Lesson 17: More about Sampling Variability

Goals

- Compare and contrast (orally) a distribution of sample means and the distribution of the population.
- Generalize that an estimate for the center of a population distribution is more likely to be accurate when it is based on a larger random sample.
- Interpret (orally and in writing) a dot plot that displays the means of multiple samples from the same population.

Learning Targets

- I can use the means from many samples to judge how accurate an estimate for the population mean is.
- I know that as the sample size gets bigger, the sample mean is more likely to be close to the population mean.

Lesson Narrative

This lesson is optional. It goes beyond necessary grade-level standards to examine the accuracy of estimates for population characteristics based on many samples. The lesson builds a solid foundation for future grades to build upon, but may be shortened or skipped due to time constraints.

In this lesson, students continue to look at multiple samples from the same population. Examining the structure of dot plots composed of the means from several samples, they see that different samples from the same population can have different means, but that most of these means cluster around the mean of the population (MP7). They consider how far off their estimate might be if they didn't know the mean of the population but they did know the sample mean. Additionally, students see that larger samples usually produce sample means that are less variable from one another and can more accurately estimate population means.

Alignments

Building On

- 6.SP.A.3: Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Addressing

- 7.SP.A: Use random sampling to draw inferences about a population.
- 7.SP.A.2: 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.

Building Towards

- 7.SP.A.2: 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

- MLR2: Collect and Display
- MLR8: Discussion Supports
- Think Pair Share

Required Materials

Sticky notes

Required Preparation

For the "Reaction Population" activity, prepare a large number line for the class to use as a dot plot of their sample means from the warm-up of this lesson. Provide students with sticky notes to include as dots for this dot plot.

> ### Student Learning Goals
>
> Let’s compare samples from the same population.

17.1 Average Reactions

Warm Up: 10 minutes
Students calculate the mean of a sample collected in a previous lesson to compare with their partners. Students experience first-hand that different samples from the same population can produce different results. In the following activities of this lesson, students will use the data they have collected here to develop a deeper understanding of sampling variability.

Building On

- 6.SP.A.3

Building Towards

- 7.SP.A.2

Launch

Arrange students in groups of 2 so that different partners are used from the ones used in the Reaction Times activity in the previous lesson.

Remind students that the numbers came from a survey of all 120 seniors from a certain school. The numbers represent their reaction time in seconds from an activity in which they clicked a button as soon as they noticed that a square changed color. These 120 values are the population for this activity.

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

> **Student Task Statement**
>
> The other day, you worked with the reaction times of twelfth graders to see if they were fast enough to help out at the track meet. Look back at the sample you collected.
>
> 1. Calculate the mean reaction time for your sample.
> 2. Did you and your partner get the same sample mean? Explain why or why not.

Student Response

Answers vary. Sample responses:

1. 0.43 seconds
2. The two means were slightly different, but both close to 0.43 seconds. Since there were different samples, the means were slightly different.

Activity Synthesis

The purpose of the discussion is for students to think about how the data they collected relates to the population data.

Some questions for discussion:

- "Based on the information you currently know, estimate the mean of the population. Explain your reasoning." (Since both my partner and I got means close to 0.4, I think the population mean will be a little greater than 0.4.)
- "If each person selected 40 reaction times for their sample instead of 20, do you think this would provide a better estimate for the population mean?" (Since the sampling method is random, and thus fair, it should produce a better estimate.)

17.2 Reaction Population

Optional: 15 minutes

In the warm-up, students computed the mean of a sample. In this activity, a dot plot is created by the class that includes all of the calculated sample means. Students then compare that display to the data from the entire population to better understand the information that can be gained from a sample mean (MP7). In the discussion that follows the activity, students look at similar displays of sample means for samples of different sizes (MP8). Students should see that larger samples tend to more accurately estimate the population means than smaller samples.

