# Analyzing and Comparing Data

Q

## **ESSENTIAL QUESTION**

How can you use solve real-world problems by analyzing and comparing data?



LESSON 10.1 Comparing Data Displayed in Dot Plots

**CACC** 7.SP.3, 7.SP.4

LESSON 10.2

Comparing Data Displayed in Box Plots

CACC 7.SP.3, 7.SP.4

#### LESSON 10.3

Using Statistical Measures to Compare Populations

🥋 CACC 7.SP.3, 7.SP.4

### **Real-World Video**

Scientists place radio frequency tags on some animals within a population of that species. Then they track data, such as migration patterns, about the animals.





my.hrw.com

Go digital with your write-in student edition, accessible on any device.



my.hrw.com

Math On the Spot

Scan with your smart phone to jump directly to the online edition, video tutor, and more.



Animated Math

Interactively explore key concepts to see how math works.



#### **Personal Math Trainer**

Get immediate feedback and help as you work through practice sets.

# Are Ready

Complete these exercises to review skills you will need for this module.



#### Personal **Math Trainer** Online Practice and Help

# Fractions, Decimals, and Percents

EXAMPLE	Write $\frac{13}{20}$ as a	0.65
	decimal and a	20)13.00
	percent.	<u>-120</u>
		1 00
		<u>-1 00</u>
		0
		0.65 = 65%

Write the fraction as a division problem. Write a decimal point and zeros in the dividend.

Place a decimal point in the quotient.

Write the decimal as a percent.

### Write each fraction as a decimal and a percent.



# Find the Median and Mode

EXAMPLE	17, 14, 13, 16, 13, 11 11, 13, 13, 14, 16, 17
	$median = \frac{13 + 14}{2}$
	= 13.5
	mode = 13

Order the data from least to greatest.

The median is the middle item or the average of the two middle items.

The mode is the item that appears most frequently in the data

#### Find the median and the mode of the data.

**5.** 11, 17, 7, 6, 7, 4, 15, 9 **6.** 43, 37, 49, 51, 56, 40, 44, 50, 36 **6.** 

# **Find the Mean**

EXAMPLE 17, 14, 13, 16, 13, 11  $mean = \frac{17 + 14 + 13 + 16 + 13 + 11}{6}$  items divided by the number of items.  $=\frac{84}{6}$ = 14

The mean is the sum of the data

#### Find the mean of the data.

**7.** 9, 16, 13, 14, 10, 16, 17, 9 **8.** 108, 95, 104, 96, 97,106, 94 **8.** 

# **Reading** Start-Up

## **Visualize Vocabulary**

Use the 🖌 words to complete the right column of the chart.

Statistical Data							
Definition	Example	<b>Review Word</b>					
A group of facts.	Grades on history exams: 85, 85, 90, 92, 94						
The middle value of a data set.	85, 85, <mark>90</mark> , 92, 94						
A value that summarizes a set of values, found through addition and division.	Results of the survey show that students typically spend 5 hours a week studying.						

# **Understand Vocabulary**

Complete each sentence using the preview words.

1. A display that uses values from a data set to show how the

values are spread out is a \_\_\_\_\_

**2.** A \_\_\_\_\_\_ uses a number line to display data.

# Vocabulary

### **Review Words**

- data (datos) interquartile range (rango entre cuartiles)
- mean (media)
   measure of center (medida central)
  - measure of spread (medida de dispersión)
- median (mediana) survey (encuesta)

### **Preview Words**

box plot (diagrama de caja) dot plot (diagrama de puntos) mean absolute deviation (MAD) (desviación absoluta media, (DAM))

# **Active Reading**

Layered Book Before beginning the module, create a layered book to help you learn the concepts in this module. Label the first flap with the module title. Label the remaining flaps with the lesson titles. As you study each lesson, write important ideas, such as vocabulary and formulas, under the appropriate flap. Refer to your finished layered book as you work on exercises from this module.





# GETTING READY FOR Analyzing and Comparing Data

Understanding the standards and the vocabulary terms in the standards will help you know exactly what you are expected to learn in this module.

### 🧖 CA CC 7.SP.3

Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

### **Key Vocabulary**

# **measure of center** (medida de centro)

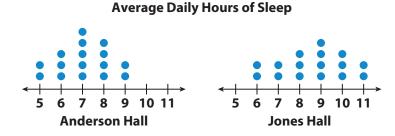
A measure used to describe the middle of a data set; the mean and median are measures of center.

# What It Means to You

You will compare two populations based on random samples.

### EXAMPLE 7.SP.3

Melinda surveys a random sample of 16 students from two college dorms to find the average number of hours of sleep they get. Use the results shown in the dot plots to compare the two populations.



Students in Jones Hall tend to sleep more than students in Anderson Hall, but the variation in the data sets is similar.

