Lesson 12: Describing the Center of a Distribution Using the Median

Student Outcomes
- Given a data set, students determine the median of the data.

Lesson Notes
The focus of this lesson is on using the median as a summary statistic to describe a data set. Students find the value of the median for data sets with an odd number of observations and for data sets with an even number of observations. Informally, they consider the variability among three different data sets to assess a claim about typical behavior. In preparation for a later lesson on finding quartiles, students calculate the median of the values below the median and the median of the values above the median. This lesson provides the background needed for the development of a box plot.

In this lesson, students construct arguments and critique the reasoning of others. They respond to the reasoning of others in some of the tasks, distinguish correct reasoning from flawed reasoning, and explain why incorrect reasoning is flawed. They also model with mathematics, apply mathematics to problems from everyday life, and interpret results in the context of the situation.

It should be noted that students should have access to calculators throughout this module.

Classwork
Students read the following paragraph silently.

How do we summarize a data distribution? What provides us with a good description of the data? The following exercises help us to understand how a numerical summary provides an answer to these questions.

Example 1 (5 minutes): The Median—A Typical Number
The activity begins with a set of data represented as a dot plot. The concept of median is then developed by having students consider a sequence of questions. Once the concept has been developed, the median is formally defined. Begin by introducing the data presented in the example.

Example 1: The Median—A Typical Number
Suppose a chain restaurant (Restaurant A) advertises that a typical number of french fries in a large bag is 82. The dot plot shows the number of french fries in a sample of twenty large bags from Restaurant A.
What could the restaurant mean when they say that the typical number of french fries in a large bag is 82?

- Answers will vary, but students may suggest that the mean is about 82.

Locate 82 on the dot plot. What do you notice about the number of data values that are above 82 and the number of data values that are below 82?

- There are the same number of data values on either side of 82—ten data values are greater than 82, and 10 data values are less than 82.

The restaurant used a summary measure called the median to describe the typical number of french fries. The median represents the middle value in a data set when the data values are arranged in order from smallest to largest. The same number of values will be above the median as are below the median.

Sometimes it is useful to know what point separates a data distribution into two equal parts, where one part represents the upper half of the data values and the other part represents the lower half of the data values. This point is called the median. When the data are arranged in order from smallest to largest, the same number of values will be above the median point as below the median.

Exercises 1–3 (4 minutes)

Students work independently on the exercises and confirm answers with a neighbor.

Exercises 1–3

1. You just bought a large bag of fries from the restaurant. Do you think you have exactly 82 french fries? Why or why not?

   The number of fries in a bag seems to vary greatly from bag to bag. No bag had exactly 82 fries, so mine probably will not. The bags that were in the sample had from 66 to 93 french fries.

2. How many bags were in the sample?

   20 bags were part of the sample.

3. Which of the following statement(s) would seem to be true for the given data? Explain your reasoning.
   a. Half of the bags had more than 82 fries in them.
   b. Half of the bags had fewer than 82 fries in them.
   c. More than half of the bags had more than 82 fries in them.
   d. More than half of the bags had fewer than 82 fries in them.
   e. If you got a random bag of fries, you could get as many as 93 fries.

   Statements (a) and (b) are true because there are 10 bags above 82 fries and 10 bags below 82 fries. Also, statement (e) is true because that happened once, so it could probably happen again.
Examples 2–4 (8 minutes)

As a class, work through the examples one at a time.

Example 2

Examine the dot plot below.

Grades on a Science Test

| 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |

a. How many data values are represented on the dot plot above?

_There are 28 data values on the dot plot._

b. How many data values should be located above the median? How many below the median? Explain.

_There should be 14 data values above the median and 14 data values below the median because the median represents the middle value in a sorted data set._

c. For this data set, 14 values are 80 or smaller, and 14 values are 85 or larger, so the median should be between 80 and 85. When the median falls between two values in a data set, we use the average of the two middle values. For this example, the two middle values are 80 and 85. What is the median of the data presented on the dot plot?

_The median of the dot plot is 82.5._

Students may need help determining the median when there are an even number of data values. If needed, guide them through the process with the discussion below.

- What is the middle data value?
  - _The median score must have 14 scores above and 14 scores below. If you start counting from the smallest score, you find that the 14th score is 80, so 14 scores are 80 or less. The next score is 85, and there are 14 scores that are 85 or greater. This means that the median must be between 80 and 85._

- What would you suggest as a score to represent the median? Explain your answer.
  - _Answers will vary. Students may select the value that is the middle of the interval from 80 to 85, or they may say that any number in this interval could be the median. Indicate that the usual thing to do when there are an even number of observations is to define the median to be the mean (average) of the middle two values. In this example, the middle two values are 80 and 85:_

\[
\frac{80 + 85}{2} = \frac{165}{2} = 82.5
\]
d. What does this information tell us about the data?

