Lesson 15

Estimating Population Measures of Center

Goals

- Calculate and interpret (orally and in writing) the mean absolute deviation of a sample.
- Generalize that an estimate for the center of a population distribution is more likely to be accurate when it is based on a random sample with less variability.
- Use the mean of a random sample to make inferences about the population, and explain (orally and in writing) the reasoning.

Learning Targets

- I can consider the variability of a sample to get an idea for how accurate my estimate is.
- I can estimate the mean or median of a population based on a sample of the population.

Access for Students with Diverse Abilities

- Action and Expression (Activity 1)
- Representation (Activity 3)

Access for Multilingual Learners

- MLR5: Co-Craft Questions (Activity 3)

Instructional Routines

- MLR5: Co-Craft Questions

Lesson Narrative

In this lesson students calculate measures of center and variation for samples from different populations and consider the meaning of these quantities in terms of the situation. Then students use these measures to estimate a center for the associated population. Students see that when there is less variability in the data from a sample, there is reason to believe that the measure of center from a sample is a better estimate for the measure of center from a population than when a sample has greater variability.

Student Learning Goal

Let's use samples to estimate measures of center for the population.

Lesson Timeline

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

Warm-up

Describing the Center

5 min

Activity Narrative

This *Warm-up* asks students to decide whether to use the mean or median based on the distribution of the data. As students compare groups in this section, the choice of measure of center will be important.

Launch

Arrange students in groups of 2.

Give students 1 minute quiet work time, followed by 2 minutes to discuss their work with a partner.

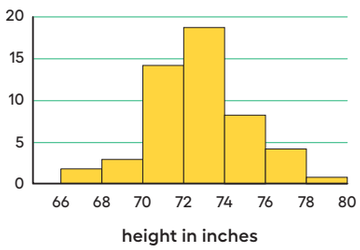
Then follow with a whole-class discussion.

Student Task Statement

Would you use the median or mean to describe the center of each data set? Explain your reasoning.

A

heights of 50 basketball players

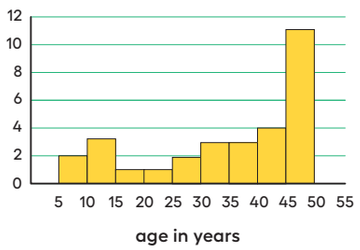


Mean. Sample reasoning:

The distribution is symmetric.

B

ages of 30 people at a family dinner party

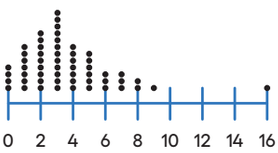


Median. Sample reasoning:

The distribution is not symmetric.

C

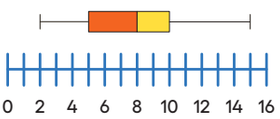
backpack weights of sixth-grade students



Median. Sample reasoning: The point at 16 would affect the mean much more than the median.

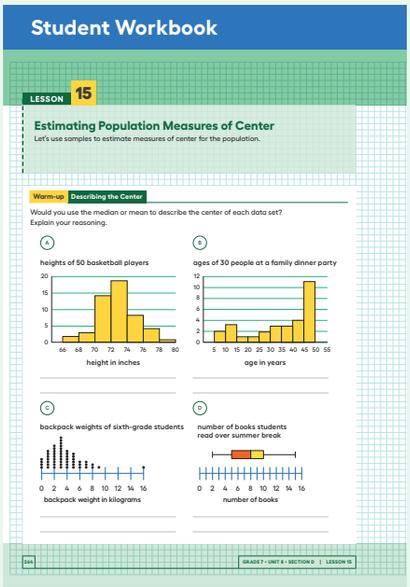
D

number of books students read over summer break



Sample reasoning: The mean can be used if it is known because the data set is approximately symmetric, but if what's given on the box plot is the only thing known, then the median should be used.