### **CACC** 7.SP.3

Informally assess... distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

### **Key Vocabulary**

# **measure of spread** (medida de la dispersión)

A measure used to describe how much a data set varies; the range, IQR, and mean absolute deviation are measures of spread.



Visit my.hrw.com to see all CA Common Core Standards explained.

🙆 my.hrw.com

# What It Means to You

You will compare two groups of data by comparing the difference in the means to the variability.

### EXAMPLE 7.SP.3

The tables show the number of items that students in a class answered correctly on two different math tests. How does the difference in the means of the data sets compare to the variability?

Items	<b>Correct o</b>	n Test 1
-------	------------------	----------

20, 13, 18, 19, 15, 18, 20, 20, 15, 15, 19, 18

Mean: 17.5; Mean absolute deviation: 2

#### **Items Correct on Test 2**

8, 12, 12, 8, 15, 16, 14, 12, 13, 9, 14, 11

Mean: 12; Mean absolute deviation: 2

The means of the two data sets differ by  $\frac{17.5-12}{2} = 2.75$  times the variability of the data sets.

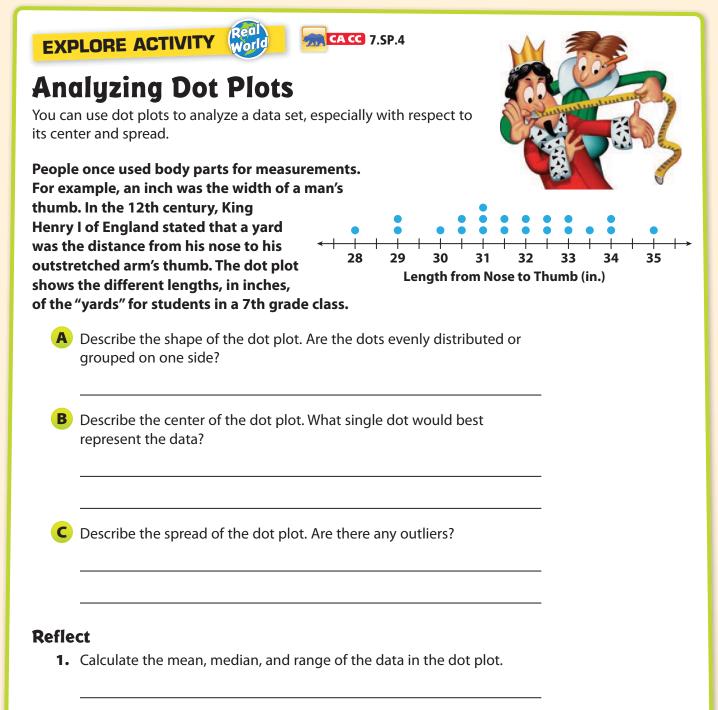
# Comparing Data Displayed in Dot Plots

Use measures of center and measures of variability for

measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *Also 7.SP.3* 

### **ESSENTIAL QUESTION**

How do you compare two sets of data displayed in dot plots?





<u>Math</u> Talk

How do the heights of field hockey players compare with

the heights of softball and

basketball players?

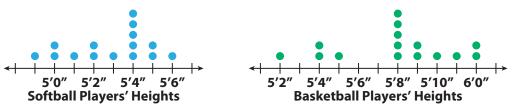
**Mathematical Practices** 

# **Comparing Dot Plots Visually**

You can compare dot plots visually using various characteristics, such as center, spread, and shape.

# EXAMPLE 1 Real Store 7.SP.3

The dot plots show the heights of 15 high school basketball players and the heights of 15 high school softball players.



### Visually compare the shapes of the dot plots.

Softball: All the data is 5'6" or less. Basketball: Most of the data is 5'8" or greater. As a group, the softball players are shorter than the basketball players.

### **B** Visually compare the centers of the dot plots.

Softball: The data is centered around 5'4". Basketball: The data is centered around 5'8". This means that the most common height for the softball players is 5 feet 4 inches, and for the basketball players 5 feet 8 inches.

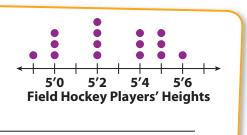
### Visually compare the spreads of the dot plots.

Softball: The spread is from 4'11" to 5'6". Basketball: The spread is from 5'2" to 6'0". There is a greater spread in heights for the basketball players.

## YOUR TURN

С

**2.** Visually compare the dot plot of heights of field hockey players to the dot plots for softball and basketball players.



Shape: \_\_\_\_\_

Center: \_\_\_\_\_

Spread: \_\_\_\_\_

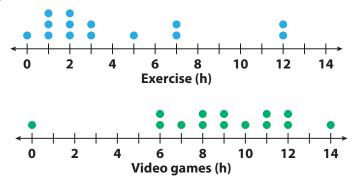


# **Comparing Dot Plots Numerically**

You can also compare the shape, center, and spread of two dot plots numerically by calculating values related to the center and spread. Remember that outliers can affect your calculations.

