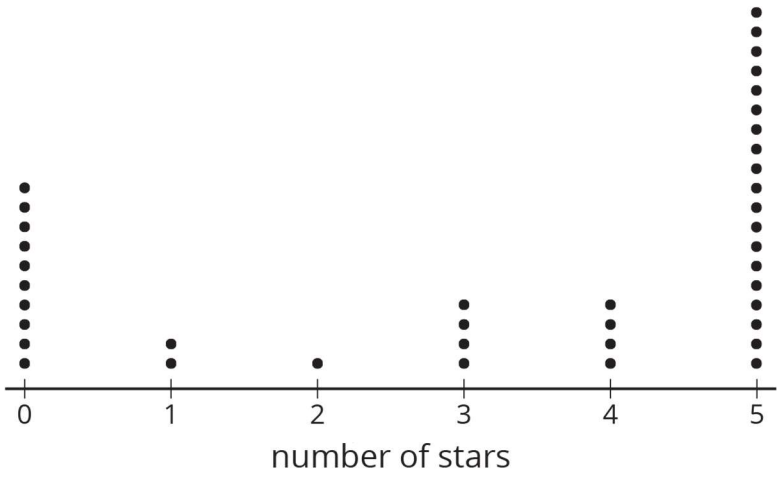
Cool Down: 5 minutes

In the cool-down, students need to understand what a representative sample means and why it might be useful to have.

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 lasagne.



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

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?

2. Is the sample representative of the population?

Student Response

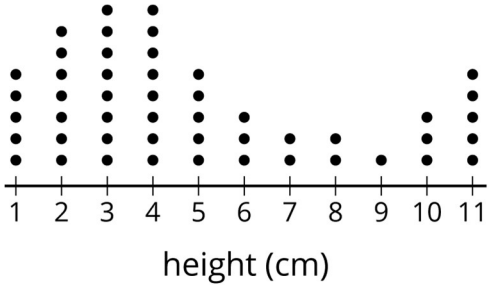
1. Answers vary. Sample response: It might be good to keep it so that students can see the wide range of reviews possible for the lasagne. It might be good to change it because there are a lot more 0 and 5 star ratings than ones in the middle, so maybe there should be more of those ratings shown.

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

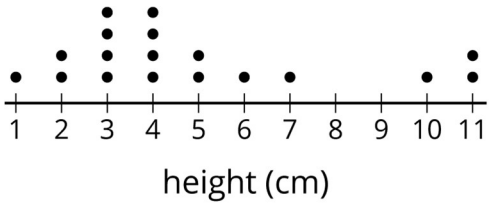
Student Lesson Summary

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

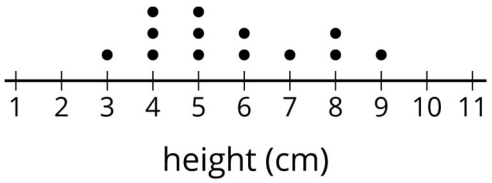
For example, consider the distribution of plant heights, in cm, for a population of plants shown in this dot plot. The mean for this population is 4.9 cm, and the range is 10 cm.



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 range is 10 cm.



Here is the distribution for another sample from the same population. This sample has a mean of 5.7 cm and a range of 6 cm. These are both very different from the population, and the distribution has a very different shape, so it is not a representative sample.



Glossary

- representative

Lesson 13 Practice Problems

Problem 1 Statement

Suppose 45% of all the students at Andre's school brought in a can of food to contribute to a canned food drive. Andre picks a representative sample of 25 students from the school and determines the sample's percentage.

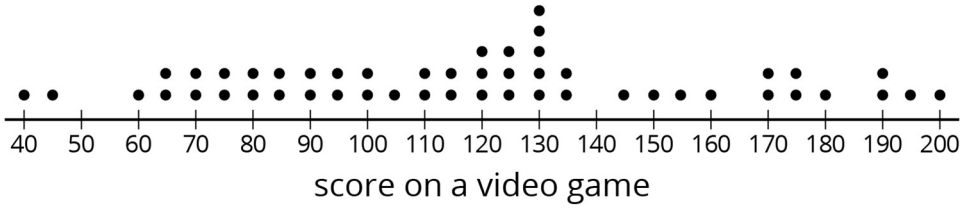
He expects the percentage for this sample will be 45%. Do you agree? Explain your reasoning.

Solution

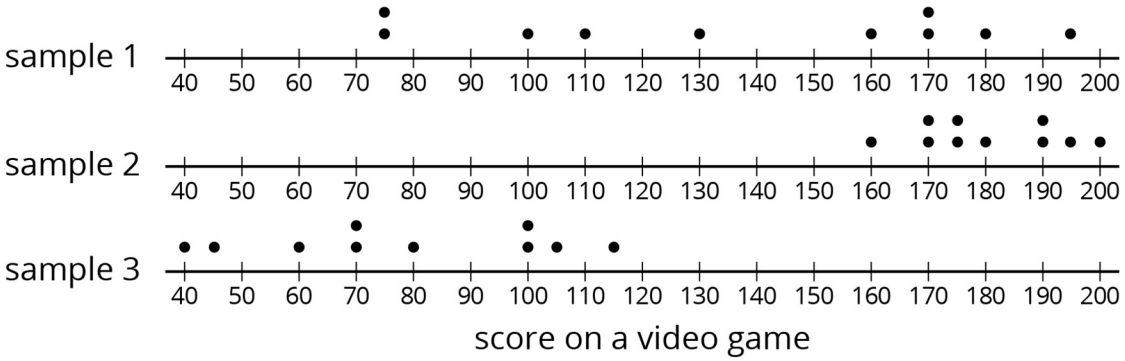
No, the percentage of students cannot be exactly 45%, since 45% of 25 is 11.25. The percentage in the sample is likely to be close to 45%, but it cannot equal the population percentage for this sample. Even if it were possible to hit 45% exactly, it is likely for there to be some variation in samples.

Problem 2 Statement

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



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.

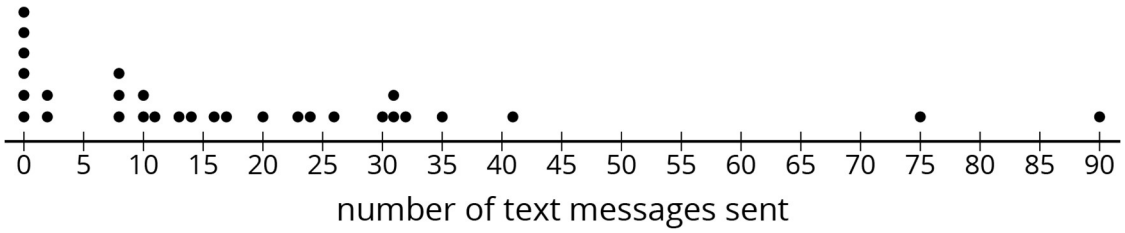


Solution

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

Problem 3 Statement

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



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

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

Solution

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

b. Answers vary. 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 Statement

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?

Solution

Answers vary. 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 Statement

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.

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

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.

Solution

a. 260 outcomes. There are 26 letters and 10 digits, and $26 \times 10 = 260$.

b. 48 outcomes. There are 3 bill colours, 2 kinds of clasps, 4 positions for the team logo, and 2 piping options. $3 \times 2 \times 4 \times 2 = 48$

c. 64 000 outcomes. $40 \times 40 \times 40 = 64\,000$