Lesson 14: Comparing mean and median

Goals

- Choose which measure of centre 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 centre.
- Generalise 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 centre 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 centre of a distribution in a given context. They learn that when the distribution is symmetrical, the mean and median have similar values. 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 (even if there is only one unusual value). 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 centre 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 centre.

Addressing

- Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
- Giving quantitative measures of centre (median and/or mean) and variability (interquartile range and/or range), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
- Relating the choice of measures of centre and variability to the shape of the data distribution and the context in which the data were gathered.

Instructional Routines

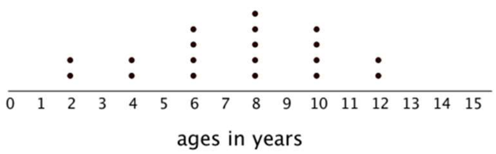
- Compare and Connect

- Discussion Supports
- Take Turns

Required Materials

Four-function calculators
Pre-printed cards, cut from copies of the blackline master

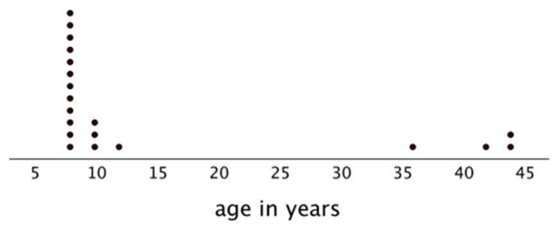
Mean or Median?
Card A
Ages of 20 dogs being treated at an animal hospital



Mean: Approximately 7.8 years old

Median: 8 years old

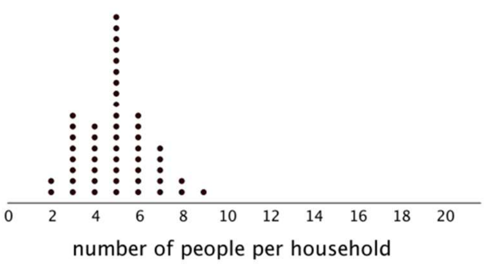
Mean or Median?
Card D
Ages of 20 people at an 8-year-old's birthday party



Mean: Approximately 15 years old

Median: 8 years old

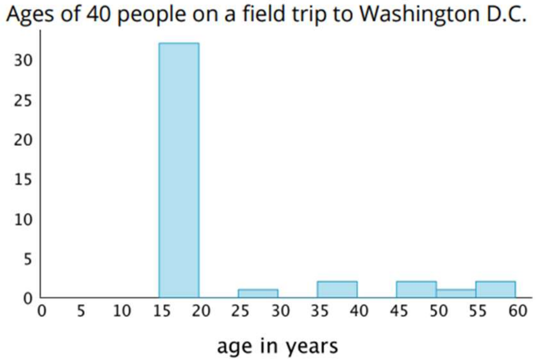
Mean or Median?
Card B
Size of household of 50 people in Ireland



Mean: Approximately 5 people per household

Median: 5 people per household

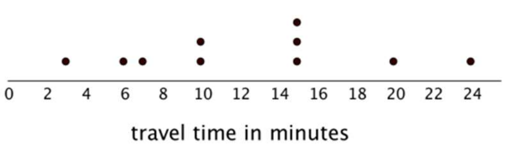
Mean or Median?
Card F
Ages of 40 people on a field trip to Washington D.C.



Mean: Approximately 21.3 years old

Median: 15 years old

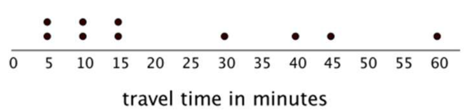
Mean or Median?
Card C
Time it took ten sixth-grade students in New Zealand to travel to school



Mean: Approximately 12.5 minutes

Median: 12.5 minutes

Mean or Median?
Card E
Time it took ten sixth-grade students in South Africa to travel to school



Mean: Approximately 24 minutes

Median: 15 minutes

Required Preparation

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.