#### Real **EXAMPLE 2** World

Numerically compare the dot plots of the number of hours a class of students exercises each week to the number of hours the students play video games each week.



### A Compare the shapes of the dot plots.

The dot plots appear almost opposite. The dot plots show that most students exercise less than 4 hours but most play video games more than 6 hours each week.

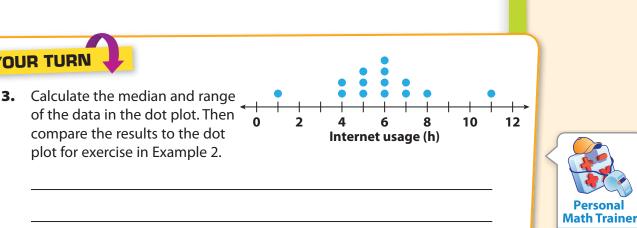
### **B** Compare the centers of the dot plots by finding the medians.

The median number of hours that students exercise is 2.5 hours, which is 6.5 hours less than the median of 9 hours that students play video games.

### C Compare the spreads of the dot plots by calculating the ranges.

The dot plots show the ranges to be similar to one another. The range for the amount of time that students exercise is 12 hours, and the range for the amount of time that students play video games is 14 hours.

**YOUR TURN** 



# **CA CC** 7.SP.4





**Math Talk Mathematical Practices** 

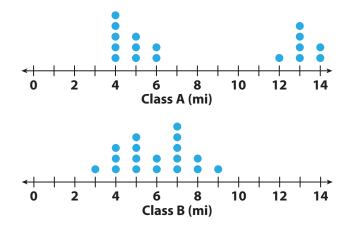
How do outliers affect the results of this data?

Lesson 10.1 313

**Online Practice** and Help my.hrw.com

### **Guided Practice**

The dot plots show the number of miles run per week for two different classes. For 1–5, use the dot plots shown. (Explore Activity, Example 1 and 2)



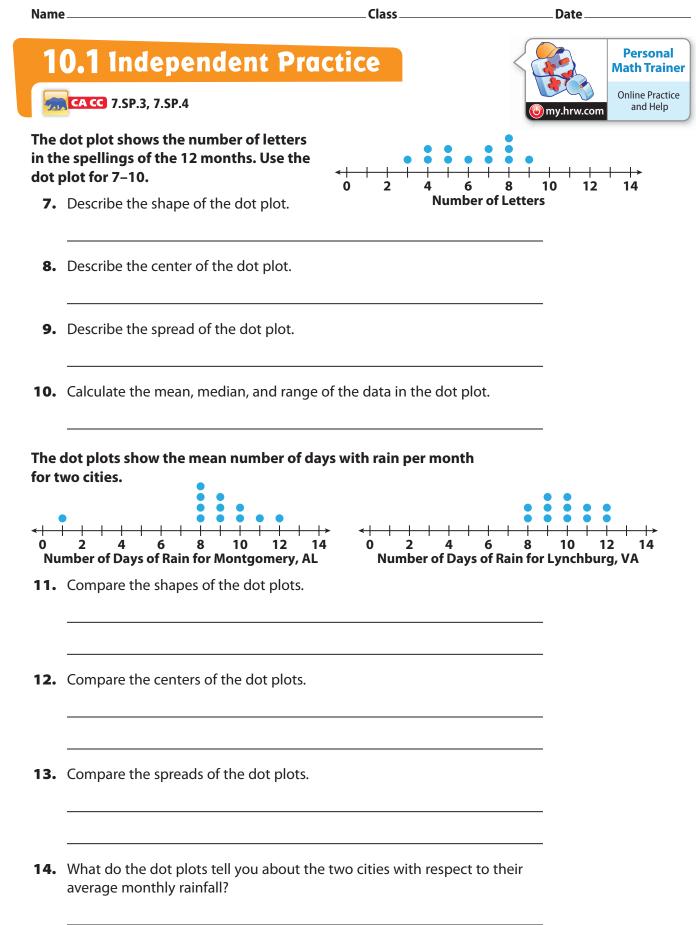
1. Compare the shapes of the dot plots.

2. Compare the centers of the dot plots.

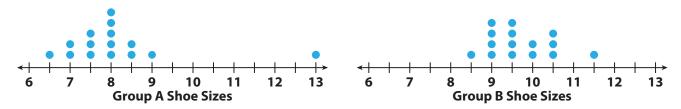
- **3.** Compare the spreads of the dot plots.
- 4. Calculate the medians of the dot plots.
- **5.** Calculate the ranges of the dot plots.

### **ESSENTIAL QUESTION CHECK-IN**

6. What do the medians and ranges of two dot plots tell you about the data?



The dot plots show the shoe sizes of two different groups of people.



