Comparing Mean and Median

Goals

- Choose which measure of center to use to describe a given data set and justify (orally and in writing) the choice.
- Explain (orally) that the median is a better estimate of a typical value than the mean for distributions that are not symmetric or contain values far from the center.
- Generalize how the shape of the distribution affects the mean and median of a data set.

Learning Targets

- I can determine when the mean or the median is more appropriate to describe the center of data.
- I can explain how the distribution of data affects the mean and the median.

Lesson Narrative

In this lesson, students investigate whether the mean or the median is a more appropriate measure of the center of a distribution in a given context. They learn that when the distribution is symmetrical, the mean and median have similar values, so the mean should usually be used for its connection to fairness. When a distribution is not symmetrical, however, the mean is often greatly influenced by values that are far from the majority of the data points. In this case, the median may be a better choice.

At this point, students may not yet fully understand that the choice of measures of center is not entirely black and white, or that the choice should always be interpreted in the context of the problem and should hinge on what insights we seek or questions we would like to answer. This is acceptable at this stage. In upcoming lessons, they will have more opportunities to include these considerations into their decisions about measures of center.

Student Learning Goal

Let's compare the mean and median of data sets.

Lesson Timeline

5 min

Warm-up

206

10 min

Activity 1

20 min

Activity 2

10 min

Lesson Synthesis

Access for Students with Diverse Abilities

- Action and Expression (Activity 2)
- Engagement (Activity 1)

Access for Multilingual Learners

 MLR3: Critique, Correct, Clarify (Activity 2)

Instructional Routines

- Card Sort
- · MLR3: Critique, Correct, Clarify
- Take Turns

Required Materials

Materials to Gather

- · Four-function calculators: Activity 1
- Pre-printed cards, cut from copies of the blackline master: Activity 2

Materials To Copy

 Mean or Median Cards (1 copy for every 4 students): Activity 2

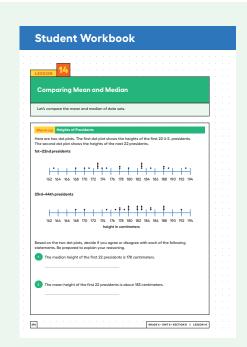
Required Preparation

Lesson:

For "The Tallest and Smallest in the World" activity, students will need the data on their heights (collected in the first lesson). Consider preparing a class dot plot that shows this data set to facilitate discussions.

5_{min}

Cool-down



Student Workbook Vivinion Melights of Presidents A typical height for a president in the second group is about 182 certimeters. U.S. presidents have become taller over time. The heights of the first 22 presidents are more alike than the heights of the second 22 presidents. The MAD of the second data set is greater than the MAD of the first set.

Warm-up

Heights of Presidents



Activity Narrative

In this *Warm-up*, students review ways to interpret and compare data shown on a dot plot. The discussion on each given statement gives the teacher an opportunity to hear how students reason about the median, mean, typical value, spread, balance point, and MAD. This discussion will be helpful in upcoming activities, as students compare median and mean values for different data sets.

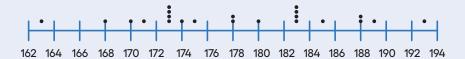
Launch

Give students 2 minutes of quiet work time. Follow with a whole-class discussion.

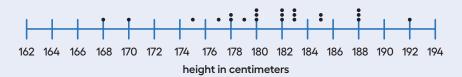
Student Task Statement

Here are two dot plots. The first dot plot shows the heights of the first 22 U.S. presidents. The second dot plot shows the heights of the next 22 presidents.

1st-22nd presidents



23rd-44th presidents



Based on the two dot plots, decide if you agree or disagree with each of the following statements. Be prepared to explain your reasoning.

1. The median height of the first 22 presidents is 178 centimeters.

Agree

The median is the average of 178 and 178 centimeters, which are the 11th and 12th data points.

2. The mean height of the first 22 presidents is about 183 centimeters.

Disagree

Though there are 4 presidents who are 183 centimeters tall, 183 is not the balance point of the data. There are many more presidents who are shorter than 183 centimeters than are taller than 183 centimeters.

3. A typical height for a president in the second group is about 182 centimeters.

Agree

The center of the data appears to be about 182 centimeters.