Addressing

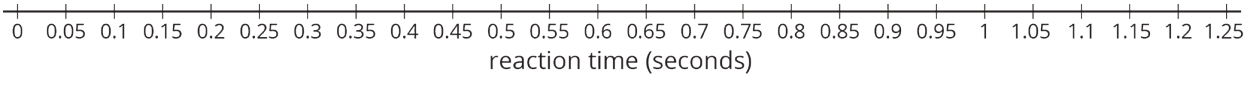
- 7.SP.A
- 7.SP.A.2

Instructional Routines

- MLR8: Discussion Supports

Launch

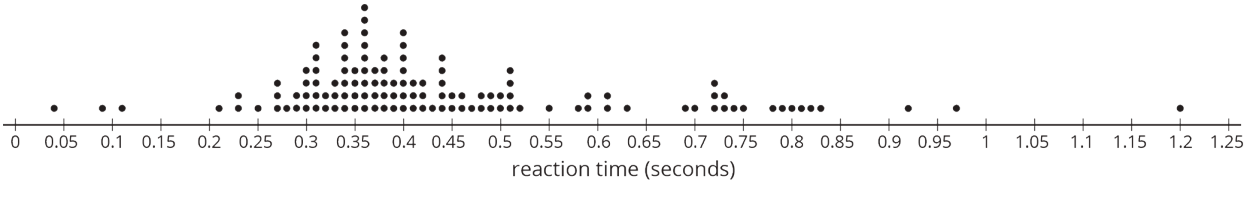
Display the basis for a dot plot of sample means for all to see.



Allow students 5 minutes to create the class dot plot and complete the first set of questions.

Ask students to include their sample means from the previous task to the display for all to see (since the samples were originated in pairs, there will be many repeats. It is ok to include the repeats or ensure only 1 of the original partners includes their point on the display). They may do this by adding a dot to the display on the board, placing a sticky note in the correct place, or passing a large sheet of paper around the class.

After students have had a chance to answer the first set of questions, pause the class to display the dot plot of the the reaction times for the entire population of 120 seniors for all to see.



Provide students 5 minutes to complete the problems. Follow with a whole-class discussion.

Anticipated Misconceptions

The center of the dot plot created by the class can be thought of as a mean of sample means. The phrase can be difficult for students to think through, so remind them what the dot they added to the dot plot represents and how they calculated that value. Consider displaying the class dot plot for the rest of the unit to help students remember this example to understand similar dot plots.

Student Task Statement

Your teacher will display a blank dot plot.

1. Plot your sample mean from the previous activity on your teacher's dot plot.
2. What do you notice about the distribution of the sample means from the class?
    a. Where is the center?
    b. Is there a lot of variability?
    c. Is it approximately symmetric?
3. The population mean is 0.442 seconds. How does this value compare to the sample means from the class?

    Pause here so your teacher can display a dot plot of the population of reaction times.

4. What do you notice about the distribution of the population?
    a. Where is the center?
    b. Is there a lot of variability?
    c. Is it approximately symmetric?
5. Compare the two displayed dot plots.
6. Based on the distribution of sample means from the class, do you think the mean of a random sample of 20 items is likely to be:
    a. within 0.01 seconds of the actual population mean?
    b. within 0.1 seconds of the actual population mean?

    Explain or show your reasoning.

Student Response

Answers vary. Sample responses:

1. No response required.
2. a. The center of the distribution is about 0.44 seconds.
    b. There is not a lot of variability. Most of the values are very close to 0.44 seconds.
    c. The distribution is approximately symmetric.
3. The center of the class distribution of sample means is about the same as the population mean.

4. a. The center of the population distribution is around 0.44 seconds

   b. There is relatively large variability.

   c. The distribution is approximately symmetric, but there are some times that are much longer than the bulk of the data.

5. The center of the two dot plots are in about the same place, but the dot plot of sample means shows much less variability.

6. a. Most of the dots in our sample mean plot are not within 0.01 seconds of the center, so I do not think the mean for a random sample of 20 would be that close.

   b. Most of the dots in our sample mean plot are within 0.1 seconds of the center, so I think a random sample mean would probably be within 0.1 seconds of the population mean.