- **15.** Compare the shapes of the dot plots.
- **16.** Compare the medians of the dot plots.
- **17.** Compare the ranges of the dot plots (with and without the outliers).
- **18.** Make A Conjecture Provide a possible explanation for the results of the dot plots.

### FOCUS ON HIGHER ORDER THINKING

**19. Analyze Relationships** Can two dot plots have the same median and range but have completely different shapes? Justify your answer using examples.

**20.** Draw Conclusions What value is most affected by an outlier, the median

or the range? Explain. Can you see these effects in a dot plot?

Work Area

# LESSON Comparing Data 10.2 Displayed in Box Plots

Real

Norld

**CACC** 7.SP.3

Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. Also 7.SP.4

**ESSENTIAL QUESTION** 

How do you compare two sets of data displayed in box plots?

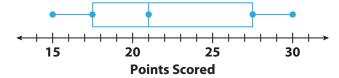
# EXPLORE ACTIVITY



# **Analyzing Box Plots**

Box plots show five key values to represent a set of data, the least and greatest values, the lower and upper quartile, and the median. To create a box plot, arrange the data in order, and divide them into four equal-size parts or guarters. Then draw the box and the whiskers as shown.

The number of points a high school basketball player scored during the games he played this season are organized in the box plot shown.





A F	ir	

nd the least and greatest values.

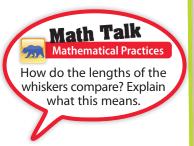
Least value: \_\_\_\_\_ Greatest value: \_\_\_\_\_

Find the median and describe what it means for the data.

**C** Find and describe the lower and upper quartiles.

**D** The interguartile range is the difference between the lower and upper guartiles, which is represented by the length of the box. Find the interquartile range.

 $Q_3 - Q_1 =$ \_\_\_\_\_ = \_\_\_\_



### Reflect

1. Why is one-half of the box wider than the other half of the box?



# **Box Plots with Similar Variability**

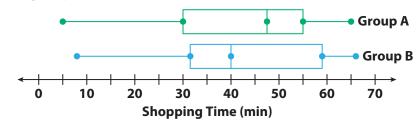
You can compare two box plots numerically according to their centers, or medians, and their spreads, or variability. Range and interquartile range (IQR) are both measures of spread. Box plots with similar variability should have similar boxes and whiskers.

### EXAMPLE 1

CACC 7.SP.3

### My Notes

The box plots show the distribution of times spent shopping by two different groups.



- A Compare the shapes of the box plots. The positions and lengths of the boxes and whiskers appear to be very similar. In both plots, the right whisker is shorter than the left whisker.
- **B** Compare the centers of the box plots.

Group A's median, 47.5, is greater than Group B's, 40. This means that the median shopping time for Group A is 7.5 minutes more.

C Compare the spreads of the box plots.

The box shows the interquartile range. The boxes are similar.

Group A: 55 - 30 = 25 min Group B: About 59 - 32 = 26 min

The whiskers have similar lengths, with Group A's slightly shorter than Group B's.