Student Workbook


LESSON 15

Estimating Population Measures of Center

Let's use samples to estimate measures of center for the population.

Warm-up Describing the Center

Would you use the median or mean to describe the center of each data set? Explain your reasoning.

A heights of 50 basketball players

B ages of 30 people at a family dinner party

C backpack weights of sixth-grade students

D number of books students read over summer break

Activity Synthesis

Select students to share their chosen measure of center and reasoning for their choice. Ask students what measures of variability should be used with each measure of center. Ask students,

- *"How does the symmetry of the distribution affect which measure of center we might choose to represent what is typical for a group?"*

  If the distribution is not symmetric, median is often a better choice for the measure of center because values that are very far from the center of the distribution do not affect the median as much as they do the mean. If the distribution is symmetric, the mean is better because it uses information from all of the data collected rather than ignoring some values to find the middle.

- *"What measures of variability should be used with each measure of center?"*

  Mean absolute deviation, or MAD, is used with mean and interquartile range, while IQR is used with median.

Activity 1

Three Different TV Shows

5 min

Activity Narrative

In this activity, students analyze data from samples of viewers for different TV shows. The data in this activity is used to begin the analysis as well as to get students thinking about the different shows the sample could represent. The purpose of the activity is to get students thinking about how measures of center from a sample might be used to make decisions about the population of a group.

Launch

Arrange students in groups of 3. Tell students that each person in the group should work on a different sample then share their results with their group.

Give students 1 minute quiet work time and then 1 minute to share their work with the group.

Followed with a whole-class discussion.

Student Task Statement

Here are the ages (in years) of a random sample of 10 viewers for 3 different television shows. The shows are titled, *Science Experiments YOU Can Do, Learning to Read, and Trivia the Game Show*.

Sample 1	6	6	5	4	8	5	7	8	6	6
Sample 2	15	14	12	13	12	10	12	11	10	8
Sample 3	43	60	50	36	58	50	73	59	69	51

1. Calculate the mean for one of the samples. Make sure each person in your group works with a different sample. Record the answers for all three samples.
   - Sample 1 mean: 6.1 years
   - Sample 2 mean: 11.7 years
   - Sample 3 mean: 54.9 years
2. Which show do you think each sample represents? Explain your reasoning.

   Sample responses:
   - Sample 1 is probably *Learning to Read* since it is the youngest group.
   - Sample 2 is probably *Science Experiments YOU Can Do* since that sounds like a show young people would like, but they should be older than 6 years old.
   - Sample 3 is probably *Trivia the Game Show* since that is the oldest group.

Activity Synthesis

Select students to share how they determined which shows matched with which data set. The purpose of the discussion is for students to notice that the shows are meant to appeal to different age groups.

Activity 2

Who's Watching What?

10 mins

Activity Narrative

This activity continues the work begun in the previous activity for this lesson. Students compute the means for sample ages to determine what shows might be associated with each sample. They also consider the variability to assess the accuracy of population estimates. A sample from a population with less variability should provide a more accurate estimate than a sample that came from a population with more spread in the data. In the discussion, students think about why a sample is used and why an estimate of the mean is helpful, even though it may miss some important aspects of the data. The discussion following the activity also asks students to think again about why different samples from the same population may produce different results.

Provide calculators to calculate statistics such as the mean and MAD.

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.

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

Student Workbook


1 Three Different TV Shows

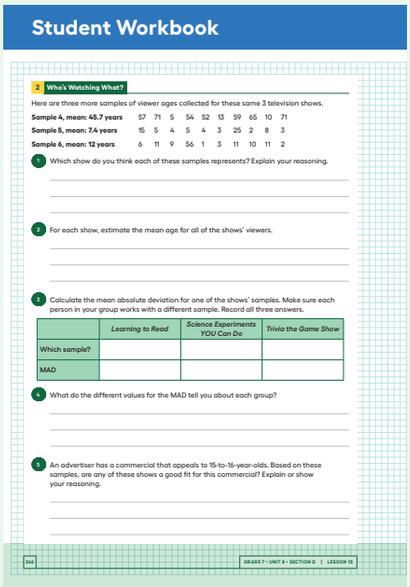
Here are the ages (in years) of a random sample of 10 viewers for 3 different television shows. The shows are titled, *Science Experiments YOU Can Do*, *Learning to Read*, and *Trivia the Game Show*.