Activity Synthesis

Display the dot plot of sample means for all to see throughout the unit. It may be helpful to refer to this display when viewing dot plots of sample means in future lessons.

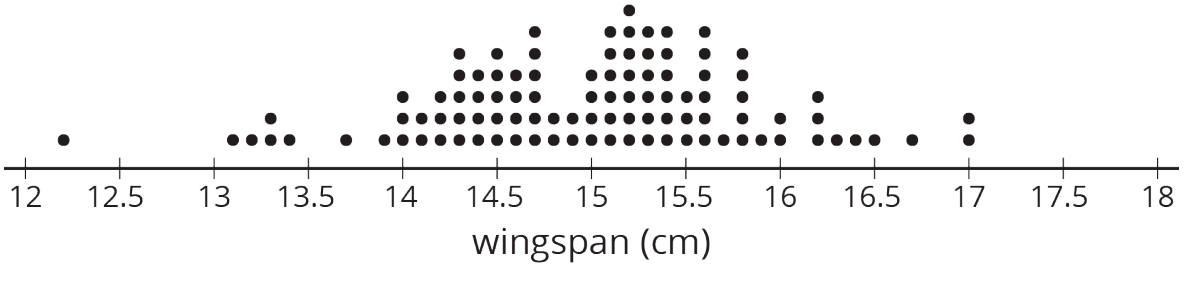
The purpose of the discussion is for students to understand that the centers of the two dot plots in the activity are close (if not the same), but the sample means have less variability and the shapes of the population distribution and the sample mean distribution are probably different. Additionally, students should use the dot plots shown here to see that larger samples have sample means that are still centered around the population mean, but have less variability than smaller samples. Therefore, larger samples should provide better estimates of the population mean than smaller samples do.

Items for further discussion:

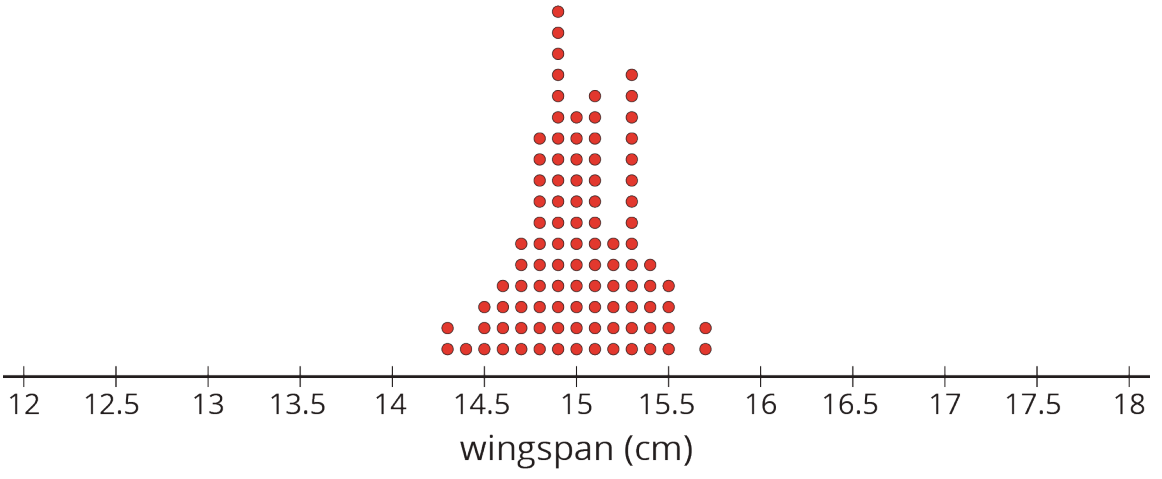
- "If the sample size were increased to 30 instead of 20, what do you think the distribution might look like?"

For a different set of data representing the wingspan in centimeters of a certain species of bird, compare the following dot plots.