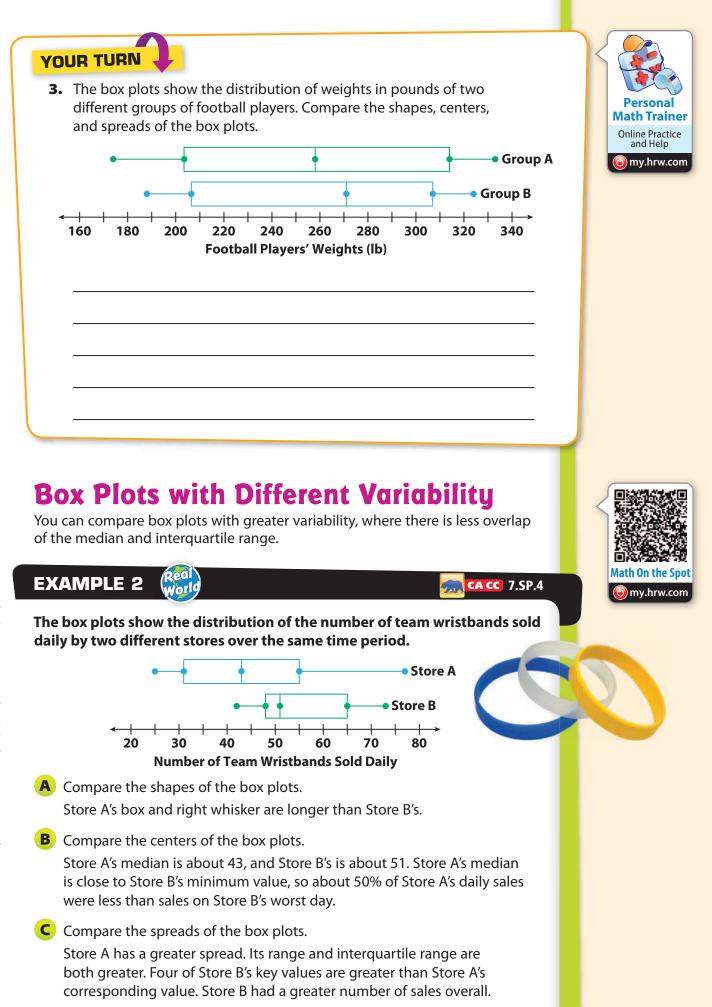
### Reflect

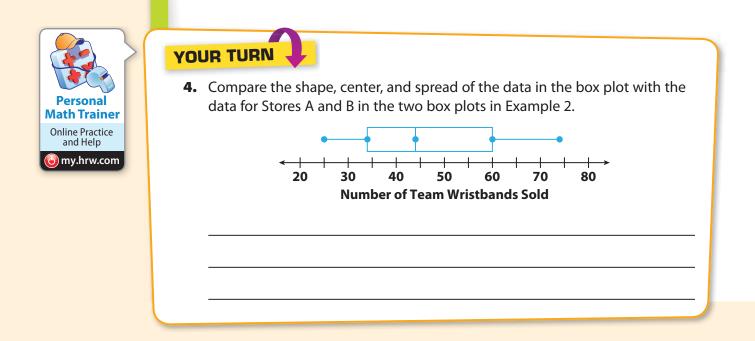
2. Which group has the greater variability in the bottom 50% of shopping times? The top 50% of shopping times? Explain how you know.

Math Talk Mathematical Practices

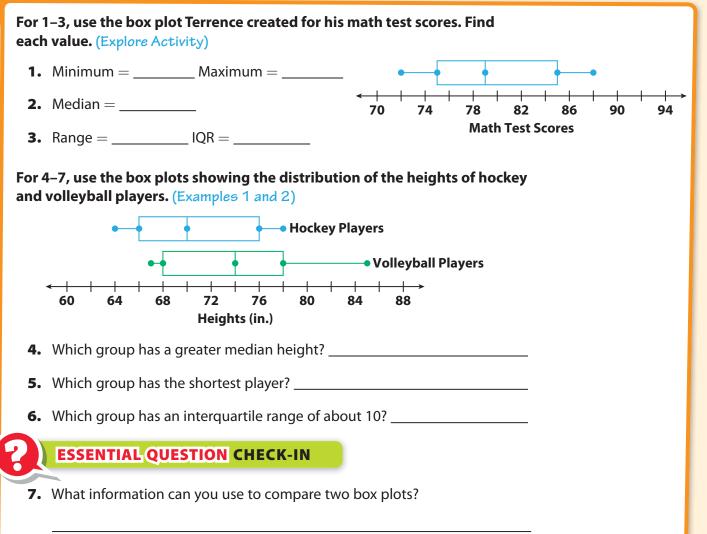
Which store has the shopper

who shops longest? Explain how you know.





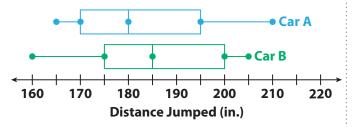
### **Guided Practice**



# **10.2** Independent Practice

**CACC** 7.SP.3, 7.SP.4

For 8–11, use the box plots of the distances traveled by two toy cars that were jumped from a ramp.



8. Compare the minimum, maximum, and median of the box plots.

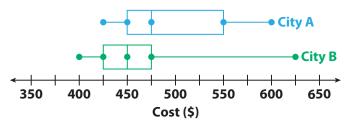
9. Compare the ranges and interguartile ranges of the data in box plots.

**10.** What do the box plots tell you about the

jump distances of two cars?

- **Online Practice** and Help my.hrw.com
- **11.** Critical Thinking What do the whiskers tell you about the two data sets?





**12.** In which city could you spend the least amount of money to lease a car? The greatest?

- **13.** Which city has a higher median price? How much higher is it?
- 14. Make a Conjecture In which city is it more likely to choose a car at random that leases for less than \$450? Why?

Personal **Math Trainer** 

Class.

Date\_

15.	Summarize Look back at the box plots for 12–14 on the previous page. What do the box plots tell you about the costs of leasing cars in those two cities?	
H.	OTT FOCUS ON HIGHER ORDER THINKING	Work Area
16.	<b>Draw Conclusions</b> Two box plots have the same median and equally long whiskers. If one box plot has a longer box than the other box plot, what does this tell you about the difference between the data sets?	
17.	<b>Communicate Mathematical Ideas</b> What can you learn about a data set from a box plot? How is this information different from a dot plot?	
		tt Publishing Company
18.	<b>Analyze Relationships</b> In mathematics, <i>central tendency</i> is the tendency of data values to cluster around some central value. What does a measure of variability tell you about the central tendency of a set of data? Explain.	© Houghton Mifflin Harcourt Publishing Company

322 Unit 5

# LESSON 10.3 Using Statistical Measures to Compare Populations

Informally assess the degree of visual overlap of two numerical data distributions

of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *Also* 7.SP.4

**ESSENTIAL QUESTION** 

How can you use statistical measures to compare populations?

