Lesson 2: Describing the Center of a Distribution

Student Outcomes

- Students construct a dot plot from a data set.
- Students calculate the mean of a data set and the median of a data set.
- Students observe and describe that measures of center (mean and median) are nearly the same for distributions that are nearly symmetrical.
- Students observe and explain why the mean and median are different for distributions that are skewed.
- Students select the mean as an appropriate description of center for a symmetrical distribution and the median as a better description of center for a distribution that is skewed.

Lesson Notes

In Grade 6, students were introduced to center as a description of a typical value in a data set. In Grades 6 and 7, students used the mean or the median as a description of a typical value of data distribution, looking at the shape of the distribution to determine whether the median or the mean was a better description of a typical value. This lesson continues to expand their understanding and interpretation of the center of a distribution.

MP.1 Students are provided three data sets and asked to construct dot plots for each in order to study the distributions. A symmetrical or a nearly symmetrical distribution emerges from the first data set. Students determine that the mean and median are nearly the same in this distribution. In the second and third data sets, a non-symmetrical distribution is given. Students determine that the mean and the median are not the same. They are asked to explain why the two measures of center are not equal.

In closing, students are asked to generalize what shape of distribution would have a median that is less than the mean, and what shape of distribution would have a median that is greater than the mean.

Solving the problems in the exercises provides an opportunity to address the modeling cycle outlined in the conceptual category of modeling of the high school Common Core State Standards. The following graphic from the Common Core State Standards for high school mathematics summarizes the modeling cycle. As students work through the exercise, observe and point out to them how to **formulate** a summary of the data by constructing a dot plot. Students observe that the dot plot identifies a center that is appropriate for the data distribution. Students **compute** a measure of center, **interpret** the center given the context of the data set, and then **validate** that the center provides a description of a typical value by re-examining the data distribution. In some cases, students are expected to revisit their decision, select a new measure of center, interpret the value of the new center, and validate that it provides a description of a typical value. Students should be made familiar with the modeling cycle as it is designed to guide their thinking not only in this module but in all of their mathematics course work. If time permits at the end of this lesson, ask students to reflect on this cycle and how solving the questions in the exercise connect to this cycle.
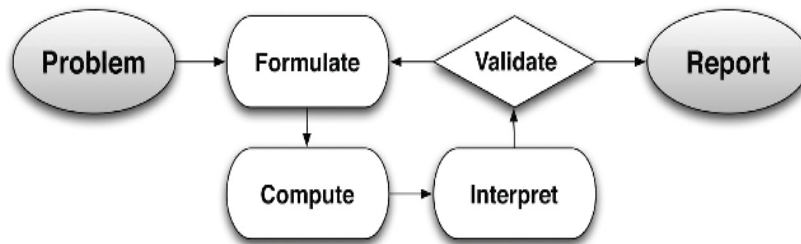
Classwork

In previous work with data distributions, you learned how to derive the mean and the median of a data distribution. This lesson builds on your previous work with a center.

Exploratory Challenge (25 minutes)

Introduce each data set by discussing the data. Discuss the following questions before students complete the problems in the exercise to help them understand the context of the problems. Direct students to examine the data sets presented in the exercise of the lesson.

Exploratory Challenge

Consider the following three sets of data.

Data Set 1: Pet owners

Students from River City High School were randomly selected and asked, "How many pets do you currently own?" The results are recorded below.

0	0	0	0	1	1	1	1	1	1	1	1	1	1	2
2	2	2	3	3	4	5	5	6	6	7	8	9	10	12

- Can we assume that all students will interpret the question "How many pets do you currently own?" in the same way?

Data Set 2: Length of the east hallway at River City High School

Twenty students were selected to measure the length of the east hallway. Two marks were made on the hallway's floor, one at the front of the hallway, and one at the end of the hallway. Each student was given a meter stick and asked to use the meter stick to determine the length between the marks to the nearest tenth of a meter. The results are recorded below.

8.2	8.3	8.3	8.4	8.4	8.5	8.5	8.5	8.5	8.5
8.6	8.6	8.6	8.6	8.7	8.7	8.8	8.8	8.9	8.9

- Why would the same hallway have different reported measures of length?
- What measures of the length of the hallway do you think are the most accurate from the data set?