Sample 1	6	6	5	4	8	5	7	8	6	6
Sample 2	15	14	12	13	12	10	12	11	10	8
Sample 3	43	60	50	36	58	50	73	59	69	51

1. Calculate the mean for one of the samples. Make sure each person in your group works with a different sample. Record the answers for all three samples.

2. Which show do you think each sample represents? Explain your reasoning.

Student Workbook



2 Who's Watching What?

Here are three more samples of viewer ages collected for these same 3 television shows.

Sample 4, mean: 45.7 years 57 71 5 54 52 13 59 65 10 71
Sample 5, mean: 7.4 years 15 5 4 5 4 3 25 2 8 3
Sample 6, mean: 12 years 6 11 9 56 1 3 11 10 11 2

1. Which show do you think each of these samples represents? Explain your reasoning.

2. For each show, estimate the mean age for all of the shows' viewers.

3. Calculate the mean absolute deviation for one of the shows' samples. Make sure each person in your group works with a different sample. Record all three answers.

	Learning to Read	*Science Experiments YOU Can Do*	*Trivia the Game Show*
Which sample?			
MAD			

4. What do the different values for the MAD tell you about each group?

5. An advertiser has a commercial that appeals to 15-to-16-year-olds. Based on these samples, are any of these shows a good fit for this commercial? Explain or show your reasoning.

Launch

Keep students in groups of 3.

Ask students if they have ever noticed that the advertisements that show up on different types of shows or websites are different depending on the main content. How might you expect the advertisements to be different for the three shows from a previous activity: *Science Experiments YOU Can Do, Learning to Read, and Trivia the Game Show?* If you had a business helping 15-to-16-year-olds learn how to drive, which show do you think would be best to advertise on?

The samples given in this activity are related to the shows mentioned in a previous activity.

After students have completed the first two problems, ask them to indicate which of the shows seem to go with each of the 6 samples (Samples 1 through 3 from the previous activity as well as 4 through 6 in this activity). Discuss any disagreements until the class can agree on which samples correspond to which shows. Tell half of the groups that they will use Samples 1 through 3 from a previous activity for the last three problems, and tell the other half of the groups that they will use Samples 4 through 6 for the last three problems.

Student Task Statement

Here are three more samples of viewer ages collected for these same 3 television shows.

Sample 4, mean: 45.7 years	57	71	5	54	52	13	59	65	10	71
Sample 5, mean: 7.4 years	15	5	4	5	4	3	25	2	8	3
Sample 6, mean: 12 years	6	11	9	56	1	3	11	10	11	2

1. Which show do you think each of these samples represents? Explain your reasoning.

- Sample 4 is *Trivia the Game Show.*

  Sample reasoning: The mean age is of an adult, and that makes sense for this show the most.

- Sample 5 is *Learning to Read.*

  Sample reasoning: Most of the people in this sample are young children who may be learning to read.

- Sample 6 is *Science Experiments YOU Can Do.*

  Sample reasoning: Most of the people in this sample are older children who might be more interested in science experiments than learning to read or a trivia game show.

2. For each show, estimate the mean age for all of the shows' viewers.

Mean age of Learning to Read: about 7 years.

Mean age of *Science Experiments YOU Can Do*: about 12.

Mean age of *Trivia the Game Show*: about 45.

3. Calculate the mean absolute deviation for one of the shows' samples. Make sure each person in your group works with a different sample. Record all three answers.