**CACC** 7.SP.3

# **Comparing Differences in Centers** to Variability

Recall that to find the mean absolute deviation (MAD) of a data set, first find the mean of the data. Next, take the absolute value of the difference between the mean and each data point. Finally, find the mean of those absolute values.

#### 

The tables show the number of minutes per day students in a class spend exercising and playing video games. What is the difference of the means as a multiple of the mean absolute deviations?

Minutes Per Day Exercising

0, 7, 7, 18, 20, 38, 33, 24, 22, 18, 11, 6

 Minutes Per Day Playing Video Games

 13, 18, 19, 30, 32, 46, 50, 34, 36, 30, 23, 19

STEP 1

Calculate the mean number of minutes per day exercising.

0 + 7 + 7 + 18 + 20 + 38 + 33 + 24 + 22 + 18 + 11 + 6 = 204

 $204 \div 12 = 17$  Divide the sum by the number of students.

**STEP 2** Calculate the mean absolute deviation for the number of minutes exercising.

|0-17| = 17 |7-17| = 10 |7-17| = 10 |18-17| = 1|20-17| = 3 |38-17| = 21 |33-17| = 16 |24-17| = 7

20 - 17 = 3 |30 - 17| = 21 |33 - 17| = 10 |24 - 17| = 7

|22-17| = 5 |18-17| = 1 |11-17| = 6 |6-17| = 11

Find the mean of the absolute values.

17 + 10 + 10 + 1 + 3 + 21 + 16 + 7 + 5 + 1 + 6 + 11 = 108

 $108 \div 12 = 9$  Divide the sum by the number of students.

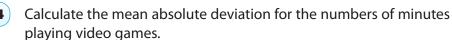




**STEP 3** Calculate the mean number of minutes per day playing video games. Round to the nearest tenth.

13 + 18 + 19 + 30 + 32 + 46 + 50 + 34 + 36 + 30 + 23 + 19 = 350

 $350 \div 12 \approx 29.2$  Divide the sum by the number of students.



13-29.2  = 16.2	18-29.2  = 11.2	19-29.2  = 10.2
30-29.2  = 0.8	32-29.2  = 2.8	46-29.2  = 16.8
50-29.2  = 20.8	34-29.2  = 4.8	36-29.2  = 6.8
30-29.2  = 0.8	23-29.2  = 6.2	19-29.2  = 10.2

Find the mean of the absolute values. Round to the nearest tenth.

16.2 + 11.2 + 10.2 + 0.8 + 2.8 + 16.8 + 20.8 + 4.8 + 6.8 + 0.8 + 6.2 + 10.2 = 107.6

 $107.6 \div 12 \approx 9$  Divide the sum by the number of students.

**STEP 5** Find the difference in the means.

29.2 - 17 = 12.2 Subtract the lesser mean from the greater mean.

**STEP 6** Write the difference of the means as a multiple of the mean absolute deviations, which are similar but not identical.

 $12.2 \div 9 \approx 1.36$  Divide the difference of the means by the MAD.

The means of the two data sets differ by about 1.4 times the variability of the two data sets.

## YOUR TURN

ò

1. The high jumps in inches of the students on two intramural track and field teams are shown below. What is the difference of the means as a multiple of the mean absolute deviations?

High Jumps for Students on Team 1 (in.)

44, 47, 67, 89, 55, 76, 85, 80, 87, 69, 47, 58

High Jumps for Students on Team 2 (in.)

40, 32, 52, 75, 65, 70, 72, 61, 54, 43, 29, 32





My Notes

# Using Multiple Samples to Compare Populations

Many different random samples are possible for any given population, and their measures of center can vary. Using multiple samples can give us an idea of how reliable any inferences or predictions we make are.



**CA CC** 7.SP.4

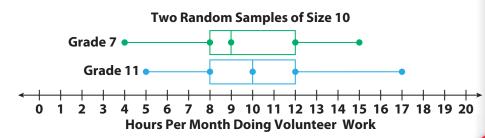
### EXAMPLE 2

STEP 1

**STEP 2** 

ò

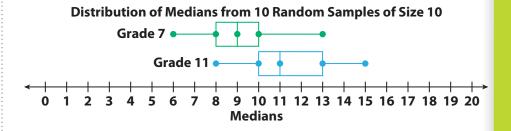
A group of about 250 students in grade 7 and about 250 students in grade 11 were asked, "How many hours per month do you volunteer?" Responses from one random sample of 10 students in grade 7 and one random sample of 10 students in grade 11 are summarized in the box plots.



