

8.3 Comparing Populations

Learning Target: Compare populations using measures of center and variation.

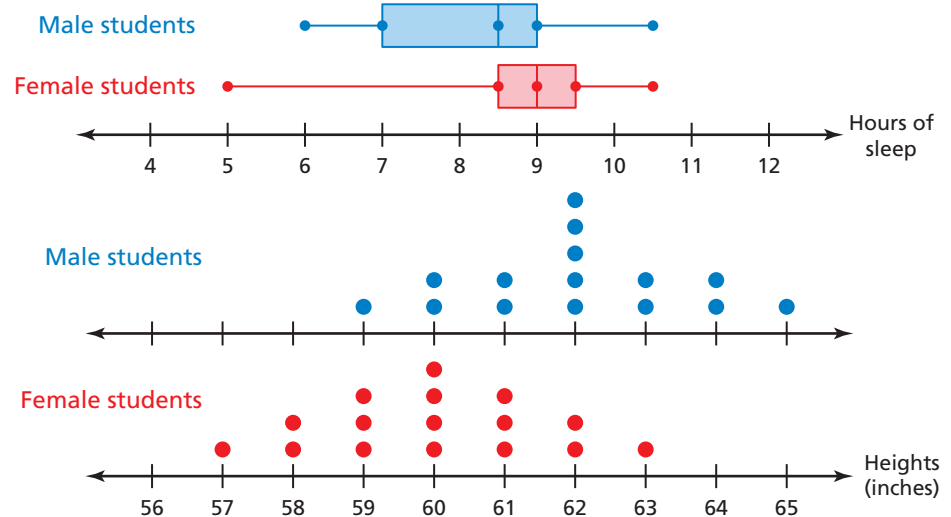
- Success Criteria:**
- I can find the measures of center and variation of a data set.
 - I can describe the visual overlap of two data distributions numerically.
 - I can determine whether there is a significant difference in the measures of center of two data sets.

EXPLORATION 1

Comparing Two Data Distributions

Work with a partner.

- a. Does each data display show *overlap*? Explain.



Math Practice

Recognize Usefulness of Tools

What are the advantages of each type of data display? Which do you prefer? Explain.

Ages of People in Two Exercise Classes

10:00 A.M. Class		8:00 P.M. Class
	1	8 9
	2	1 2 2 7 9 9
	3	0 3 4 5 7
9 7 3 2 2 2	4	0
7 5 4 3 1	5	
7 0 0	6	
0	7	

Key: 2 | 4 | 0 = 42 and 40 years

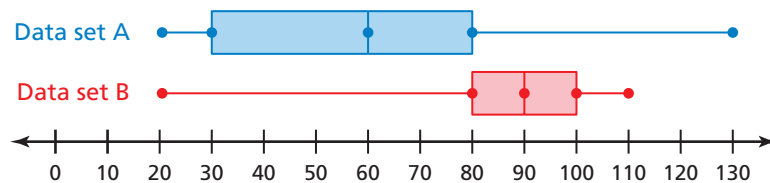
- b. How can you describe the overlap of two data distributions using words? How can you describe the overlap numerically?
- c. In which pair of data sets is the difference in the measures of center the most significant? Explain your reasoning.

8.3 Lesson

Use the mean and the mean absolute deviation (MAD) to compare two populations when both distributions are symmetric. Use the median and the interquartile range (IQR) when either one or both distributions are skewed.

EXAMPLE 1 Comparing Populations

Two data sets contain an equal number of values. The double box-and-whisker plot represents the values in the data sets.



- a. Compare the data sets using measures of center and variation.

Both distributions are skewed. Use the median and the IQR.

Data set A	Data set B
Median = 60	Median = 90
IQR = $80 - 30 = 50$	IQR = $100 - 80 = 20$

▶ So, Data set B has a greater measure of center, and Data set A has a greater measure of variation.

- b. Which data set is more likely to contain a value of 95?

About 25% of the data values in Data set A are between 80 and 130. About 50% of the data values in Data set B are between 80 and 100.

▶ So, Data set B is more likely to contain a value of 95.

- c. Which data set is more likely to contain a value that differs from the center by at least 30?

The IQR of Data set A is 50 and the IQR of Data set B is 20. This means it is more common for a value to differ from the center by 30 in Data set A than in Data set B.