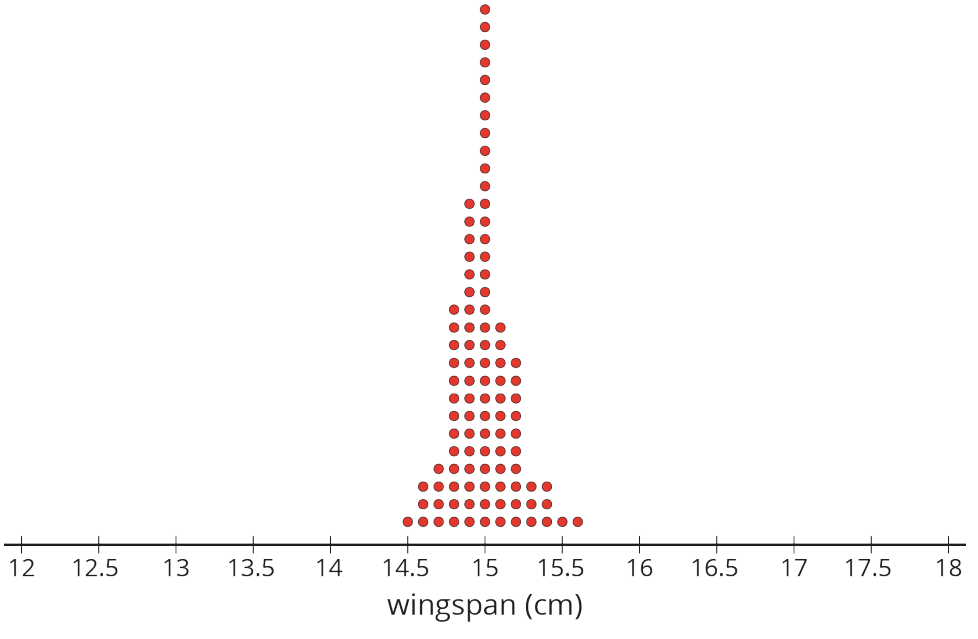
- Here is a dot plot of the population data.



- Here is a dot plot of the sample means for 100 different random samples each of size 10



- Here is a dot plot of the sample means for 100 different random samples each of size 30.



- "What do you notice about the distributions of the sample means as the sample size increases?" (The variability decreases, but the center stays in about the same place.)
- "Use the dot plots to explain why the sample mean from a random sample of size 30 would be a better estimate of the population mean than the sample mean from a random sample of size 10." (Since the variability is less, the sample mean is probably closer to the true mean with size 30 than with size 10.)

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Provide students with a graphic organizer to support their participation during the synthesis. Invite students to describe what to look for to determine variability and shape of the population distribution and to include examples for each.
Supports accessibility for: Conceptual processing; Organization

Access for English Language Learners

Conversing: MLR8 Discussion Supports. As students compare the wingspan dot plots, encourage students to use sentence frames such as: "When the sample size is 10, I noticed that the distribution . . .", "When the sample size is 30, I noticed that the distribution . . .", and "It's better to have a larger/smaller sample size to estimate the population mean because . . . ." The listener should press for more detail and explanation by asking for evidence from the dot plots. This will help students compare the dot plots and make conclusions about the mean of the population through communicating with their partner.
Design Principle(s): Support sense-making; Cultivate conversation

17.3 How Much Do You Trust the Answer?

Optional: 10 minutes
This activity is a follow-up to the context used in the activities Three Different TV Shows and Who's Watching What?, but also follows the ideas of sample means from this lesson.

In this activity, students continue to look at how variability in the sample means can be used to think about the accuracy of an estimate of a population mean. If the means from samples tend to be very spread out, then there is reason to believe that the mean from a single sample may not be especially close to the mean for the population. If the means from samples tend to be tightly grouped, then there is reason to believe that the mean from a single sample is a good estimate of the mean for the population (MP7).

Addressing

- 7.SP.A.2

Instructional Routines

- MLR2: Collect and Display
- Think Pair Share

Launch

Arrange students in groups of 2.

Tell students that, although the dot plots seem to have a similar shape, attention should be given to the scale of the horizontal axis.

Allow students 3–5 minutes of quiet work time followed by partner and whole-class discussions.