How can we tell if the grade 11 students do more volunteer work than the grade 7 students?

The median is higher for the students in grade 11. But there is a great deal of variation. To make an inference for the entire population, it is helpful to consider how the medians vary among multiple samples.

The box plots below show how the medians from 10 different random samples for each group vary.



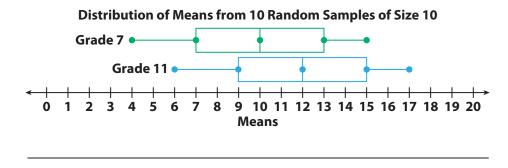
The medians vary less than the actual data. Half of the grade 7 medians are within 1 hour of 9. Half of the grade 11 medians are within 1 or 2 hours of 11. Although the distributions overlap, the middle halves of the data barely overlap. This is fairly convincing evidence that the grade 11 students volunteer more than the grade 7 students.





Why doesn't the first box plot establish that students in grade 11 volunteer more than students in grade 7? YOUR TURN

2. The box plots show the variation in the means for 10 different random samples for the groups in the example. Why do these data give less convincing evidence that the grade 11 students volunteer more?



## **Guided Practice**

**Math Trainer** 

Online Practice and Help

my.hrw.com

The tables show the numbers of miles run by the students in two classes. Use the tables in 1-2. (Example 1)

#### Miles Run by Class 1 Students

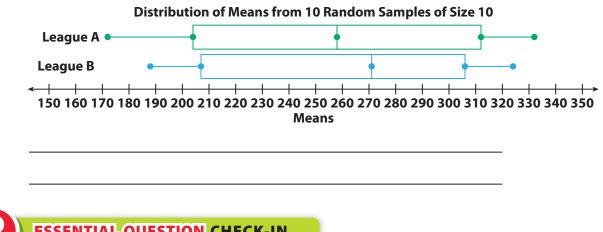
12, 1, 6, 10, 1, 2, 3, 10, 3, 8, 3, 9, 8, 6, 8

Miles Run by Class 2 Students 11, 14, 11, 13, 6, 7, 8, 6, 8, 13, 8, 15, 13, 17, 15

1. For each class, what is the mean? What is the mean absolute deviation (MAD)?



**3.** The box plots show the distributions of mean weights of 10 samples of 10 football players from each of two leagues, A and B. What can you say about any comparison of the weights of the two populations? Explain.



### **ESSENTIAL QUESTION CHECK-IN**

**4.** Why is it a good idea to use multiple random samples when making comparative inferences about two populations?

#### Class

Date

Personal

**Math Trainer** 

**Online Practice** 

# **10.3 Independent Practice**

**CACC** 7.SP.3, 7.SP.4

Josie recorded the average monthly temperatures for two cities in the state where she lives. Use the data for 5–7.

Average Monthly Temperatures for City 1 (°F)

23, 38, 39, 48, 55, 56, 71, 86, 57, 53, 43, 31

Average Monthly Temperatures for City 2 (°F)

8, 23, 24, 33, 40, 41, 56, 71, 42, 38, 28, 16



- **5.** For City 1, what is the mean of the average monthly temperatures? What is the mean absolute deviation of the average monthly temperatures?
- 6. What is the difference between each average monthly temperature for

City 1 and the corresponding temperature for City 2? \_\_\_\_

- 7. Draw Conclusions Based on your answers to Exercises 5 and 6, what do you think the mean of the average monthly temperatures for City 2 is? What do you think the mean absolute deviation of the average monthly temperatures for City 2 is? Give your answers without actually calculating the mean and the mean absolute deviation. Explain your reasoning.
- 8. What is the difference in the means as a multiple of the mean absolute

deviations? \_

**9.** Make a Conjecture Mark took 10 random samples of 10 students from two schools. He asked how many minutes they spend per day going to and from school. The tables show the medians and the means of the samples. Compare the travel times using distributions of the medians and means.

### School A

Medians: 28, 22, 25, 10, 40, 36, 30, 14, 20, 25 Means: 27, 24, 27, 15, 42, 36, 32, 18, 22, 29

Medians: 22, 25, 20, 14, 20, 18, 21, 18, 26, 19 Means: 24, 30, 22, 15, 20, 17, 22, 15, 36, 27 **10.** Justify Reasoning Statistical measures are shown for the ages of middle school and high school teachers in two states.

**State A:** Mean age of middle school teachers = 38, mean age of high school teachers = 48, mean absolute deviation for both = 6

**State B:** Mean age of middle school teachers = 42, mean age of high school teachers = 50, mean absolute deviation for both = 4

In which state is the difference in ages between members of the two groups more significant? Support your answer.