4. U.S. presidents have become taller over time.

Agree

The center of the data in the second group is higher than in the first group.

5. The heights of the first 22 presidents are more alike than the heights of the second 22 presidents.

Disagree

The spread of the data for the first 22 presidents is wider than that for the other group, so overall their heights are more different.

6. The MAD of the second data set is greater than the MAD of the first set.

Disagree

Compared to the first group, the data points for the second group are clustered closer together, so their average distance from the mean is likely smaller, not greater.

Activity Synthesis

For each statement, ask students to indicate if they agree or disagree. If all students agree or all students disagree, ask a couple of students to explain their reasoning. If the class is divided on a statement, ask students on both sides to share their reasoning until the class comes to an agreement. As students share, record and display their responses for all to see. If possible, record their reasoning on the dot plots to highlight important terms that students use.

To help facilitate the discussion, consider asking:

"Who can restate ______'s reasoning in a different way?"
"Did anyone have the same reasoning but would explain it differently?"
"Did anyone reason about the statement in a different way?"
"Does anyone want to add on to ______'s reasoning?"
"Do you agree or disagree? Why?"

Activity 1

The Tallest and the Smallest in the World

10 min

Activity Narrative

In this lesson, students begin to notice how the distribution of data affects the mean and median of a data set. Using their class height data, they examine how both measures of center are affected when a value that is far from the center is added to a data set. Students find that adding these unusually large or small values pulls the mean up or down while having little or no effect on the median. They begin to see that for data sets with some far-off values, the median might be a better choice for describing a typical value because the sizes of those extreme values (whether very large or very small) do not affect the median as much as they do the mean.

Provide technology such as a spreadsheet or calculator that can quickly find the mean and median. The focus of the activity should be on interpreting why the measures differ with outliers rather than on calculation.

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

Engagement: Provide Access by Recruiting Interest.

Invite students to share the approximate height of the tallest person they know.

Supports accessibility for: Conceptual Processing, Memory

Building on Student Thinking

When calculating the mean after a new person joins the class, some students might enter individual heights into a calculator once again, rather than using the sum from their original mean calculation to save time (for example, by adding 251 to the sum of heights of students and then dividing by a class size that includes one additional student). Urge them to think about how they might use the previous calculation to make the process more efficient.

Student Workbook



Launch

Consider preparing a large-scale dot plot that shows student height data and displaying it for all to see.

Arrange students in groups of 3–4. Provide each student with the data on students' heights (collected in the first lesson of the unit) and access to calculators or other technology that can quickly calculate the mean and median of a data set. Give groups 5–7 minutes to complete the questions, and then move on to the whole-class discussion.

Student Task Statement

Your teacher will provide some height data for your class. Use the data to complete the questions.

1. Find the mean height of your class in centimeters.

Sample response: 160 cm

2. Find the median height in centimeters. Show your reasoning.

Sample response: 156 cm

- **3.** Suppose that the world's tallest adult, who is 251 centimeters tall, joined your class.
 - a. Find the new mean.

Sample response: 164 cm

b. Find the new median.

Sample response: 156 cm

c. Which measure of center—the mean or the median—changes more when this new person joins the class? Explain why the value of one measure changes more than the other.

Sample response: The mean changes more because the tall person's height is counted equally with all the other students. For the median, the middle value in the list of heights is still the same or very close.

Activity Synthesis

Use the whole-class discussion to further explore how unusually high or low values affect the mean and the median. Invite several students to share the new mean and median, and to explain how these measures would change if the world's shortest person joined the class. (If a class dot plot that shows student heights is made, add the point that represents the smallest person in the world.) Discuss:

- What effect does the tallest person in the world have on the mean? Why?"
 - The mean is pulled higher by the inclusion of this value. In terms of fair share, more would need to be taken from that value so that all of the heights are equal, so the mean would be greater.
- © "Suppose a new student who has a height close to the mean joins the class. Would that person's height affect the mean? Why or why not?"

Because every value is included, it would affect the mean in some way unless it was exactly the same as the mean. But, because the person's height is close to the mean, it should not change much.

"Does adding two values—one unusually high and one unusually low—affect the median? Why or why not?"