	Learning to Read	*Science Experiments YOU Can Do*	*Trivia the Game Show*
Which sample?	1; 5	2; 6	3; 4
MAD	0.94 years; 5.16 years	1.56 years; 8.8 years	8.9 years; 21.82 years

4. What do the different values for the MAD tell you about each group?

Sample response: The different values tell us how spread out the values are for each sample. For example, most of the ages for *Learning to Read* are close to the mean, but for *Trivia the Game Show*, the ages are spread out more.

5. An advertiser has a commercial that appeals to 15-to-16-year-olds. Based on these samples, are any of these shows a good fit for this commercial? Explain or show your reasoning.

Sample response: It might work best with *Science Experiments YOU Can Do*, but since the average age is 11.7 years old (or 12 years old) and the MAD is 1.56 years (or 8.8 years), there may or may not not be very many 15-to-16-year-olds watching this show.

Activity Synthesis

The purpose of the discussion is to understand why it might be helpful to estimate the mean of a population based on a sample.

Some questions for discussion:

- *"Why do you think a sample was used in this situation rather than data from the population?"*

  There are probably millions of people who watch these shows, and it would be difficult to collect data about their ages from all of them.

- *"How could we improve the estimate of the mean for the populations?"*

  If we include more viewers in the sample, then the estimate will probably improve with the additional information.

- *"If a sample has a large MAD, what does that imply about the population?"*

  It implies that the data in the population is very spread out.

- *"If a sample has a small MAD, what is the relationship between the data and the mean?"*

  Most of the data is close to the mean.

- *"Which estimate of the mean for the population do you expect to be more accurate: the mean from a sample with a large MAD or the mean from a sample with a small MAD? Explain your reasoning."*

  The mean from a sample with a small MAD is probably more accurate. If the sample does not have much variation, then we might expect the population to not have much variation either. That means the data might all be close to the mean.

- *"What do you notice about the mean and MAD for different samples related to the same show?"*

  The means are close, but the MADs can be different by a lot.

- *"Why are the answers for two samples different for the same show?"*

  Different people are included in the samples, so the numbers may change some, but if they are representative, they should be close. In these examples, even though the MADs may seem very different, the relative size compared to the other shows is similar.

Some students may wonder why they need to calculate the mean when it might be obvious how to match the titles by just looking at the data. This example included 10 ages in each sample so that the important information could be calculated quickly. In a more realistic scenario, the sample may include hundreds of ages. A computer could still calculate the mean quickly, but scanning through all of the data may not make the connection to the correct show as obvious.

- *"Notice that there is a 56-year-old in sample 6. What are some reasons you think they might be watching this show?"*

  Maybe a grandparent is watching with their grandchild. Maybe an older person is interested in how science is shown on TV.

- *"The questions ask you to consider means, but are there any data sets for which median might be a better measure of center? Explain your reasoning."*

  Yes, sample 6 has a wide range of data, ranging from a 1-year-old to a 56-year-old, but most of the data are around 10 years, so the median might be better to use for that sample.

- *"A lot of families might be watching Learning to Read with their children or older people may be using the show to learn English. How might this affect the mean? How could you recognize that there are two main age groups that watch this show?"*

  It would bring the mean up from what it would be if only kids were watching the show. The mean would not make the two age groups obvious, so looking at a dot plot or histogram might be more helpful with this group.

Activity 3
Movie Reviews

15 min

Activity Narrative

In this activity, students use data from a sample of movie reviews to estimate information about all the reviews for the movie. Based on the distribution of the data, students are asked to choose an appropriate measure of center and measure of variation, then apply their calculations to the entire population. Finally, students gauge their trust in the measure of center they have chosen based on the associate measure of variation.

Students must make use of the structure of the distribution to determine the best measure of center and variability to use.

Launch

Arrange students in groups of 3. Introduce the context of movie reviews. Use *Co-Craft Questions* to orient students to the context and elicit possible mathematical questions.

