

8.2

Which Measure Is Better?

Determining the Best Measure of Center for a Data Set

LEARNING GOALS

In this lesson, you will:

- Calculate and interpret the mean of a data set.
- Calculate and interpret the median of a data set.
- Estimate the mean and median of a data set from its data distribution.
- Determine which measure of central tendency (mean or median) is best to use for a data set.

KEY TERMS

- statistic
- measure of central tendency

You have probably been able to recite your ABCs since you started school. Now you may even be learning a new language that might use new letters. Some languages have different alphabets, where each letter represents sounds that are unique to that language even if the letters are the same as English. There are also some alphabets, such as the Russian alphabet or the Chinese alphabet, which use different letter symbols altogether.

Today, you will get the opportunity to learn new letters from another alphabet. The letters of the Greek alphabet are often used in mathematics to represent different mathematical ideas. You should already know the letter pi (π), which represents the ratio of the circumference of a circle to its diameter. By the time you finish this chapter you will know at least two more Greek letters! Keep an eye out for them as you work through the lessons.

8

PROBLEM 1 How Sweet It Is



Previously you analyzed a data set by creating a graphical representation of the data. However, you can also analyze a data set by describing numerical characteristics, or **statistics**, of the data. A statistic that describes the "center" of a data set is called a *measure of central tendency*. A **measure of central tendency** is the numerical values used to describe the overall clustering of data in a set. Two measures of central tendency that are typically used to describe a set of data are the mean and the median.

The arithmetic mean, or mean, represents the sum of the data values divided by the number of values. A common notation for the mean is \bar{x} , which is read "x bar."



The E-like symbol is actually the Greek letter sigma and in mathematical terms it means the "summation" or "sum of."

The formula shown represents the mean of a data set.

$$\text{mean} \rightarrow \bar{x} = \frac{\sum x}{n}$$

the sum of the data values

the number of data values

The mean of the data set 5, 10, 9, 7, 5 can be written using this formula.

$$\bar{x} = \frac{5 + 10 + 9 + 7 + 5}{5}$$

$$\bar{x} = 7.2$$

The mean of this data set is 7.2.

Why don't I write the sigma when writing the data values in the formula?

Recall that Lesson 8.1 *How Much Sugar Is Too Much?*, Matthew collected data on the sugar amount in one serving of various breakfast cereals. The data collected is shown.

8

Cereal Name	Sugar Amount in One Serving (grams)	Cereal Name	Sugar Amount in One Serving (grams)
Cocoa Rounds	13	Multi-Grain Squares	7
Flakes of Corn	4	All Branola	5
Frosty Flakes	11	Munch Crunch	12
Grape Nuggets	7	Branola Flakes	5
Golden Nuggets	10	Complete Flakes	4
Honey Nut Squares	10	Corn Crisps	3
Raisin Branola	7	Rice Crisps	4
Healthy Living Flakes	7	Shredded Wheatleys	1
Wheatleys	8	Puffs	22
Healthy Living Crunch	6	Fruit Circles	11

1. Represent the sugar amount in different cereals using the formula for the mean. Then determine the mean of the data set.

8



You can use your graphing calculator to determine the mean of a data set.

Step 1: Press **STAT** then press **ENTER** to select **1:Edit**.

Step 2: Enter the data values for the data set in List 1.

Step 3: Press **STAT** and scroll over to highlight **CALC**. Press **ENTER** to select **1:1-Var Stats**. Press **ENTER** again.

Step 4: The calculator should now show many values relating to the data set. You can scroll down for more values including the five number summary.

Be sure to check that your lists are clear of old data. Delete any data that might be in your lists before entering new data.

Do the values need to be entered in order?



2. Enter the data set for the sugar amount in various breakfast cereals into a graphing calculator. Then for each given symbol, state what it represents and its calculated value.

a. \bar{x}

b. Σx

c. n



3. Compare your answers in Question 2 with the answers you wrote using the formula for determining the mean in Question 1. What do you notice?



4. Determine the median sugar amount in grams in one serving of cereal. Interpret the meaning in terms of this problem situation.

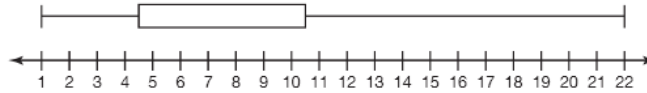
Does the order of the data matter when determining the median?



5. The box-and-whisker plot you constructed in the previous lesson is shown. Locate and label the mean and median values on the dot plot.

8

Sugar in Breakfast Cereals




6. Compare the mean and median. Which measure best represents the data set?

Constructing a box-and-whisker plot can take some time when using paper and pencil. Technology can make constructing a box-and-whisker plot more efficient.



You can use a graphing calculator to construct a box-and-whisker plot.

- Step 1:** Press **STAT** and then press **ENTER** to select **1:Edit**.
- Step 2:** Enter the data values of the data set in List 1.
- Step 3:** Press **2nd** and **STAT PLOT**, which is above the **Y=** button.
- Step 4:** Select **1:** and press **ENTER**. Then highlight **PLOT 1** and press **ENTER** to turn Plot 1 on. Then scroll down to **Type:** and select the box-and-whisker icon.  Press **ENTER**.
- Step 5:** Make sure the **XList** is using the correct list. Then press **GRAPH**.