Access for Students with Disabilities

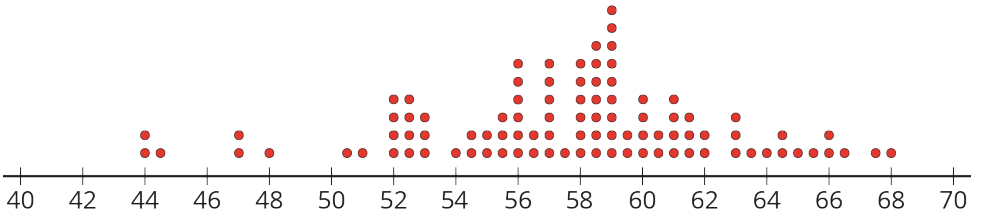
Action and Expression: Internalize Executive Functions. Chunk this task into more manageable parts. After students have solved the first 2-3 population mean estimate problems, check-in with either select groups of students or the whole class. Invite students to share the strategies they have used so far as well as any questions they have before continuing.
Supports accessibility for: Organization; Attention

Anticipated Misconceptions

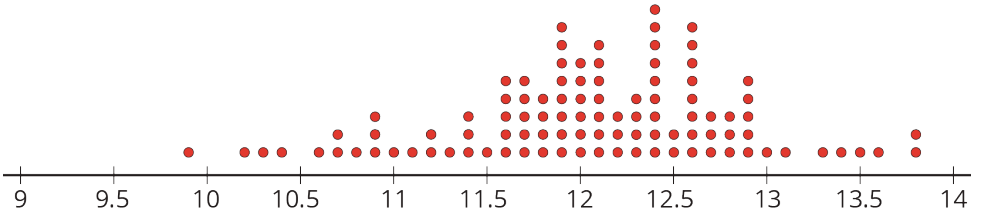
Students may get confused about what the dot plots represent. Refer to the class dot plot of sample means from the reaction time activity to help them understand that a single dot on the dot plot represents a mean from a single sample. Each dot plot shows means from several different samples.

Student Task Statement

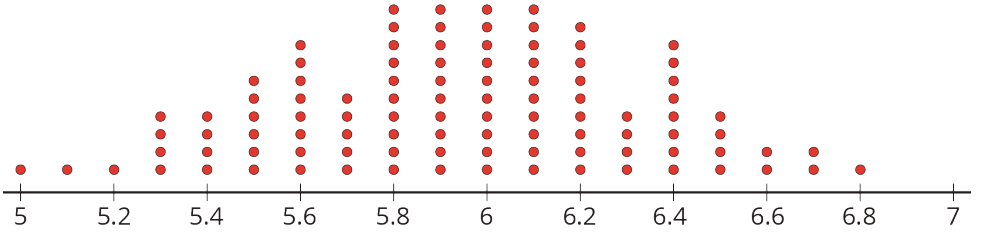
The other day you worked with 2 different samples of viewers from each of 3 different television shows. Each sample included 10 viewers. Here are the mean ages for 100 different samples of viewers from each show.



sample means for Trivia the Game Show



sample means for Science Experiments YOU Can Do



sample means for Learning to Read

1. For each show, use the dot plot to estimate the *population* mean.

   a. Trivia the Game Show

   b. Science Experiments YOU Can Do

   c. Learning to Read

2. For each show, are most of the sample means within 1 year of your estimated population mean?

3. Suppose you take a new random sample of 10 viewers for each of the 3 shows. Which show do you expect to have the new sample mean closest to the population mean? Explain or show your reasoning.

Student Response

1. "Trivia the Game Show": 58 years old. "Science Experiments YOU Can Do": 12 years old. "Learning to Read": 6 years old.

2. "Science Experiments YOU Can Do" and "Learning to Read" have most of the sample means within 1 year of the estimated population mean.

3. "Learning to Read" should produce the best estimates of population mean since it has the least variability in the sample means.

Are You Ready for More?