Display only the problem stem and related image, without revealing the questions.

Give students 1–2 minutes to write a list of mathematical questions that could be asked about the situation before comparing questions with a partner.

Invite several partners to share one question with the class and record responses. Ask the class to make comparisons among the shared questions and their own. Ask,

- *"What do these questions have in common? How are they different?"*
  Listen for and amplify language related to the learning goal, such as "measure of center," "mean," "median," and "measure of variability."

Reveal the questions about measure of center, and give students 1–2 minutes to compare it to their own question and those of their classmates.

Invite students to identify similarities and differences by asking:

- *"Which of your questions is most similar to or different from the ones provided? Why?"*

  "Is there a main mathematical concept that is present in both your questions and those provided? If so, describe it."

Instructional Routines

MLR5: Co-Craft Questions

ilclass.com/r/10695544

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

Access for Multilingual Learners (Activity 3)

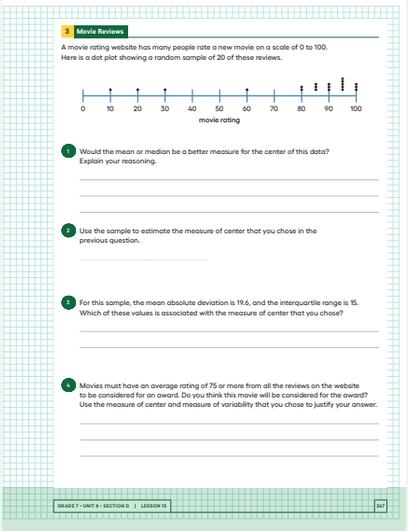
MLR5: Co-Craft Questions.
This activity uses the *Co-Craft Questions* math language routine to advance reading and writing as students make sense of a context and practice generating mathematical questions.

Access for Students with Diverse Abilities (Activity 3, Launch)

Representation: Develop Language and Symbols.
Activate or supply background knowledge. To help students recall the terms mean absolute deviation and interquartile range, ask students how to measure how spread out a distribution is and how symmetry can affect their choice.

Supports accessibility for: Memory, Language

Student Workbook


3 Movie Reviews

A movie rating website has many people rate a new movie on a scale of 0 to 100. Here is a dot plot showing a random sample of 20 of these reviews.

0 10 20 30 40 50 60 70 80 90 100
movie rating

1. Would the mean or median be a better measure for the center of this data? Explain your reasoning.

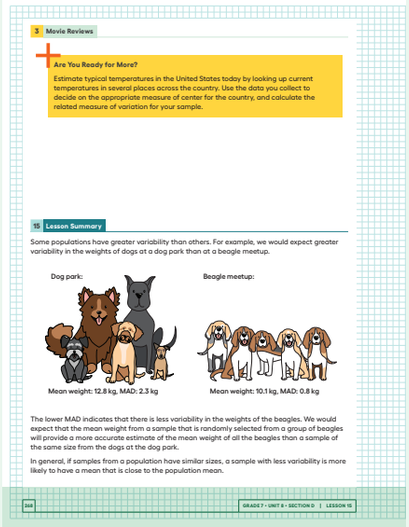
2. Use the sample to estimate the measure of center that you chose in the previous question.

3. For this sample, the mean absolute deviation is 19.6, and the interquartile range is 15. Which of these values is associated with the measure of center that you chose?

4. Movies must have an average rating of 75 or more from all the reviews on the website to be considered for an award. Do you think this movie will be considered for the award? Use the measure of center and measure of variability that you chose to justify your answer.

GRADE 7 • UNIT 8 • SECTION D | LESSON 15 267

Student Workbook


3 Movie Reviews

Are You Ready for More?

Estimate typical temperatures in the United States today by looking up current temperatures in several places across the country. Use the data you collect to decide on the appropriate measure of center for the country, and calculate the related measure of variation for your sample.