Data Set 3: Age of cars

Twenty-five car owners were asked the age of their cars in years. The results are recorded below.

0	1	2	2	3	4	5	5	6	6	6	7	7
7	7	7	7	8	8	8	8	8	8	8	8	

- What number would you use to describe the typical age of cars in years by the car owners in this group?

Students now work through Exercise 1–9 of their lesson. Students determine for each distribution if the mean and the median are different or similar and how those characteristics relate to the distribution. Students relate their answers in the context of the data set.

Several of the questions within the exercise can be posed verbally as part of a discussion with students. Other questions may require written responses. Review the questions and decide what format is best for your students.

1. **Make dot plot of each of the data sets. Use the following scales.**
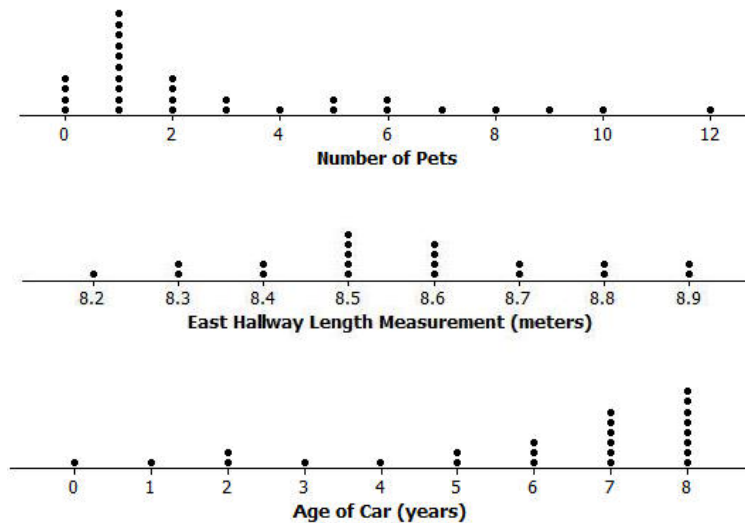
2. **Calculate the mean number of pets owned by the thirty students from River City High School. Calculate the median number of pets owned by the thirty students.**

    The mean number of pets owned is 3.2.

    The median number of pets owned is 2.

3. **What do you think is a typical number of pets for students from River City High School? Explain how you made your estimate.**

    The median is a better description of a typical value. The median is less than the mean because more of the data is located on the left side of the scale. The value of the mean is "pulled up" by the larger values in the data set.

4. **Why do you think that different students got different results when they measured the same distance of the east hallway?**

    Measurement errors occurred as a result of the placement of the ruler at the starting point, the alignment of the ruler after each length was made, and the different interpretation made by the person making the measurement.

5. **What is the mean length of the east hallway data set? What is the median length?**

    The mean length is approximately 8.57 meters.

    The median length is approximately 8.55 meters.

    They are approximately the same.

6. **A construction company will be installing a handrail along a wall from the beginning point to the ending point of the east hallway. The company asks you how long the handrail should be. What would you tell the company? Explain your answer.**

    It should be approximately 8.6 meters. Since the distribution is nearly symmetric, the mean should be used as a measure of center.

7. **Describe the distribution of the age of cars.**

    The distribution is not symmetric. It is skewed to the left. Most of the data is on the right-hand side with a long tail to the left.

8. **What is the mean age of the twenty-five cars? What is the median age? Why are the mean and the median different?**

    The mean age is approximately 5.84 years old, and the median age is 7 years old. They are different. The mean is smaller because of the small values in the tail of the data distribution.

9. **What number would you use as an estimate of the typical age of a car for the twenty-five car owners? Explain your answer.**

    The median is a better description of a typical age because the distribution is skewed.

Closing (10 minutes)

Ask the following of students:

▪   How is the choice made between mean and median to describe the typical value related to the shape of the data distribution?

▪   Sketch a dot plot in which the median is greater than the mean. Could you think of a context that might result in data where you think that would happen?