Market research shows that advertisements for retirement plans appeal to people between the ages of 40 and 55. Younger people are usually not interested and older people often already have a plan. Is it a good idea to advertise retirement plans during any of these three shows? Explain your reasoning.

Student Response

Answers vary. Sample response: It might be worth it to advertise during "Trivia the Game Show". Although the mean appears to be about 59, about half of the viewers are still in the desired age range of 40 to 55. It might depend on the total number of viewers to make the decision.

Activity Synthesis

The purpose of the discussion is to think about what the information in the given dot plots tells us about the accuracy of an estimate of a population mean based on the sample mean from a single sample.

Consider asking these questions for discussion:

- "In the first dot plot for 'Trivia the Game Show,' what does the dot at 68 represent?" (There is a sample of 10 viewers that has a mean age of 68 years.)
- "Based on the dot plot shown for 'Learning to Read,' do you think any 7 year olds were included in the data?" (Yes. Since one of the sample means is 6.8, there are probably a few 7 year olds and some 6 year olds in that sample.)
- "Why is this kind of dot plot useful?" (It shows how variable the means from samples can be, so it gives a good idea of how accurate a sample mean might be as an estimate of a population mean.)
- "Why might it be difficult to obtain a dot plot like this?" (Each dot represents a sample of 10. It might be hard to get 100 different samples of size 10.)

Access for English Language Learners

Representing, Speaking: MLR2 Collect and Display. Create a visual display with the heading "Dot Plots of Sample Means." As students respond to the discussion questions, write down the words and phrases students use on the visual display. Listen for and amplify words and phrases such as "each dot represents…," "variability of the sample means," and "population mean." As students review the language collected in the visual display, encourage students to clarify the meaning of a word or phrase. For example, a phrase such as "each dot represents a sample" can be clarified by restating it as "each dot represents a sample mean." This will help students make sense of dot plots of sample means as they draw conclusions about populations.
Design Principle(s): Support sense-making: Maximize meta-awareness

Lesson Synthesis

Consider asking these discussion questions to emphasize the main ideas of the lesson:

- "Do different samples always have the same measure of center?" (No, but they should be relatively close.)
- "If multiple researchers select their own random samples and compute the mean of their sample, how can this collection of means be used to estimate the mean for the population?" (They could create a dot plot with their sample means and find the center of that distribution to estimate the true population mean.)
- "Explain why larger samples are generally better than smaller ones." (Larger samples tend to produce more accurate estimates for the population mean since the sample means have less variability than sample means computed from smaller samples.)

17.4 How Much Mail?

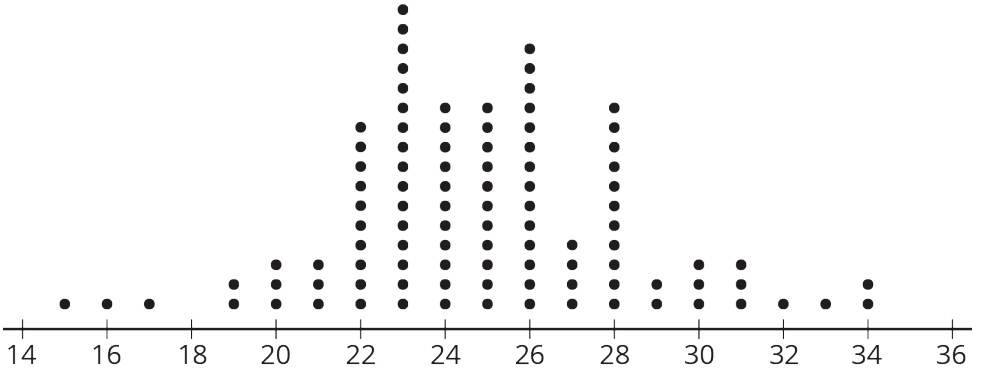
Cool Down: 5 minutes
This cool-down assesses whether students understand what information can be gained from looking at a dot plot of sample means as well as the idea that larger samples tend to product sample means that are less variable.