15 Lesson Summary

Some populations have greater variability than others. For example, we would expect greater variability in the weights of dogs at a dog park than at a beagle meetup.

Dog park:

Mean weight: 12.8 kg, MAD: 2.3 kg

Beagle meetup:

Mean weight: 10.1 kg, MAD: 0.8 kg

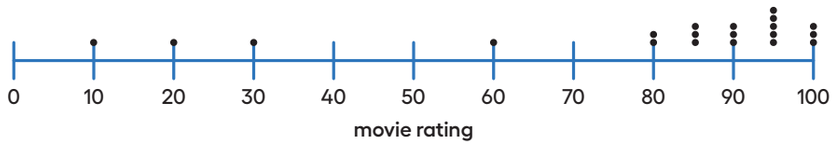
The lower MAD indicates that there is less variability in the weights of the beagles. We would expect that the mean weight from a sample that is randomly selected from a group of beagles will provide a more accurate estimate of the mean weight of all the beagles than a sample of the same size from the dogs at the dog park.

In general, if samples from a population have similar sizes, a sample with less variability is more likely to have a mean that is close to the population mean.

268 GRADE 7 • UNIT 8 • SECTION D | LESSON 15

Student Task Statement

A movie rating website has many people rate a new movie on a scale of 0 to 100. Here is a dot plot showing a random sample of 20 of these reviews.



1. Would the mean or median be a better measure for the center of this data? Explain your reasoning.

   Median

   Sample reasoning: Most of the data is above 80, but there are a few points that are much lower, so the data is not symmetric.

2. Use the sample to estimate the measure of center that you chose in the previous question.

   90 is the median.

3. For this sample, the mean absolute deviation is 19.6, and the interquartile range is 15. Which of these values is associated with the measure of center that you chose?

   The IQR of 15 is better to use with the median.

4. Movies must have an average rating of 75 or more from all the reviews on the website to be considered for an award. Do you think this movie will be considered for the award? Use the measure of center and measure of variability that you chose to justify your answer.

   Sample response: Yes. Because the first quartile for this data set is 80, at least 75% of the ratings in the sample are 80 or higher. Therefore, it is likely that the overall rating for the entire population is 75 or higher.

Are You Ready for More?

Estimate typical temperatures in the United States today by looking up current temperatures in several places across the country. Use the data you collect to decide on the appropriate measure of center for the country, and calculate the related measure of variation for your sample.

Answers vary.

Activity Synthesis

The purpose of the discussion is for students to review how to choose a measure of center and its associated measure of variation. Additionally, students use the measure of variation to help them think about how much to trust their population characteristic estimate.

Consider asking these discussion questions:

- *"Which measure of center did you choose and why?"*

  I chose the median since the distribution is not symmetric.

- *"Based on the situation, do you think other movie reviews would have non-symmetric distributions as well?"*

  Yes, usually a lot of people will agree whether a movie is good or bad, but a few people will have strong opinions on the other end of the scale.

- *"A random sample of 20 reviews for another movie also has a median of 90, but its IQR is 30. Do you think this movie is more or less likely to be considered for the award?"*

  I think it is less likely to be considered since there is more variability in the sample, so it is harder to estimate the median for all of the reviews.

- *"A random sample of 20 reviews for a third movie has a median of 50 and an IQR of 20. Is it possible this third movie will be considered for an award?"*

  It seems unlikely, but it is possible. The random sample may have randomly selected the 20 worst reviews, and all the other reviews gave it a 100 rating.

Lesson Synthesis

Consider asking these discussion questions to review the main ideas from the lesson:

- *"How do you determine which measure of center will best describe the data in a sample?"*

  If the distribution is symmetric, use the mean. If the distribution is skewed or has some data values far from the center of the distribution, then the median is often more representative of typical values.

- *"When you have the data from a sample, how can you estimate the value of a measure of center for the population?"*

  If the sample is random, calculate the appropriate measure of center for the sample and use that to estimate the same characteristic for the population.