It does not. The middle value on the list does not change when one extreme value is added on each side.

Activity 2

Card Sort: Mean or Median?

20 min

Activity Narrative

In this lesson, students study distributions using dot plots and a histogram for which the mean and median can be the same, close, or far apart, and make conjectures about how the distributions affect the mean and median. Along the way, students recognize that the mean and median are equal or close when the distribution is roughly symmetrical and are farther apart when the distribution is non-symmetrical.

Students sort different distributions during this activity. A sorting task gives students opportunities to analyze representations, statements, and structures closely and to make connections.

As students work, encourage them to refine their descriptions of the distributions using more precise language and mathematical terms.

Launch



Arrange students in groups of 2 and distribute pre-cut cards. Allow students to familiarize themselves with the representations on the cards.

- Give students 1 minute to sort the cards into categories of their choosing.
- Pause the class after students have sorted the cards.
- Select groups to share their categories and how they sorted their cards/started sorting their cards.
- Discuss as many different types of categories as time allows.

Attend to the language that students use to describe their categories and distributions, giving them opportunities to describe their distributions more precisely. Highlight the use of terms like "symmetric" and "asymmetric." After a brief discussion, invite students to complete the remaining questions.

If not mentioned by students, highlight that in three of the distributions, the mean and median of the data are approximately equal. In the other three distributions, the mean and median are quite different. Discuss:

"What do you notice about the shape and features of distributions that have a roughly equal mean and median?"

They are roughly symmetrical and each have one peak in the middle, with roughly the same number of values to the left and right. They may have gaps, but the gaps are somewhat evenly spaced out.

"What about the shape and features of a distribution that has a very different mean and median?"

They are not at all symmetrical. They may have one peak, but it is off to one side, or they don't really show any peaks. They may have gaps or data values that are unusually high or low. There is more variability in these data sets.

Instructional Routines

Card Sort

ilclass.com/r/10783726

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



Instructional Routines

MLR3: Critique, Correct, Clarify

ilclass.com/r/10695504

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



Instructional Routines

Take Turns

ilclass.com/r/10573524

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



Access for Multilingual Learners (Activity 2)

MLR3: Critique, Correct, Clarify.

This activity uses the *Critique*, *Correct*, *Clarify* math language routine to advance representing and conversing as students critique and revise mathematical arguments.

Access for Students with Diverse Abilities (Activity 2, Student Task)

Action and Expression: Develop Expression and Communication.

To help get students started, display sentence frames such as "Cards in this category all have ..." or "The differences between these categories are ..."

Supports accessibility for: Language, Organization

"In the second group, why might the mean and the median be so different?"

The mean is pulled toward the direction of unusually large or small values. The median simply tells us where the middle of the data lies when sorted, so it is not as affected by these values that are far from where most data points are.

Student Task Statement

1. Your teacher will give you six cards. Each has either a dot plot or a histogram. Sort the cards into 2 piles based on the distributions shown. Be prepared to explain your reasoning.

Sample responses:

- The data points are more or less clustered together and their distribution is roughly symmetric. There are no values far from the center.
- The distributions are either clumped up on one side, have several clusters, or have values that are very far away from the rest of the data. They are not symmetric.
- The data points with significantly larger or smaller values affected the mean, which moves the balance point away from the data's middle value and toward one end or the other.
- **2.** Discuss your sorting decisions with another group. Did you have the same cards in each pile? If so, did you use the same sorting categories? If not, how are your categories different?

Pause here for a class discussion.

No answer required.

- **3.** Use the information on the cards to answer these questions.
 - **a.** Card A: What is a typical age of the dogs being treated at the animal clinic?

8 years old

- b. Card B: What is a typical number of people in the Irish households?5 people per household
- **c.** Card C: What is a typical travel time for the New Zealand students?

12.5 minutes

- **d.** Card D: Would 15 years old be a good description of a typical age of the people who attended the birthday party?
 - No, 15 years old would not be a good description of a typical age. The vast majority of the partygoers are 8 years old.
- **e.** Card E: Is 15 minutes or 24 minutes a better description of a typical time it takes the students in South Africa to get to school?
 - 15 would be a better description of a typical travel time because it is the middle value of the data set.
- **f.** Card F: Would 21.3 years old be a good description of a typical age of the people who went on a field trip to Washington, D.C.?
 - No, 21.3 years old would not be a good description of a typical age of the people on the field trip. Three-quarters of the people on the trip are 15 or 16 years old.