Addressing

- 7.SP.A
- 7.SP.A.2

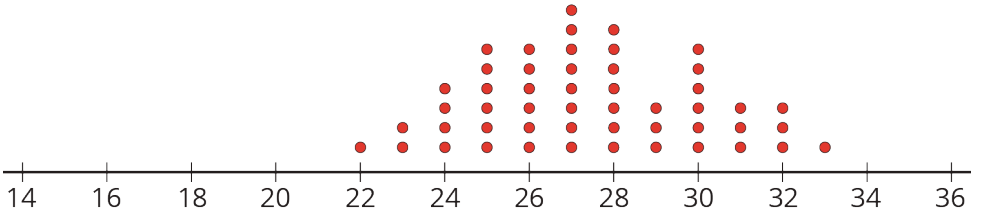
Student Task Statement

Jada collects data about the number of letters people get in the mail each week. The population distribution is shown in the dot plot.

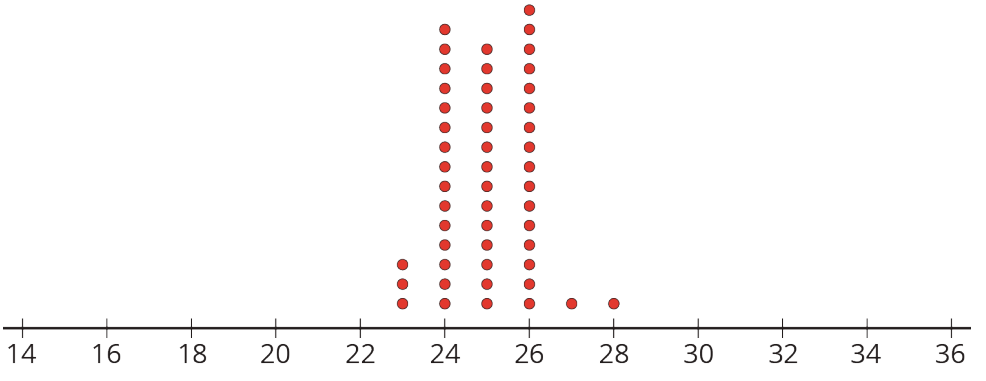


Which of the following dot plots are likely to represent the means from samples of size 10 from this population? Explain your reasoning.

Dot Plot 1



Dot Plot 2

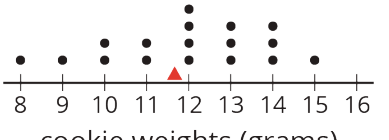


Student Response

Dot plot 2 since it has the same center, but less variability than the population data. Dot plot 1 also has less variability, but has a different center than the population data, so it is probably not generated by sample means from the original population.

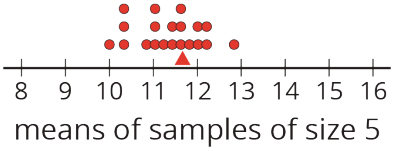
Student Lesson Summary

This dot plot shows the weights, in grams, of 18 cookies. The triangle indicates the mean weight, which is 11.6 grams.



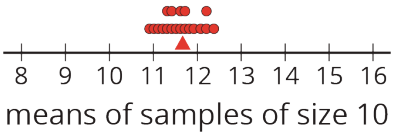
cookie weights (grams)

This dot plot shows the *means* of 20 samples of 5 cookies, selected at random. Again, the triangle shows the mean for the *population* of cookies. Notice that most of the sample means are fairly close to the mean of the entire population.



means of samples of size 5

This dot plot shows the means of 20 samples of 10 cookies, selected at random. Notice that the means for these samples are even closer to the mean for the entire population.



means of samples of size 10

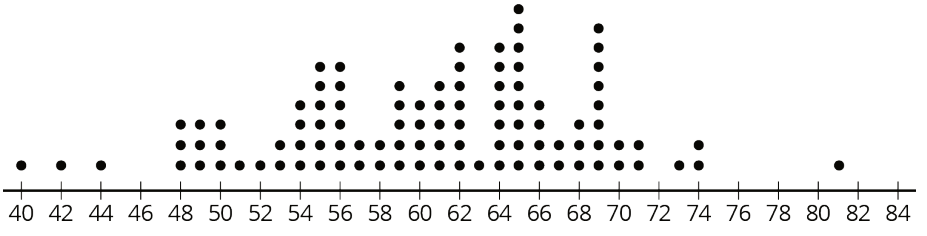
In general, as the sample size gets bigger, the mean of a sample is more likely to be closer to the mean of the population.