- *"What does the variability of the sample tell you about your estimate for the measure of center of the population?"*

  The greater the variability, the less certain I am of the estimate. If the data is spread widely in the sample, it might be spread even more widely in the population, and this sample may not capture everything going on in the population, so the estimate may not be very accurate.

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.

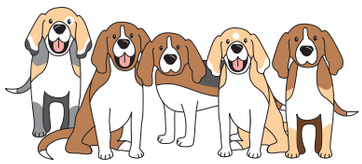
Lesson Summary

Some populations have greater variability than others. For example, we would expect greater variability in the weights of dogs at a dog park than at a beagle meetup.

Dog park:

Mean weight: 12.8 kg, MAD: 2.3 kg

Beagle meetup:



Mean weight: 10.1 kg, MAD: 0.8 kg

The lower MAD indicates that there is less variability in the weights of the beagles. We would expect that the mean weight from a sample that is randomly selected from a group of beagles will provide a more accurate estimate of the mean weight of all the beagles than a sample of the same size from the dogs at the dog park.

In general, if samples from a population have similar sizes, a sample with less variability is more likely to have a mean that is close to the population mean.

Cool-down

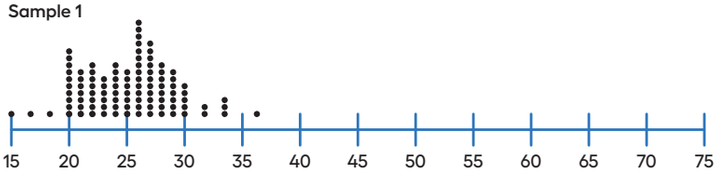
More Accurate Estimate

5 min

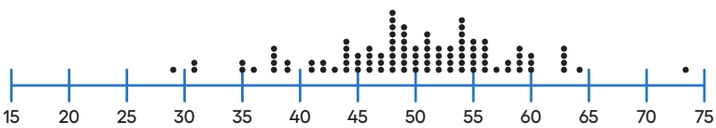
Student Task Statement

Here are dot plots that represent samples from two different populations.

Sample 1



Sample 1



1. Estimate the mean of each population using these samples.

   Correct responses should be close to 25 and 50 respectively.

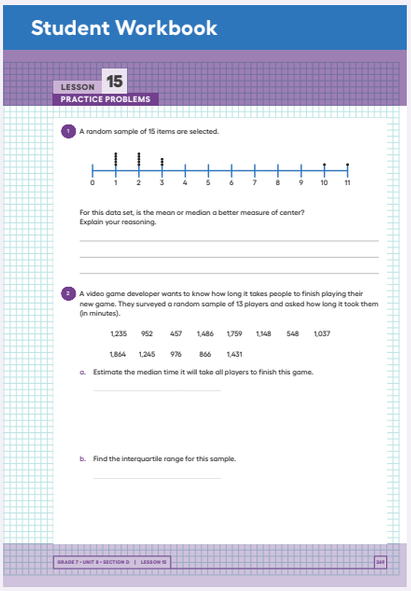
2. Based on the dot plots, which estimate is more likely to be accurate? Explain your reasoning.

   The estimate for Sample 1 is probably more accurate since there is much less variability in the data.

Lesson 15

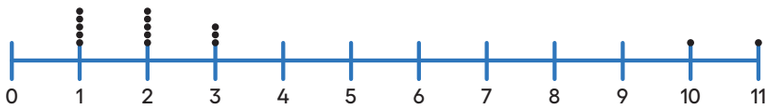
Practice Problems

4 Problems



Problem 1

A random sample of 15 items are selected.



For this data set, is the mean or median a better measure of center? Explain your reasoning.

Median

Sample reasoning: The data is not symmetric and has a couple of values far away from most of the other numbers.

Problem 2

A video game developer wants to know how long it takes people to finish playing their new game. They surveyed a random sample of 13 players and asked how long it took them (in minutes).