- **4.** How would you decide which measure of center to use for the dot plots on Cards A–C? What about for those on Cards D–F?
 - For Cards A-C, I could use either one, since the two measures of center are either identical or very close.
 - For Cards D-F, I chose the median because they represent the center of the data set better than the mean.

Are You Ready for More?

Most teachers use the mean to calculate a student's final grade, based on that student's scores on tests, quizzes, homework, projects, and other graded assignments.

Diego thinks that the median might be a better way to measure how well a student did in a course. Do you agree with Diego? Explain your reasoning.

Sample responses:

- I think that using the mean makes the most sense. Each assignment affects
 the final grade, so every assignment reflects the student's learning on all
 of the material the best.
- I think that median makes the most sense. When the mean is used, a grade
 of 0 (or 100) can greatly influence the final grade in one direction if that
 is far away from most of a student's grades. With the median, a few very
 low (or very high) grades will not influence the center as much, so the final
 grade will better reflect the general understanding of the student.

Activity Synthesis

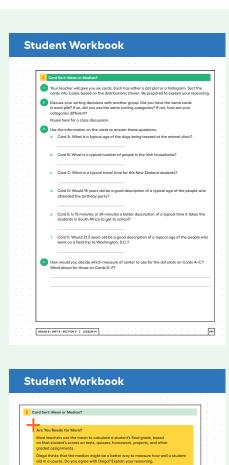
Use the whole-class discussion to reinforce the idea that the distribution of a data set can tell us which measure of center best summarizes what is typical for the data set. Briefly review the answers to the statistical questions, and then focus the conversation on the last questions (how students knew which measure of center to use in each situation). Select a couple of students to share their responses. Discuss:

- "For data sets with non-symmetrical distributions, why does the median turn out to be a better measure of center?"
 - Non-symmetrical data sets often have unusual values that pull the mean away from the center of data. The median is less influenced by these values.)
- "Does it matter which measure we choose to describe a typical value? For example, in Card F, would it matter if we said that a typical age for the people who went on the field trip to D.C. was about 21 years old?"

Yes, it does matter in some cases. In that example, it wouldn't really make sense to say that 21 years is a typical age because the vast majority of the people on the trip were teenagers.)

Use *Critique*, *Correct*, *Clarify* to give students an opportunity to improve a sample written response to the question, "When are median and mean likely to be close?", by correcting errors, clarifying meaning, and adding details.

- Display this first draft:
 - "They're the same when most of the points are in the middle."





Student Workbook **Description of the median are express of measuring the center of a distribution. They tell us signify efference things, however. The dot glot shows the median are express on 30 pages. The mean number of stickers is 20 (marked with a triangle). The median number of stickers is 20.5 (marked with a dismand). The mean tells us that if the number of stickers were distributed so that soch page has the same number, then each page would have 21 We could also this of 21 stickers to before. The median tells us that if the number of stickers were distributed so that soch page has the same number, then each page would have 21 We could also this of 21 stickers to before. The median tells us that of the page has nown than 50 stickers and half box less than 28 stickers on appage some through 55 distributed to this cose, both the mean and the median could describe a typical number of stickers on appage some through 55 distributed to the cose, both the mean and the median could describe a typical number of stickers on appage some through 55 distributed to the cose, both the median is 33 stickers. In this case, the median is closer to where meat of the data pairs as the stickers on a page. In this case, the median is closer to where meat of the data pairs as a stickers on a page. In the case, the median is closer to where meat of the data pairs as a stickers on a page. In the case, the median is done to where meat of the data pairs as inhumsced in this cose, build down by a hord full of pages with very less stickers, as its father own expression of the typical number of stickers in opage. The mean number of stickers is inhumsced in this case, pulled down by a hord full of pages with very less stickers, as its father own spring from mean cath pairs, and the stickers on a page. The mean number of stickers is inhumsced in this case, pulled down by a hord full of pages with very less stickers, as it is father own spring from mean cath pairs. In general, when a distribution is symmetrical or apparationally

- · Ask,
- "What parts of this response are unclear, incorrect, or incomplete?"
 As students respond, annotate the display with 2–3 ideas to indicate the parts of the writing that could use improvement.
 - Give students 2–4 minutes to work with a partner to revise the first draft.
 - Display and review these criteria:
- Specific words and phrases: "symmetric" or "approximately symmetric" distributions
- Select 1–2 individuals or groups to read their revised draft aloud slowly
 enough to record for all to see. Scribe as each student shares, then invite
 the whole class to contribute additional language and edits to make the
 final draft even more clear and more convincing.