For the Mean or Median activity, one copy of the blackline master for each group of 3-4 students cut into cards for sorting and examining.

Student Learning Goals

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

14.1 Heights of Presidents

Warm Up: 5 minutes

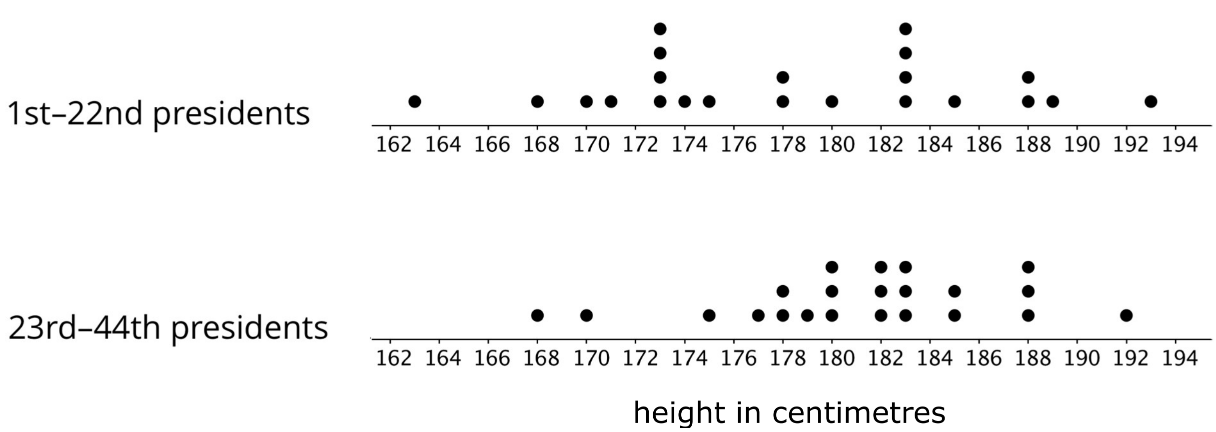
In this warm-up, students review ways to interpret and compare data shown on a dot plot. The discussion on each given statement gives the teacher an opportunity to hear how students reason about the median, mean, typical value, spread, balance point, and range. This discussion will be helpful in upcoming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



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 centimetres.

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

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

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

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

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

Student Response

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

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

3. Agree. The centre of the data appears to be about 182 centimetres.

4. Agree. The centre of the data in the second group is higher than in the first group.

5. 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. Disagree. Compared to the first group, the data points for the second group are clustered closer together, so their mean 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 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?"

14.2 The Tallest and the Smallest in the World

15 minutes

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 centre are affected when a value that is far from the centre 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.

Instructional Routines

- Compare and Connect

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. Give groups 8–10 minutes to complete the first three questions, and then ask them to pause for a brief class discussion before moving on to the last set of questions.

During this discussion, compare the mean that students calculated and the median they found, and solicit students' ideas on why the mean changed more than the median if the tallest person in the world joined the class. (If a class dot plot that shows student heights is made, add the point that represents the tallest person in the world.)

At this time students should begin to see that the value of the additional data point greatly affects the mean because finding the mean entails redistributing data values so that they are all the same (or moving the balance point toward the new data point to keep the distribution balanced). The new data point does not affect the median as much in this case because finding the median entails finding the point in a distribution that divides the data in half, and the numbers around the middle of the distribution are likely very close.

Give groups another few minutes to complete the rest of the task.

Representation: Internalise Comprehension. Represent the same information through different modalities by using physical objects to represent abstract concepts. Some students may benefit by starting with a physical demonstration of student's actual heights in the class.
Supports accessibility for: Conceptual processing; Visual-spatial processing Representing,

Conversing: Compare and Connect. Use this routine to support small-group discussion while students work on the task. Ask students, "What changes and what stays the same for each set of data when the heights of the world's tallest, and world's smallest adults are included?" Listen for students who generalise how the mean and median change when including the largest height. Amplify students' use of the targeted language such as, mean, median, spread of data, high/low/typical values, etc. This will help students use mathematical language to describe the effect of extreme data points on the mean and median.
Design Principle(s): Maximise meta-awareness

Anticipated Misconceptions

When calculating the mean after a new person joined 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 (e.g. 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 Task Statement

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

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

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

3. Suppose that the world's tallest adult, who is 251 centimetres tall, joined your class.

    a. Discuss the following questions with your group and explain your reasoning.