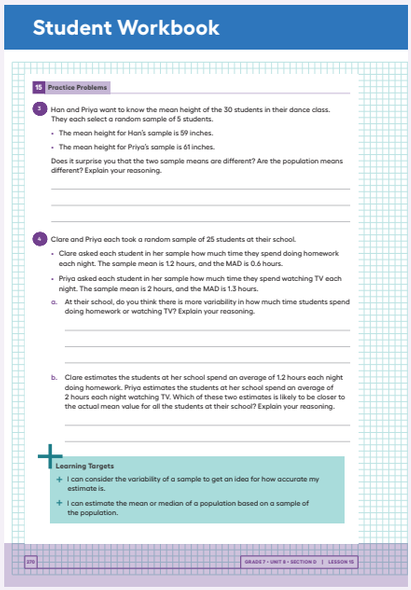
1,235	952	457	1,486	1,759	1,148	548	1,037
1,864	1,245	976	866	1,431			

a. Estimate the median time it will take all players to finish this game.

Median: 1,148 minutes

b. Find the interquartile range for this sample.

IQR: 549.5 minutes


Student Workbook

15 Practice Problems

3 Han and Priya want to know the mean height of the 30 students in their dance class. They each select a random sample of 5 students.

- The mean height for Han's sample is 59 inches.
- The mean height for Priya's sample is 61 inches.

Does it surprise you that the two sample means are different? Are the population means different? Explain your reasoning.

4 Clare and Priya each took a random sample of 25 students at their school.

- Clare asked each student in her sample how much time they spend doing homework each night. The sample mean is 1.2 hours, and the MAD is 0.6 hours.
- Priya asked each student in her sample how much time they spend watching TV each night. The sample mean is 2 hours, and the MAD is 1.3 hours.

a. At their school, do you think there is more variability in how much time students spend doing homework or watching TV? Explain your reasoning.

b. Clare estimates the students at her school spend an average of 1.2 hours each night doing homework. Priya estimates the students at her school spend an average of 2 hours each night watching TV. Which of these two estimates is likely to be closer to the actual mean value for all the students at their school? Explain your reasoning.

Learning Targets

+ I can consider the variability of a sample to get an idea for how accurate my estimate is.
+ I can estimate the mean or median of a population based on a sample of the population.

275 GRADE 7 • UNIT 8 • SECTION D | LESSON 15

Problem 3

Han and Priya want to know the mean height of the 30 students in their dance class. They each select a random sample of 5 students.

- The mean height for Han's sample is 59 inches.
- The mean height for Priya's sample is 61 inches.

Does it surprise you that the two sample means are different? Are the population means different? Explain your reasoning.

No

Sample reasoning: Even though they both selected a random sample, their samples probably included different people from the population, so the two sample means would not necessarily be the same, even though there is only one population mean.

Problem 4

Clare and Priya each took a random sample of 25 students at their school.

- Clare asked each student in her sample how much time they spend doing homework each night. The sample mean is 1.2 hours, and the MAD is 0.6 hours.
- Priya asked each student in her sample how much time they spend watching TV each night. The sample mean is 2 hours, and the MAD is 1.3 hours.

a. At their school, do you think there is more variability in how much time students spend doing homework or watching TV? Explain your reasoning.

There is more variability in the times spent watching TV.

Sample reasoning: The MAD is a measure of variability, and the MAD for time spent watching TV is greater than the MAD for the time spent doing homework.

b. Clare estimates the students at her school spend an average of 1.2 hours each night doing homework. Priya estimates the students at her school spend an average of 2 hours each night watching TV. Which of these two estimates is likely to be closer to the actual mean value for all the students at their school? Explain your reasoning.

Clare's estimate is more likely to be closer.

Sample reasoning: The times spent doing homework don't vary as much. It is harder to get an accurate estimate of the population mean when there is a lot of variability in the population values