The median tells us half of the students in the class scored below an 82.5 on the science test, and the other half of the students scored above an 82.5 on the science test.

Example 3
Use the information from the dot plot in Example 2.

a. What percentage of students scored higher than the median? Lower than the median?

50% of the students scored higher than the median, and 50% of the students scored lower than the median.

b. Suppose the teacher made a mistake, and the student who scored a 65 actually scored a 71. Would the median change? Why or why not?

The median would not change because there would still be 14 scores below 82.5 and 14 scores above 82.5.

c. Suppose the student who scored a 65 actually scored an 89. Would the median change? Why or why not?

The median would change because now there would be 13 scores below 82.5 and 15 scores above 82.5, so 82.5 would not be the median.

Example 4
A grocery store usually has three checkout lines open on Saturday afternoons. One Saturday afternoon, the store manager decided to count how many customers were waiting to check out at 10 different times. She calculated the median of her ten data values to be 8 customers.

a. Why might the median be an important number for the store manager to consider?

Answers will vary. For example, students might point out that this means that half the time there were more than 8 customers waiting to check out. If there are only 3 checkout lines open, there would be a lot of people waiting to check out. She might want to consider having more checkout lines open on Saturday afternoons.

b. Give another example of when the median of a data set might provide useful information. Explain your thinking.

Answers will vary. Possible responses: When the data are about how much time students spend doing homework, it would be interesting to know the amount of time that more than half of the students spend on homework. If you are looking at the number of points earned in a competition, it would be good to know what number separates the top half of the competitors from the bottom half.

How do you find the median if there are an odd number of data points?

- Put the data values in order from smallest to largest, or construct a dot plot of the data.
- Find the middle number in the ordered list or on the dot plot. One way to do this is to divide the number of observations by 2 and then round up to get an integer. This identifies the position of the median. For example, if there are 15 observations, dividing by 2 gives us 7.5, which rounds up to 8. Then, starting with the smallest observation, count up to find the 8th number in the ordered list. This number is the median.
How do you find the median if there are an even number of data points?

- Put the data values in order from smallest to largest, or construct a dot plot of the data.
- Find the middle two numbers in the ordered list or on the dot plot. One way to do this is to divide the number of observations by 2. This identifies the position of the first of the two middle numbers. For example, if there are 18 observations, dividing by 2 gives us 9 observations. Then, starting with the smallest observation, count up to find the 9th number in the ordered list. This number and the next number in the list are the two middle values.
- Find the mean of the two middle values. This number is the median.

Exercises 4–5 (10 minutes): A Skewed Distribution

In this set of exercises, students have to put the data in order from smallest to largest before they find the median. There are 19 values, so the median is the 10th value with 9 data values above and 9 data values below. Another way to determine the median after ordering the data is to cross out the maximum and minimum values, then the next largest and smallest values, and so on until students are left with just one number in the middle if there are an odd number of data values, or two numbers if there are an even number of data values. If this process results in a single number, that number is the median. If this process results in two numbers in the middle, students would then find the mean of the two values to get the value of the median.

The questions are designed to help students confront some common misconceptions and errors: not ordering the data before counting to the middle, confusing median and mode (most frequent value), and confusing median and midrange (halfway between the maximum and the minimum). They also compute the mean and compare the median to the mean, noting that the median might be more reflective of a typical value because several bags with low numbers of french fries pulled the mean down.

Consider the following questions as students are completing the exercises:

- Why is it necessary to order the data before you find the median?
  - In order to find the value that has half of the data values smaller and half of the data values larger, the data values must first be put in order.
- Is the median connected to the range (maximum — minimum) of the data? Why or why not?
  - The median is not connected to the range because it is possible for the range to change while the median stays the same.
- What is the difference in the effect of a few very extreme values on the mean and on the median?
  - Extreme values will have a big effect on the mean but will not affect the median.

Exercises 4–5: A Skewed Distribution

4. The owner of the chain decided to check the number of french fries at another restaurant in the chain. Here are the data for Restaurant B: 82, 83, 83, 79, 85, 82, 78, 76, 76, 75, 78, 74, 70, 60, 82, 82, 83, 83, 83
   a. How many bags of fries were counted?