▶ So, Data set A is more likely to contain a value that differs from the center by at least 30.

Try It

1. Which data set is more likely to contain a value of 70?
2. Which data set is more likely to contain a value that differs from the center by no more than 3?

Remember



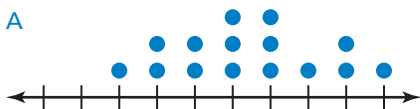
In a box-and-whisker plot, about $\frac{1}{4}$ of the data are in each whisker and about $\frac{1}{2}$ of the data are in the box.

Less visual overlap indicates a more significant difference in the measures of center.

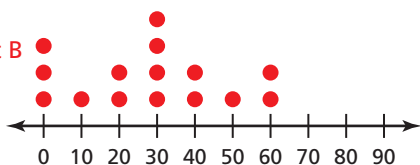
When two populations have similar variabilities, the visual overlap of the data can be described by writing the difference in the measures of center as a multiple of the measure of variation. Greater values indicate less visual overlap.

EXAMPLE 2 Describing Visual Overlap

Data set A



Data set B



The double dot plot shows two data sets. Express the difference in the measures of center as a multiple of the measure of variation.

Both distributions are approximately symmetric. Use the mean and the MAD to describe the centers and variations.

Data set A

$$\text{Mean} = \frac{810}{15} = 54$$

$$\text{MAD} = \frac{244}{15} \approx 16$$

$$\frac{\text{difference in means}}{\text{MAD}} = \frac{26}{16} \approx 1.6$$

Data set B

$$\text{Mean} = \frac{420}{15} = 28$$

$$\text{MAD} = \frac{236}{15} \approx 16$$

So, the difference in the means is about 1.6 times the MAD.

Try It

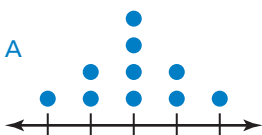
- WHAT IF?** Each value in the dot plot for Data set A increases by 30. How does this affect your answers? Explain.



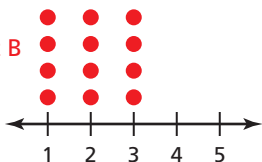
Self-Assessment for Concepts & Skills

Solve each exercise. Then rate your understanding of the success criteria in your journal.

Data set A



Data set B



- COMPARING POPULATIONS** The double dot plot shows two data sets. Compare the data sets using measures of center and variation. Then express the difference in the measures of center as a multiple of the measure of variation.
- WHICH ONE DOESN'T BELONG?** You want to compare two populations represented by skewed distributions. Which measure does *not* belong with the other three? Explain your reasoning.

Median of first data set

Median of second data set

IQR of first data set

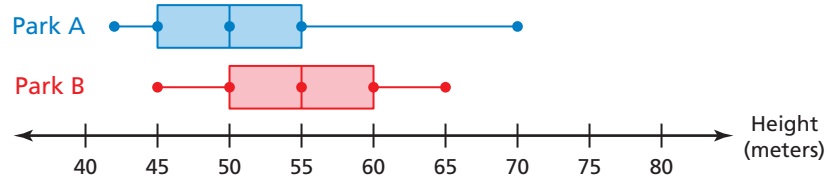
MAD of second data set

When the difference in the measures of center is at least 2 times the measure of variation, the difference is significant.

EXAMPLE 3 Modeling Real Life



The double box-and-whisker plot represents the heights of rollercoasters at two amusement parks. Are the rollercoasters significantly taller at one park than at the other park?



The distribution for Park A is skewed, so use the median and the IQR to describe the centers and variations.

Park A

Median = 50

IQR = 55 - 45 = 10

Park B

Median = 55

IQR = 60 - 50 = 10

Because the variabilities are similar, you can describe the visual overlap by expressing the difference in the medians as a multiple of the IQR.

$$\frac{\text{difference in medians}}{\text{IQR}} = \frac{5}{10} = 0.5$$

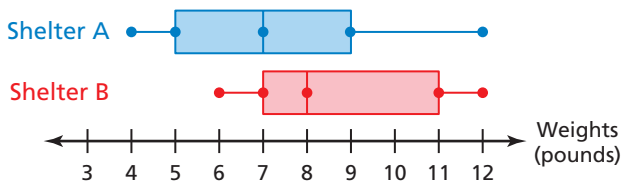
Because the quotient is less than 2, the difference in the medians is not significant.