7. Let's consider the data set without the value of 22.
- Remove the value of 22 from the data set. Use your graphing calculator to create a box-and-whisker plot for the new data set.
 - Plot above the given box-and-whisker plot your new box-and-whisker plot on the same graph in Question 5.
 - How does the removal of the value 22 affect the distribution of the data set?

8



- d. Did the mean and median change with the removal of the value 22? Does your choice for the best measure of center from Question 6 still hold true?

PROBLEM 2 Does Height Really Matter?



The Mountain View High School basketball team has its first game of the season on their home court. Coach Maynard doesn't know much about the visiting team, but he does have a list of the heights of their top ten players. Coach Maynard wants to compare the heights of his top ten players to those on the visiting team.

Home Team Heights (inches)	Visiting Team Heights (inches)
69	68
70	68
67	68
68	69
66	69
65	67
70	72
70	71
71	66
71	67

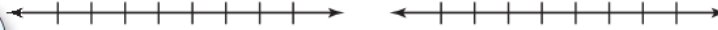
© 2012 Carnegie Learning



1. Represent the data for each team on a dot plot.

8

Don't forget
to label each dot
plot!



2. Analyze each dot plot you created.

a. Describe the data distribution of each graph and explain what it means in terms of the players' heights on each team.

b. Based on the dot plots, predict whether the mean or median will be greater for each data set. Explain your reasoning.

c. Verify your prediction by calculating the mean and median heights for each team. Was your prediction correct?

d. Which measure of central tendency best describes each data set? Explain your reasoning.



3. Describe the relationship that seems to exist between the data distribution and the values of the mean and median.

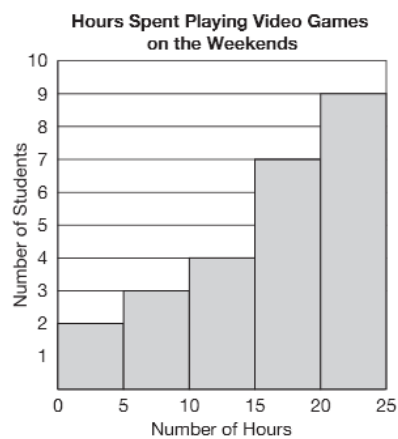
8



When the distribution of data is approximately symmetric, the mean is generally the more appropriate measure of center to use. When the distribution of data is skewed left or skewed right, the median is the more appropriate measure of center to use. The reason why the mean is more appropriate in a symmetric data distribution is due to the fact that most data points are close to the mean. There are not many if any data values that are much greater or lesser than the mean. In a skewed left or right distribution, most data values are closer to the median with few data points being much greater or lesser than the median. Therefore, the median is not affected by these values.



4. The histogram from Lesson 8.1 *Weekend Gamers* is shown.



- a. Predict whether the mean or median number of hours spent playing video games will be greater. Explain your reasoning.

Remember to use characteristics of the graph to explain your reasoning.



- b. Suppose the two measures of central tendency for the given histogram are 16.1 hours and 17.5 hours. Which value is the mean and which value is the median? Explain your reasoning.



© 2012 Carnegie Learning

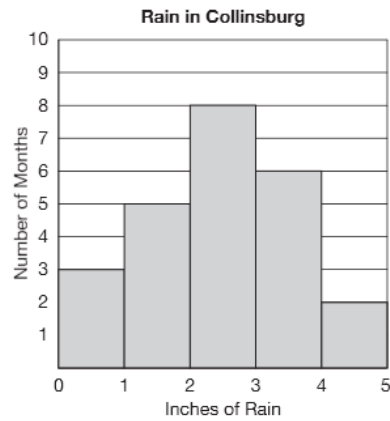
Talk the Talk

8

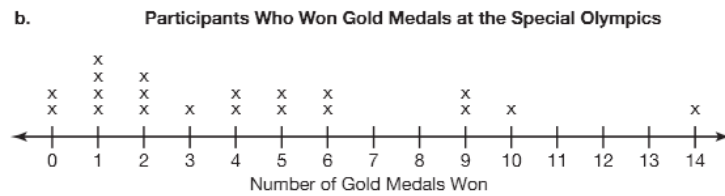


1. Identify which measure of central tendency would be most appropriate to describe the for each given graph. Then determine the mean and median if possible. If it is not possible, explain why not.

a.

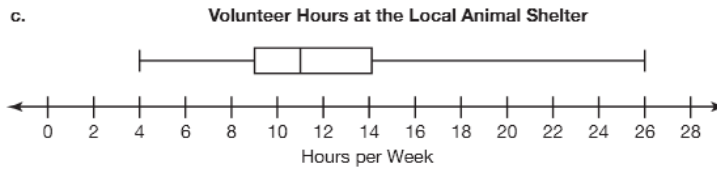


b.

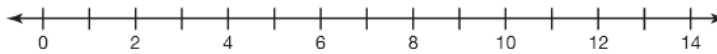


© 2012 Carnegie Learning

8



2. Consider the data set 0, 10, 10, 12, 14.
- a. Construct and label a dot plot of the data.



- b. Calculate the mean and median. Which measure do you think best represents the data set?
- c. Remove the value of 0 from the data set. How does this affect the distribution of the data set?
- d. Recalculate the mean and median without the value of 0. Does your choice in part (b) for the best measure of central tendency still hold true? Explain why or why not.

Why do you think the value 0 was selected to be removed from the data set?



Be prepared to share your solutions and methods.

© 2012 Carnegie Learning