Lesson Summary

▪   **A dot plot provides a graphical representation of a data distribution, helping us to visualize the distribution.**

▪   **The mean and the median of the distribution are numerical summaries of the center of a data distribution.**

▪   **When the distribution is nearly symmetrical, the mean and the median of the distribution are approximately equal. When the distribution is not symmetrical (often described as skewed), the mean and the median are not the same.**

▪   **For symmetrical distributions, the mean is an appropriate choice for describing a typical value for the distribution. For skewed data distributions, the median is a better description of a typical value.**

Exit Ticket (10 minutes)

Name _____   Date_____

Lesson 2: Describing the Center of a Distribution

Exit Ticket

Each person in a random sample of ten ninth graders was asked two questions:
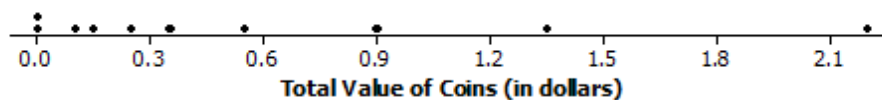
- How many hours did you spend watching TV last night?
- What is the total value of the coins you have with you today?

Here are the data for these ten students:

Student	Hours of TV	Total Value of Coins (in dollars)
1	2	0.00
2	1	0.89
3	0	2.19
4	3	0.15
5	4	1.37
6	1	0.36
7	2	0.25
8	2	0.00
9	4	0.54
10	3	0.10

1. Construct a dot plot of the data on Hours of TV. Would you describe this data distribution as approximately symmetric or as skewed?

2. If you wanted to describe a typical number of hours of TV watched for these ten students, would you use the mean or the median? Calculate the value of the measure you selected.

3. Here is a dot plot of the data on total value of coins.

Calculate the values of the mean and the median for this data set.

4. Why are the values of the mean and the median that you calculated in question 3 so different? Which of the mean and the median would you use to describe a typical value of coins for these ten students?
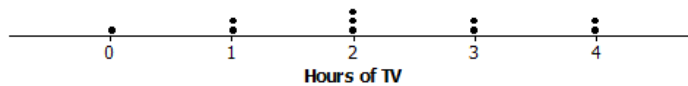
Exit Ticket Sample Solutions

Each person in a random sample of ten ninth graders was asked two questions:

- How many hours did you spend watching TV last night?

- What is the total value of the coins you have with you today?

Here are the data for these ten students:

Student	Hours of TV	Total Value of Coins (in dollars)
1	2	0.00
2	1	0.89
3	0	2.19
4	3	0.15
5	4	1.37
6	1	0.36
7	2	0.25
8	2	0.00
9	4	0.54
10	3	0.10

1. Construct a dot plot of the data on hours of TV. Would you describe this data distribution as approximately symmetric or as skewed?


This distribution is approximately symmetric.

2. If you wanted to describe a typical number of hours of TV for these ten students, would you use the mean or the median? Calculate the value of the measure you selected.

Because the distribution is approximately symmetric, I would use the mean. The value of the mean is $\frac{22}{10} = 2.2$ hours.

3. Here is a dot plot of the data on total value of coins.
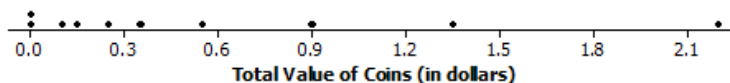
Calculate the values of the mean and the median for this data set.

Mean $= \frac{5.85}{10} = 0.585$, or about 59 cents. Arranging the data in order from smallest to largest, you get

| 0.00 | 0.00 | 0.10 | 0.15 | 0.25 | 0.36 | 0.54 | 0.89 | 1.37 | 2.19 |

The median is the average of the two middle numbers $\frac{0.25 + 0.36}{2} = 0.305$, or about 31 cents.

4. Why are the values of the mean and the median that you calculated in question 3 so different? Which of the mean and the median would you use to describe a typical value of coins for these ten students?

The values of the mean and median are different because the data distribution is skewed, and the mean is pulled up by the large values in the data set. Because the data distribution is skewed, the median would be a better choice for describing a typical value.