        - How would the mean height of the class change?

        - How would the median height change?

    b. Find the new mean.

    c. Find the new median.

    d. Which measure of centre—the mean or the median—changed more when this new person joined the class? Explain why the value of one measure changed more than the other.

4. The world's smallest adult is 63 centimetres tall. Suppose that the world's tallest and smallest adults both joined your class.

    a. Discuss the following questions with your group and explain your reasoning.

        - How would the mean height of the class change from the original mean?

- How would the median height change from the original median?

b. Find the new mean.

c. Find the new median.

d. How did the measures of centre—the mean and the median—change when these two people joined the class? Explain why the values of the mean and median changed the way they did.

Student Response

Answers vary for all questions.

3. d. Sample response: The mean changed more than the median when the new student is added. This is because the very large value of 251 is far from the centre. It pulled the mean and the balance point up by quite a bit. The median did not change very much because the middle value is simply the next point in the data set (or the average of the two middle data points).

4. d. Sample response: The new mean is a little higher than the original mean, and the median is unchanged. This is because the height difference between the world's tallest person and the original mean is larger than the difference between the world's smallest person and the original mean, so the new centre or balance point is pulled toward the higher end. The median is not changed because a new data point is added on each end, so the original middle value is also the new middle value.

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 changed 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 smallest person in the world have on the mean? Why?"
- "Which would affect the mean more: the height of the tallest person, or the height of the smallest person? Why?"
- "Suppose a new student who has a height close to the mean joined the class. Would her height affect the mean? Why or why not?"
- "Does adding two values—one unusually high and one unusually low—affect the median? Why or why not?"

Students should see that adding a data point to each end does not change the median much, if at all; the original middle value would still be the middle value as the halfway point is unchanged. Conversely, the mean will change greatly due to values that are extremely greater or less than most of the data.

14.3 Mean or Median?

15 minutes

In the previous activity, students analysed the effects of unusually high or low values on the mean and median. Here they study distributions (displayed 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 recognise 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.

Instructional Routines

- Discussion Supports
- Take Turns

Launch

Arrange students in groups of 3–4. Provide each group with a cut-up set of cards from the blackline master. Give groups 4–5 minutes to take turns sorting the cards and completing the first two problems. Then, pause the activity to discuss the sorting decisions and observations of the class.

Ask a few groups how they sorted the cards. 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 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 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.)
- "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.)

Afterwards, give students another 3–4 minutes to answer and discuss the remaining questions with their group.

Action and Expression: Develop Expression and Communication. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their ideas. For example, "Cars in this category all have . . ." or "The differences between these categories are . . ."

Supports accessibility for: Language; Organisation Speaking: Discussion Supports. Use this routine to support small-group discussion as students decide which measure of centre to use for the dot plots. Display the sentence frame: "The measure of centre that best represents card ___ is _____ because . . ." Listen for ways students relate the choice of measures of centre and variability to the shape of the data distribution.
Design Principle(s): Support sense-making

Student Task Statement

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

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.

3. Use the information on the cards to answer the following questions.

   a. Card A: What is a typical age of the dogs being treated at the animal clinic?

   b. Card B: What is a typical number of people in the Irish households?

   c. Card C: What is a typical travel time for the New Zealand students?

   d. Card D: Would 15 years old be a good description of a typical age of the people who attended the birthday party?

   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?

   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.?

4. How did you decide which measure of centre to use for the dot plots on Cards A–C? What about for those on Cards D–F?

Student Response

1. Answers vary. Sample responses:

   – The data points are more or less clustered together and their distribution is roughly symmetric. There are no values far from the centre.