**11. Analyze Relationships** The tables show the heights in inches of all the adult grandchildren of two sets of grandparents, the Smiths and the Thompsons. What is the difference in the medians as a multiple of the ranges?

Heights of the Smiths' Adult Grandchildren (in.)

64, 65, 68, 66, 65, 68, 69, 66, 70, 67

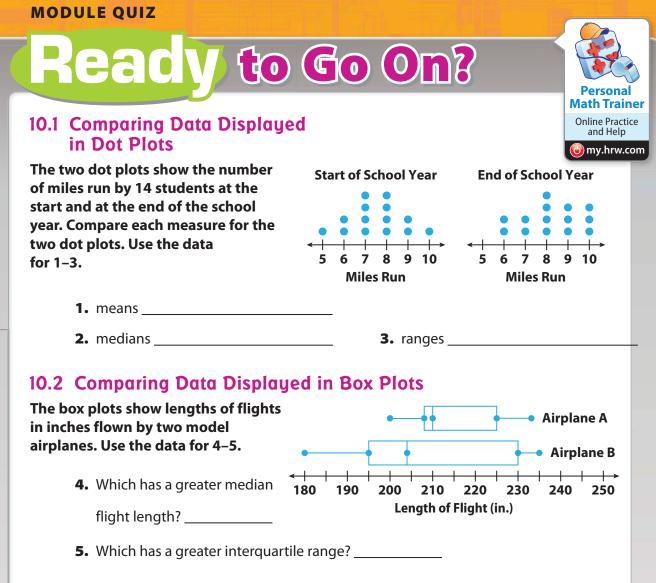
Heights of the Thompsons' Adult Grandchildren (in.)

75, 80, 78, 77, 79, 76, 75, 79, 77, 74



- **12. Critical Thinking** Jill took many samples of 10 tosses of a standard number cube. What might she reasonably expect the median of the medians of the samples to be? Why?
- **13. Analyze Relationships** Elly and Ramon are both conducting surveys to compare the average numbers of hours per month that men and women spend shopping. Elly plans to take many samples of size 10 from both populations and compare the distributions of both the medians and the means. Ramon will do the same, but will use a sample size of 100. Whose results will probably produce more reliable inferences? Explain.
- **14. Counterexamples** Seth believes that it is always possible to compare two populations of numerical values by finding the difference in the means of the populations as a multiple of the mean absolute deviations. Describe a situation that explains why Seth is incorrect.

Work Area



## **10.3 Using Statistical Measures to Compare Populations**

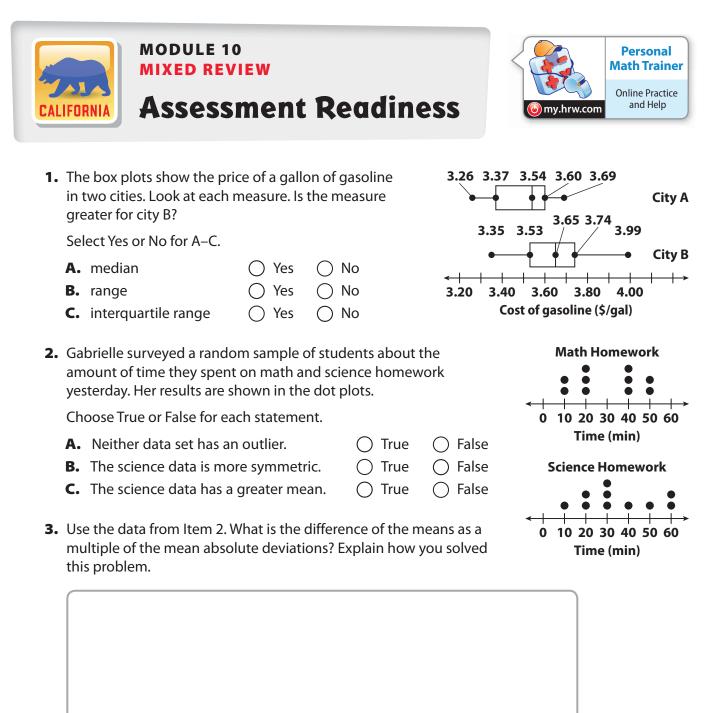
**6.** Roberta grows pea plants, some in shade and some in sun. She picks 8 plants of each type at random and records the heights.

Shade plant heights (in.)	7	11	11	12	9	12	8	10
Sun plant heights (in.)	21	24	19	19	22	23	24	24

Express the difference in the means as a multiple of their ranges.

## ESSENTIAL QUESTION

7. How can you use and compare data to solve real-world problems?



**4.** Tyrone buys a pair of shoes on sale for 20% off. The regular price of the shoes is \$59.95, and the sales tax rate is 7.5%. How much will Tyrone pay for the shoes, including sales tax? Explain how you solved this problem.

C Houghton Mifflin Harcourt Publishing Company