Lesson Synthesis

The purpose of this discussion is to distinguish between mean and median. Ask students,

"For what kinds of distributions is the median the preferred measure of center?"

When there are a few values that are far from the center of the distribution or when the distribution is not very symmetric.

"In 2019, the mean household income in the United States was 99,312 and the median household income was 69,560. Why do you think the mean is almost 43% greater than the median?"

There are probably a few households for which the income is much greater than is typical for most people, so the mean is pulled in that direction.

Tell students that, in most situations, the mean is preferred if the two measures of center are close. The mean gives equal weight to each data point which means there is less influence of bias. In some cases, though, a few very different values on one side or the other of the center shifts the mean too far from the typical values to be useful. In these cases, the median can be useful.

Lesson Summary

Both the mean and the median are ways of measuring the center of a distribution. They tell us slightly different things, however.

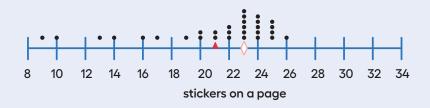
The dot plot shows the number of stickers on 30 pages. The mean number of stickers is 21 (marked with a triangle). The median number of stickers is 20.5 (marked with a diamond).



The mean tells us that if the number of stickers were distributed so that each page has the same number, then each page would have 21. We could also think of 21 stickers as a balance point for the number of stickers on all of the pages in the set.

The median tells us that half of the pages have more than 20.5 stickers and half have less than 20.5 stickers. In this case, both the mean and the median could describe a typical number of stickers on a page because they are fairly close to each other and to most of the data points.

Here is a different set of 30 pages with stickers. It has the same mean as the first set, but the median is 23 stickers.



In this case, the median is closer to where most of the data points are clustered and is therefore a better measure of center for this distribution. That is, it is a better description of the typical number of stickers on a page. The mean number of stickers is influenced (in this case, pulled down) by a handful of pages with very few stickers, so it is farther away from most data points.

In general, when a distribution is symmetrical or approximately symmetrical, the mean and median values are close. But when a distribution is not roughly symmetrical, the two values tend to be farther apart.

Cool-down

Which Measure of Center to Use?

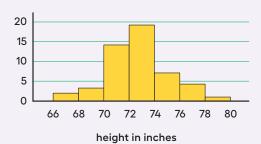
5 min

Student Task Statement

For each dot plot or histogram:

- **1.** Predict if the mean is greater than, less than, or approximately equal to the median. Explain your reasoning.
- **2.** Which measure of center—the mean or the median—better describes a typical value for the distributions?

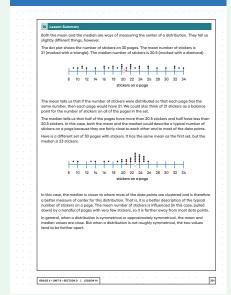
Heights of 50 basketball players



Sample responses:

- I. The mean would be approximately equal to the median, because the data are roughly symmetric.
- 2. Since I think the values would be pretty close, either the mean or the median would describe a typical height pretty well.

Student Workbook



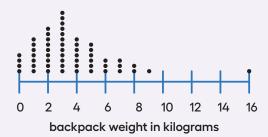
Responding To Student Thinking

Points to Emphasize

Look for opportunities to emphasize the difference between mean and median when looking at distributions throughout the rest of the unit.