   19 bags of fries were counted.
b. Sallee claims the median is 75 because she sees that 75 is the middle number in the data set listed on the previous page. She thinks half of the bags had fewer than 75 fries because there are 9 data values that come before 75 in the list, and there are 9 data values that come after 75 in the list. Do you think she would change her mind if the data were plotted in a dot plot? Why or why not?

Yes. You cannot find the median unless the data are organized from least to greatest. Plotting the number of fries in each bag on a dot plot would order the data correctly. You would probably get a different halfway point because the data above are not ordered from least to greatest.

c. Jake said the median was 83. What would you say to Jake?

83 is the most common number of fries in the bags (5 bags had 83 fries), but it is not in the middle of the data.

d. Betse argued that the median was halfway between 60 and 85, or 72.5. Do you think she is right? Why or why not?

She is wrong because the median is not calculated from the distance between the largest and smallest values in the data set. This is not the same as finding a point that separates the ordered data into two parts with the same number of values in each part.

e. Chris thought the median was 82. Do you agree? Why or why not?

Chris is correct because if you order the numbers, the middle number will be the 10th number in the ordered list, with at most 9 bags that have more than 82 fries and at most 9 bags that have fewer than 82 fries.

5. Calculate the mean, and compare it to the median. What do you observe about the two values? If the mean and median are both measures of center, why do you think one of them is smaller than the other?

The mean is 78.6, and the median is 82. The bag with only 60 fries decreased the value of the mean.

Exercises 6–8 (10 minutes): Finding Medians from Frequency Tables

MP.4

In these exercises, students find the median using data summarized in a frequency table. The median falls halfway between the 13th and 14th data value when the data are ordered from smallest to largest. They also find the medians of the top and bottom halves of the data set, the 7th value from the top and from the bottom, as a precursor to finding quartiles and the interquartile range in a later lesson. Consider having students write out the individual counts in a long ordered list. For example, the first 13 counts would be as follows:

Median of the lower half

75 75 76 77 77 78 78 79 79 79 79 79 79 ...
In these exercises, students need to deal with repeated data values when finding the median. In this case, have students find the median by counting from the top and bottom of the list, noting that values for bags with the same count can fall on both sides of the median. It might help to think about the individual bags: One of the bags with 78 fries is in the first half, one of the bags with 78 fries is in the second half, and one of the bags divides the two halves and marks the median of the data set. At this point, the important idea is that students get a sense of how to find a median: Order the values, and find a middle value in the ordered list of data values.

### Exercises 6–8: Finding Medians from Frequency Tables

6. A third restaurant (Restaurant C) tallied the number of fries for a sample of bags of french fries and found the results below.

<table>
<thead>
<tr>
<th>Number of Fries</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td></td>
</tr>
</tbody>
</table>

   a. How many bags of fries did they count?

   *They counted 26 bags of fries.*

   b. What is the median number of fries for the sample of bags from this restaurant? Describe how you found your answer.

   *79.5; I took half of 26, which is 13, and then counted 13 tallies from 86 to reach 80. I also counted 13 tallies from 75 to reach 79. The point halfway between 79 and 80 is the median.*

7. Robere wanted to look more closely at the data for bags of fries that contained a smaller number of fries and bags that contained a larger number of fries. He decided to divide the data into two parts. He first found the median of the whole data set and then divided the data set into the bottom half (the values in the ordered list that are before the median) and the top half (the values in the ordered list that are after the median).

   a. List the 13 values in the bottom half. Find the median of these 13 values.

   *75 75 76 77 77 78 78 78 79 79 79 79 79*

   *The median of the lower half is 78.*

   b. List the 13 values of the top half. Find the median of these 13 values.

   *80 80 80 81 82 84 84 84 85 85 85 86*

   *The median of the top half is 84.*
8. Which of the three restaurants seems most likely to really have 82 fries in a typical bag? Explain your thinking.

Answers will vary. The data sets for Restaurants A and B both have a median of 82. Look for answers that consider how much the data values vary around 82. Restaurant B seems to have the most bags closest to a count of 82. The data set for Restaurant C has a median of 79.5, but the data values are not very spread out, and most are close to 82, so some students might make a case for Restaurant C.

Closing (3 minutes)

- Does the median have to be a value in the data set?
  - The median does not have to be a value in the data set.

- Is calculating the median the same as calculating the middle of the range? Explain.
  - The median is not the same as calculating the middle of the range. Finding the point that is halfway between the largest and the smallest data value is not the same as finding a value that will have half of the data values above and half of the data values below. For example, think about the data set consisting of 1, 2, 3, 4, 19. The median is 3, but halfway between the largest and smallest data values is 10.