   – 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. No answer required.

3.

   a. 8 years old.

   b. 5 people per household.

   c. 12.5 minutes.

   d. 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. 15 would be a better description of a typical travel time because it is the middle value of the data set.

   f. 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. Answers vary. Sample responses:

   - For Cards A–C, I could use either one, since the two measures of centre are either identical or very close.

   - For Cards D–F, I chose the median because they represent the centre 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.

Student Response

Answers vary. Sample responses:

- I think that 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 centre 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 centre best summarises 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 centre 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 centre for non-symmetrical data sets?" (Non-symmetrical data sets often have unusual values that pull the mean away from the centre 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.)

Lesson Synthesis

We see in the lesson that sometimes the two measures of centre could be the same or very close, but other times they could be very different.

- "When are the mean and median likely to be close together?" (When the distribution is approximately symmetrical.)
- "When are they likely to be different?" (When the distribution is not roughly symmetrical or has unusually high or low values that are far from others.)
- "Why might the median be a more useful measure of centre when the distribution is not symmetrical?" (Values far from the middle tend to have a greater influence on the mean than the median, so individual values can have a greater impact.)
- "In the situations we saw today, did it matter which measure we choose to describe a typical value?" (Yes, it did matter in some cases. For the 8-year-old's birthday party, it would not make sense to say that 15 years is a typical age for the partygoers.)
- "A student reports that 7 is a typical number of pets that students in her class has. Do you think she used the mean number of pets or the median? How do you know?" (The mean. That number seems too high to be a typical number of pets. The data may have included one or more students who have a tank of fish or other small animals that get counted individually, which would pull the mean up.)
- "Can you think of other real-world situations where reporting the mean or median can be misleading?" (Example: Salaries at a company with many low-level workers and one executive who gets paid a lot more than anyone else would be better reported with a median. If they had a job opening and reported the mean, it might make people think they will make more than they probably will.)

14.4 Which Measure of Centre to Use?

Cool Down: 5 minutes

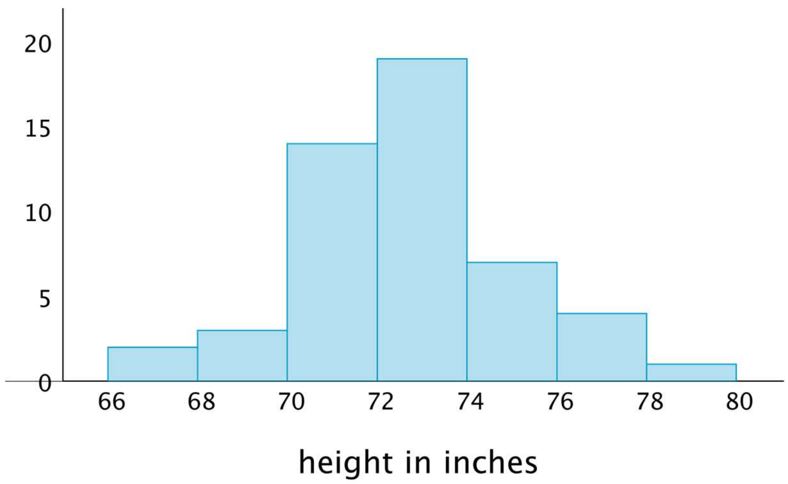
Student Task Statement

For each dot plot or histogram:

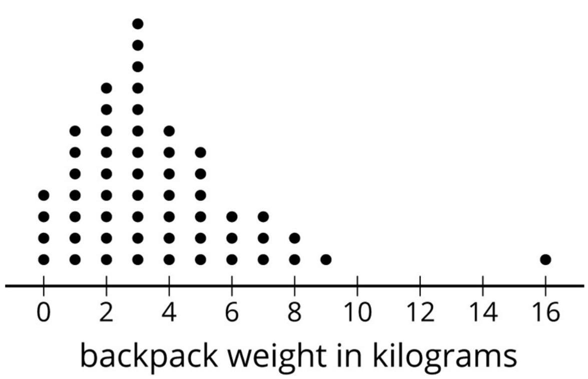
a. Predict if the mean is greater than, less than, or approximately equal to the median. Explain your reasoning.

b. Which measure of centre—the mean or the median—better describes a typical value for the following distributions?