Lesson 17 Practice Problems

Problem 1

Statement

One thousand baseball fans were asked how far they would be willing to travel to watch a professional baseball game. From this population, 100 different samples of size 40 were selected. Here is a dot plot showing the mean of each sample.



Based on the distribution of sample means, what do you think is a reasonable estimate for the mean of the population?

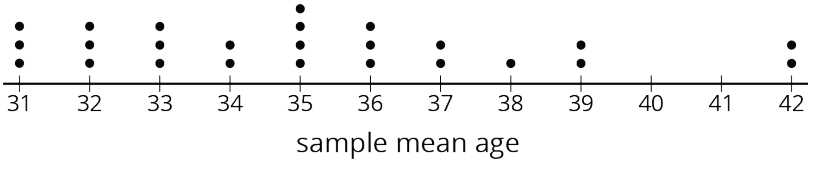
Solution

Reasonable answers are between 59 and 65.

Problem 2

Statement

Last night, everyone at the school music concert wrote their age on a slip of paper and placed it in a box. Today, each of the students in a math class selected a random sample of size 10 from the box of papers. Here is a dot plot showing their sample means, rounded to the nearest year.



a. Does the number of dots on the dot plot tell you how many people were at the concert or how many students are in the math class?

b. The mean age for the population was 35 years. If Elena picks a new sample of size 10 from this population, should she expect her sample mean to be within 1 year of the population mean? Explain your reasoning.

c. What could Elena do to select a random sample that is more likely to have a sample mean within 1 year of the population mean?

Solution

a. The math class

b. No. Only 9 of the 25 sample means in the dot plot are within 1 year of the population mean. While it could happen, it is more likely Elena's sample mean will be more than 1 year away from the population mean.

c. Select more than 10 papers for her sample

Problem 3

Statement

A random sample of people were asked which hand they prefer to write with. "l" means they prefer to use their left hand, and "r" means they prefer to use their right hand. Here are the results:

l r r r r r r r r

r l r r r r

Based on this sample, estimate the proportion of the population that prefers to write with their left hand.

Solution

$\frac{2}{15}$ (or about 13%)

(From Unit 8, Lesson 16.)

Problem 4

Statement

Andre would like to estimate the mean number of books the students at his school read over the summer break. He has a list of the names of all the students at the school, but he doesn't have time to ask every student how many books they read.

What should Andre do to estimate the mean number of books?

Solution

Answers vary. Sample response: Andre can estimate the mean by using the list he has to select a random sample of the students, then asking the students in the sample how many books they read over the summer. He could choose a sample size, but it would be better to take a larger sample if he can, because this probably gives him a more accurate estimate.

(From Unit 8, Lesson 15.)

Problem 5

Statement

A hockey team has a 75% chance of winning against the opposing team in each game of a playoff series. To win the series, the team must be the first to win 4 games.

a. Design a simulation for this event.

b. What counts as a successful outcome in your simulation?

c. Estimate the probability using your simulation.

Solution

a. Answers vary. Sample response: Make a spinner with 4 equal sections labeled 1, 2, 3, and 4. Spin the spinner, and record the outcomes as wins (if the spinner lands on 1, 2, or 3) or losses (when the spinner lands on 4) until one team wins the series. Repeat this process several times.

b. Answers vary. Sample response: Count the times that 1, 2, or 3 appear as winning a game and 4 as losing. A trial is a success if 1, 2, or 3 appear four times before 4 appears four times.

c. Answers vary. The actual probability is over 99%.

(From Unit 8, Lesson 10.)