Lesson Summary

The median is the middle value (or the mean of the two middle values) in a data set that has been ordered from smallest to largest. The median separates the data into two parts with the same number of data values below the median as above the median in the ordered list. To find a median, you first have to order the data. For an even number of data values, you find the average of the two middle numbers. For an odd number of data values, you use the middle value.

Exit Ticket (5 minutes)
Lesson 12: Describing the Center of a Distribution Using the Median

Exit Ticket

1. What is the median age for the following data set representing the ages of students requesting tickets for a summer band concert? Explain your reasoning.

   13 14 15 15 16 16 17 18

2. What is the median number of diseased trees from a data set representing the numbers of diseased trees on each of 12 city blocks? Explain your reasoning.

   11 3 3 4 6 12 9 3 8 8 8 1

3. Describe how you would find the median for a set of data that has 35 values. How would this be different if there were 36 values?
Exit Ticket Sample Solutions

1. What is the median age for the following data set representing the ages of students requesting tickets for a summer band concert? Explain your reasoning.

\begin{align*}
13 & 14 & 15 & 16 & 17 & 18 & 18
\end{align*}

The median is the 5th value in the ordered list, or 16 years, as there are 4 values less than 16 and 4 values greater than or equal to 16 (excluding the 5th value).

2. What is the median number of diseased trees from a data set representing the numbers of diseased trees on each of 12 city blocks? Explain your reasoning.

\begin{align*}
11 & 3 & 3 & 4 & 6 & 12 & 9 & 3 & 8 & 8 & 1
\end{align*}

To find the median, the values first need to be ordered: 1 3 3 3 4 6 8 8 9 11 12.

Because there are an even number of data values, the median would be the mean of the 6th and 7th values: \(\frac{6 + 8}{2}\), or 7 diseased trees.

3. Describe how you would find the median for a set of data that has 35 values. How would this be different if there were 36 values?

Answers will vary. First, you would order the data from least to greatest. Because there are 35 values, you would look for the 18th value from the top or bottom in the ordered list. This would be the median with 17 values above and 17 values below. If the set had 36 values, you would find the average of the middle two data values, which would be the average of the 18th and the 19th values in the ordered list.

Problem Set Sample Solutions

1. The amount of precipitation in each of the western states in the United States is given in the table as well as the dot plot.

<table>
<thead>
<tr>
<th>State</th>
<th>Amount of Precipitation (inches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
<td>38.4</td>
</tr>
<tr>
<td>OR</td>
<td>27.4</td>
</tr>
<tr>
<td>CA</td>
<td>22.2</td>
</tr>
<tr>
<td>MT</td>
<td>15.3</td>
</tr>
<tr>
<td>ID</td>
<td>18.9</td>
</tr>
<tr>
<td>WY</td>
<td>12.9</td>
</tr>
<tr>
<td>NV</td>
<td>9.5</td>
</tr>
<tr>
<td>UT</td>
<td>12.2</td>
</tr>
<tr>
<td>CO</td>
<td>15.9</td>
</tr>
<tr>
<td>AZ</td>
<td>13.6</td>
</tr>
<tr>
<td>NM</td>
<td>14.6</td>
</tr>
<tr>
<td>AK</td>
<td>58.3</td>
</tr>
<tr>
<td>HI</td>
<td>63.7</td>
</tr>
</tbody>
</table>

Lesson 12: Describing the Center of Distribution Using the Median

a. How do the amounts vary across the states?
   Answers will vary. The spread is pretty large: 54.2 inches. Nevada has the lowest precipitation at 9.5 inches per year. Hawaii, Alaska, and Washington have more rain than most of the states. Hawaii has the most precipitation with 63.7 inches, followed by Alaska at 58.3 inches.

b. Find the median. What does the median tell you about the amount of precipitation?
   The median is 15.9 inches. Half of the western states have more than 15.9 inches of precipitation per year, and half have less.

c. Do you think the mean or median would be a better description of the typical amount of precipitation? Explain your thinking.
   The mean at 24.8 inches reflects the extreme values, while the median seems more typical at 15.9 inches.