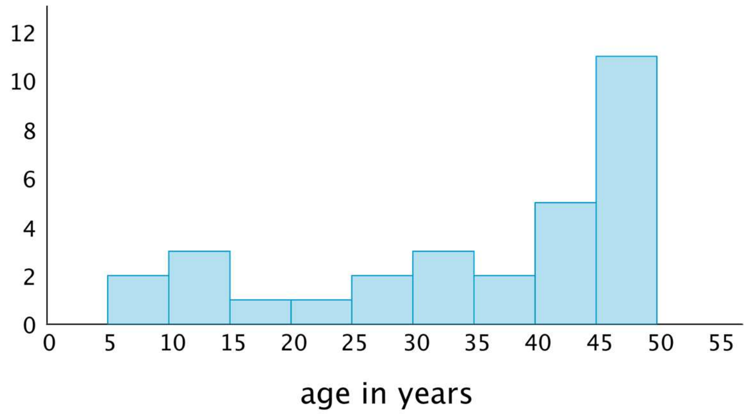
1. Heights of 50 NBA basketball players



2. Backpack weights of 55 year 7 students



3. Ages of 30 people at a family dinner party



Student Response

Answers vary. Sample responses:

1.

   a. The mean would be approximately equal to the median, because the data are roughly symmetric.

   b. Since I think the values would be pretty close, either the mean or the median would describe a typical height pretty well.

2.

   a. The mean would be higher than the median. The value of 16 kilograms would bring the mean up and move it away from the centre of the data.

   b. The median would better describe a typical backpack weight, since that value would lie in the centre of the large cluster of data points.

3.

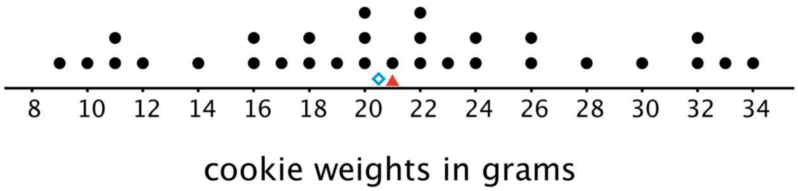
   a. The mean would be lower than the median, because even though a large fraction of the people at the dinner party are 40 or older, the ages of the people that span from 5 to 40 would bring the mean age down.

   b. The median would better describe the centre of the distribution of around 40–45 years old.

Student Lesson Summary

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

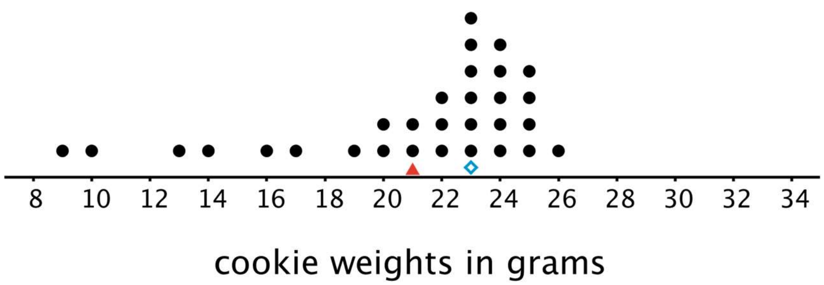
The dot plot shows the weights of 30 cookies. The mean weight is 21 grams (marked with a triangle). The median weight is 20.5 grams (marked with a diamond).



The mean tells us that if the weights of all cookies were distributed so that each one weighed the same, that weight would be 21 grams. We could also think of 21 grams as a balance point for the weights of all of the cookies in the set.