Problem Set Sample Solutions

Consider the following scenario. The company that created a popular video game, "Leaders," plans to release a significant upgrade of the game. Users earn or lose points for making decisions as the leader of an imaginary country. In most cases, repeated playing of the game improves a user's ability to make decisions. The company will launch an online advertising campaign, but at the moment, they are not sure how to focus the advertising. Your goal is to help the company decide how the advertising campaign should be focused. Five videos have been proposed for the following target audiences:

Video 1: Target females with beginning level scores

Video 2: Target males with advanced level scores

Video 3: Target all users with middle range level scores

Video 4: Target males with beginning level scores

Video 5: Target females with advanced level scores

1. Why might the company be interested in developing different videos based on user score?

Answers may vary.

2. Thirty female users and twenty-five male users were selected at random from a database of people who play the game regularly. Each of them agreed to be part of a research study and report their scores. A leadership score is based on a player's answers to leadership questions. A score of 1 to 40 is considered a beginning level leadership score, a score of 41 to 60 is considered a middle level leadership score, and a score of greater than 60 is considered an advanced level leadership score.
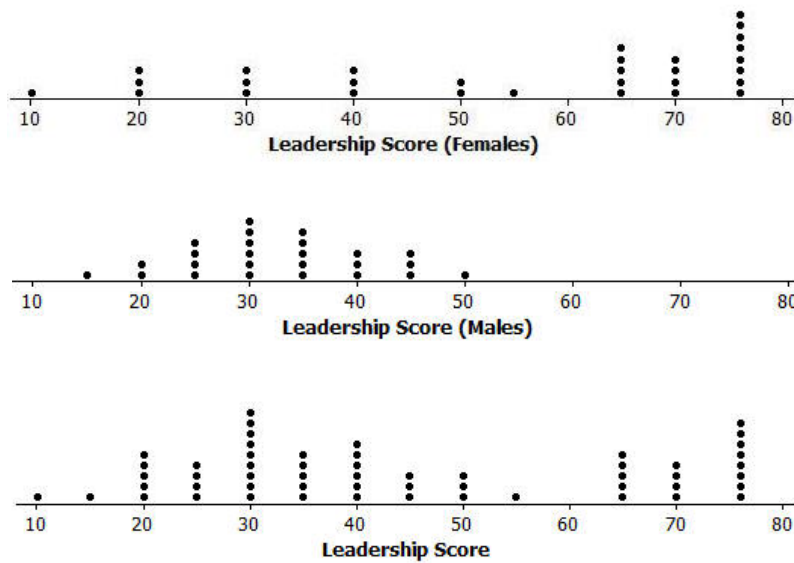
Use the following data to make a dot plot of the female scores, a dot plot of the male scores, and a dot plot of the scores for the combined group of males and females.

Female scores:

10	20	20	20	30	30	30	40	40	40
50	50	55	65	65	65	65	65	70	70
70	70	76	76	76	76	76	76	76	76

Male scores:

15	20	20	25	25	25	25	30	30	30
30	30	30	35	35	35	35	35	40	40
40	45	45	45	50					

3. **What do you think is a typical score for a female user? What do you think is a typical score for a male user? Explain how you determined these typical scores.**

    The mean score for the female users is approximately 54.9, and the median is score 65. The median is a better description of a typical score for females. Students may note that the mean score describes a female user at the middle level of leadership, and the median score describes a female user at the advanced level. The median is a better description of a typical score for females.

    The mean score for the male users is approximately 32.6, and the median score is 30. This has an approximate symmetrical distribution, and this is why the mean and the median are nearly equal. The mean is a good description of a typical score for males. Students may further indicate that this describes males at the beginning level of this game.

4. **Why is it more difficult to report a typical score for the overall group that includes both the males and females?**

    The overall distribution is difficult to summarize in terms of shape. It obviously combines the symmetrical shape from the male users and the skewed shape from the female users. The mean score of all users is approximately 44.8. The median score is 40. The mean and the median are not the same, but the values are relatively close. The mean is often interpreted as a balance point for this type of distribution. This interpretation of a mean is developed in the next lesson; however, it is also discussed in the Grade 6 standards and could be introduced again with this Problem Set.

5. **Production costs will only allow for two video advertisements to be developed. Which two videos would you recommend for development? Explain your recommendations.**

    In recommending which videos to produce based on the dot plots and measure of center, students will generally select video 4 (males with beginning level scores) and video 5 (females with advanced level scores).