2. Identify the following as true or false. If a statement is false, give an example showing why.
   a. The median is always equal to one of the values in the data set.
      *False. If the middle two values in the ordered data set are 1 and 5, the median is 3, and 3 is not in the set.*

   b. The median is halfway between the least and greatest values in the data set.
      *False. For example, looking at the number of french fries per bag for Restaurant A in Example 1, the median is 82, which is not halfway between 66 and 93 (?9.5).*

   c. At most, half of the values in a data set have values less than the median.
      *True*

   d. In a data set with 25 different values, if you change the two smallest values in the data set to smaller values, the median will not be changed.
      *True*

   e. If you add 10 to every value in a data set, the median will not change.
      *False. The median will increase by 10 as well. If the data set is 1, 2, 3, 4, 5, the median is 3. For the data set 11, 12, 13, 14, 15, the median is 13.*

3. Make up a data set such that the following is true:
   a. The data set has 11 different values, and the median is 5.
      *Answers will vary. If the numbers are whole numbers, the set would be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

   b. The data set has 10 values, and the median is 25.
      *Answers will vary. One answer is to have ten values that are all 25’s.*

   c. The data set has 7 values, and the median is the same as the least value.
      *Answers will vary. One answer is to have 1, 1, 1, 2, 3, 4.
4. The dot plot shows the number of landline phones that a sample of people have in their homes.

![Dot plot]

<table>
<thead>
<tr>
<th>Number of Phones</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. How many people were in the sample?

*There are 25 people in the sample.*

b. Why do you think three people have no landline phones in their homes?

*Possible answers: Some people might only have cell phones, or some people may not be able to afford a phone or may not want a phone.*

c. Find the median number of phones for the people in the sample.

*The median number of phones per home is 2.*

5. The salaries of the Los Angeles Lakers for the 2012–2013 basketball season are given below. The salaries in the table are ordered from largest to smallest.

<table>
<thead>
<tr>
<th>Player</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kobe Bryant</td>
<td>$27,849,149</td>
</tr>
<tr>
<td>Dwight Howard</td>
<td>$19,536,360</td>
</tr>
<tr>
<td>Pau Gasol</td>
<td>$19,000,000</td>
</tr>
<tr>
<td>Steve Nash</td>
<td>$8,700,000</td>
</tr>
<tr>
<td>Metta World Peace</td>
<td>$7,258,960</td>
</tr>
<tr>
<td>Steve Blake</td>
<td>$4,000,000</td>
</tr>
<tr>
<td>Jordan Hill</td>
<td>$3,563,600</td>
</tr>
<tr>
<td>Chris Duhon</td>
<td>$3,500,000</td>
</tr>
<tr>
<td>Jodie Meeks</td>
<td>$1,500,000</td>
</tr>
<tr>
<td>Earl Clark</td>
<td>$1,240,000</td>
</tr>
<tr>
<td>Devin Ebanks</td>
<td>$1,054,309</td>
</tr>
<tr>
<td>Darius Morris</td>
<td>$962,195</td>
</tr>
<tr>
<td>Antawn Jamison</td>
<td>$854,389</td>
</tr>
<tr>
<td>Robert Sacre</td>
<td>$473,604</td>
</tr>
<tr>
<td>Darius Johnson-Odom</td>
<td>$203,371</td>
</tr>
</tbody>
</table>

*Source: www.basketball-reference.com/contracts/LAL.html*

**Teacher Note:** Students may struggle a bit with Problems 5 and 6 because the numerical values are quite large. If this presents a challenge for students, these two problems could be done as part of a class discussion. The points being made in these two problems (how the mean can be affected by extreme values in a data set and the difference between how the mean and the median are affected by extreme values) are important, so it is worth taking time to make sure students see these problems.

a. Just looking at the data, what do you notice about the salaries?

*Possible answer: A few of the salaries for the big stars like Kobe Bryant are really big, while others are very small in comparison.*

b. Find the median salary, and explain what it tells you about the salaries.

*The median salary is $3,500,000 for Chris Duhon. Half of the players make more than $3,500,000, and half of the players make less than $3,500,000.*
c. Find the median of the lower half of the salaries and the median of the upper half of the salaries.

$962,195 \text{ is the median for the bottom half of the salaries. } \ 8,700,000 \text{ is the median for the top half of the salaries.}$

d. Find the width of each of the following intervals. What do you notice about the size of the interval widths, and what does that tell you about the salaries?

i. Minimum salary to the median of the lower half: $758,824$

ii. Median of the lower half to the median of the whole data set: $2,537,805$

iii. Median of the whole data set to the median of the upper half: $8,200,000$

iv. Median of the upper half to the highest salary: $19,149,149$

The largest width is from the median of the upper half to the highest salary. The smaller salaries are closer together than the larger ones.

6. Use the salary table from the previous page to answer the following.

a. If you were to find the mean salary, how do you think it would compare to the median? Explain your reasoning.

Possible answer: The mean will be a lot larger than the median because when you add in the really big salaries, the size of the mean will increase a lot.

b. Which measure do you think would give a better picture of a typical salary for the Lakers, the mean or the median? Explain your thinking.

Possible answer: The median seems better, as it is more typical of most of the salaries.