The median tells us that half of the cookies weigh more than 20.5 grams and half weigh less than 20.5 grams. In this case, both the mean and the median could describe a typical cookie weight because they are fairly close to each other and to most of the data points.

Here is a different set of 30 cookies. It has the same mean weight as the first set, but the median weight is 23 grams.



In this case, the median is closer to where most of the data points are clustered and is therefore a better measure of centre for this distribution. That is, it is a better description of a typical cookie weight. The mean weight is influenced (in this case, pulled down) by a handful of much smaller cookies, so it is farther away from most data points.

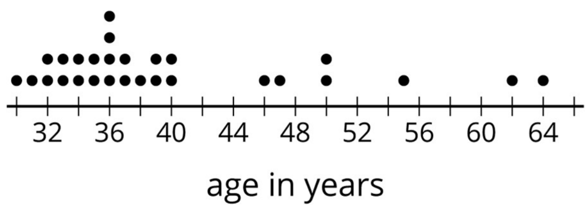
In general, when a distribution is symmetrical or approximately symmetrical, the mean and median values are close. But when a distribution is not roughly symmetrical, the two values tend to be further apart.

Lesson 14 Practice Problems

Problem 1 Statement

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.

Solution C

Problem 2 Statement

Priya asked each of five friends to attempt to throw a ball in a rubbish bin 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.

Solution B

Problem 3 Statement

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

100 100 100 100 95 100 90 100 0

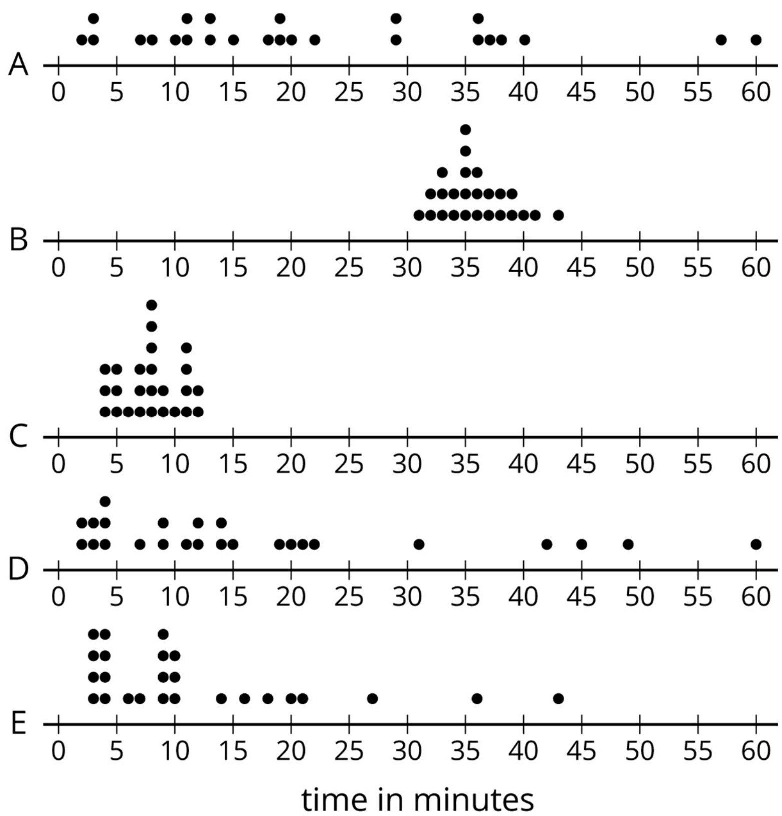
The history teacher uses the mean to calculate the grade for homework. Write an argument for Han to explain why median would be a better measure to use for his homework grades.

Solution

Answers vary. Sample response: The zero grade affects the mean much more than the median for these scores since 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 mean value does.

Problem 4 Statement

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.



Solution ["B", "C"]

Problem 5 Statement

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.

Solution

Answers vary. 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 Statement

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

Solution

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