Review this Cool-down by asking several students to share their solutions and reasoning after the activity: Unit 8, Lesson 15, Warm-up Notice and Wonder: Two Parties

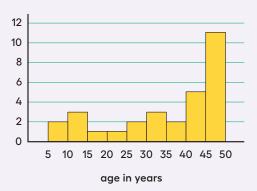
Backpack weights of 55 sixth-grade students



Sample responses:

- I. The mean would be higher than the median. The value of 16 kilograms would bring the mean up and move it away from the center of the data.
- 2. The median would better describe a typical backpack weight, since that value would lie in the center of the large cluster of data points.

Ages of 30 people at a family dinner party



Sample responses:

- I. The mean would be lower than the median, because even though a large fraction of the people at the dinner party are 40 or older, the ages of the people that span from 5 to 40 would bring the average age down.
- 2. The median would better describe the center of the distribution of around 40–45 years old.

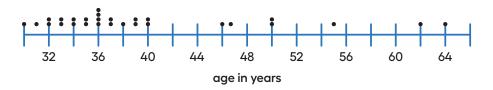
Practice Problems

6 Problems

Problem 1

Here is a dot plot that shows the ages of teachers at a school.

Which of these statements is true of the data set shown in the dot plot?



- A. The mean is less than the median.
- **B.** The mean is approximately equal to the median.
- **C.** The mean is greater than the median.
- **D.** The mean cannot be determined.

Problem 2

Priya asked each of five friends to attempt to throw a ball in a trash can until they succeeded. She recorded the number of unsuccessful attempts made by each friend as: 1, 8, 6, 2, 4. Priya made a mistake: The 8 in the data set should have been 18.

How would changing the 8 to 18 affect the mean and median of the data set?

- A. The mean would decrease and the median would not change.
- B. The mean would increase and the median would not change.
- C. The mean would decrease and the median would increase.
- **D.** The mean would increase and the median would increase.

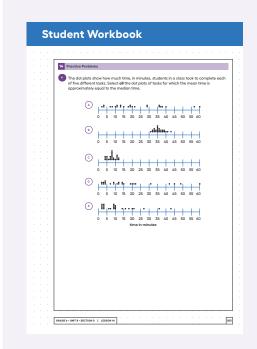
Problem 3

In his history class, Han's homework scores are:

100 100 100 100 95 100 90 100 0

The history teacher uses the mean to calculate the grade for homework. Write an argument for Han to explain why median would be a better measure to use for his homework grades.

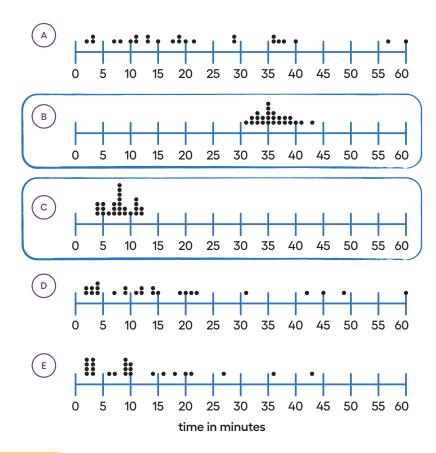
Sample response: The zero grade affects the mean much more than the median for these scores because it is so much lower than the others. Han does well on most of his homework and should not be punished so severely for the zero. The median value represents his typical understanding of the material more than the average value does.



| Treatice Problems | Trea

Problem 4

The dot plots show how much time, in minutes, students in a class took to complete each of five different tasks. Select **all** the dot plots of tasks for which the mean time is approximately equal to the median time.



Problem 5 from Unit 8, Lesson 2

Zookeepers recorded the ages, weights, genders, and heights of the 10 pandas at their zoo. Write two statistical questions that could be answered using these data sets.

Sample responses:

- · What is a typical age for the pandas at this zoo?
- What is a typical weight for the pandas at this zoo?
- Do most of the pandas weigh more than 200 pounds?
- · Are a majority of the pandas female?
- What is a typical height of the pandas at this zoo?
- Do the female pandas tend to weigh more than male pandas?

Problem 6

from Unit 7, Lesson 12

Here is a set of coordinates. Draw and label an appropriate pair of axes and plot the points. A = (1, 0), B = (0, 0.5), C = (4, 3.5), D = (1.5, 0.5)

Check student work to ensure that they made reasonable choices about axes and scale that allowed them to clearly plot all the points.

LESSON 14 • PRACTICE PROBLEMS 217