▶ The rollercoasters are not significantly taller at one park than at the other park.



Self-Assessment for Problem Solving

Solve each exercise. Then rate your understanding of the success criteria in your journal.



6. The double box-and-whisker plot represents the weights of cats at two shelters. Are the cats significantly heavier at one shelter than at the other? Explain.

7. **DIG DEEPER!** Tornadoes in Region A travel significantly farther than tornadoes in Region B. The tornadoes in Region A travel a median of 10 miles. Create a double box-and-whisker plot that can represent the distances traveled by the tornadoes in the two regions.

8.3 Practice



Go to BigIdeasMath.com to get HELP with solving the exercises.

► Review & Refresh

Twenty percent of all seventh graders have watched a horse race. Explain whether the sample closely estimates the percentage of seventh graders who have watched a horse race.

- In a sample of 15 seventh graders, 4 have watched a horse race.
- In a sample of 10 seventh graders, 6 have watched a horse race.

Find the unit rate.

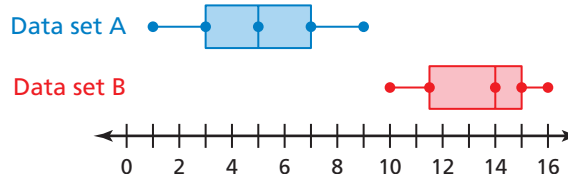
- 60 kilometers in 2 hours
- \$11.40 for 5 cans



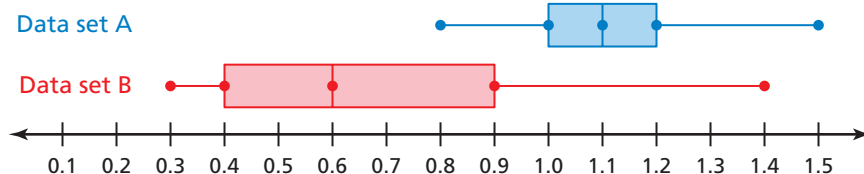
► Concepts, Skills, & Problem Solving

COMPARING TWO DATA DISTRIBUTIONS The double box-and-whisker plot represents the values in two data sets. (See Exploration 1, p. 337.)

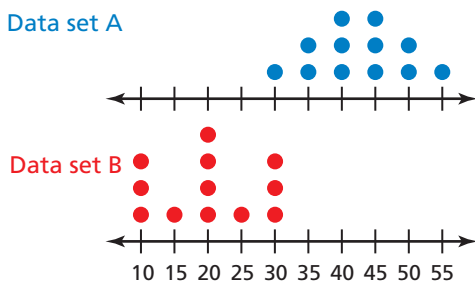
- Does the data display show *overlap*? Explain.
- Is there a significant difference in the measures of center for the pair of data sets? Explain.



COMPARING POPULATIONS Two data sets contain an equal number of values. The double box-and-whisker plot represents the values in the data sets.



- Compare the data sets using measures of center and variation.
- Which data set is more likely to contain a value of 1.1?
- Which data set is more likely to contain a value that differs from the center by 0.3?



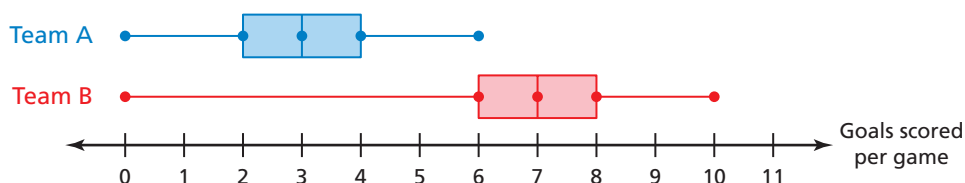
- DESCRIBING VISUAL OVERLAP** The double dot plot shows the values in two data sets. Express the difference in the measures of center as a multiple of the measure of variation.

11. **YOU BE THE TEACHER** The distributions of attendance at basketball games and volleyball games at your school are symmetric. Your friend makes a conclusion based on the calculations shown below. Is your friend correct? Explain your reasoning.

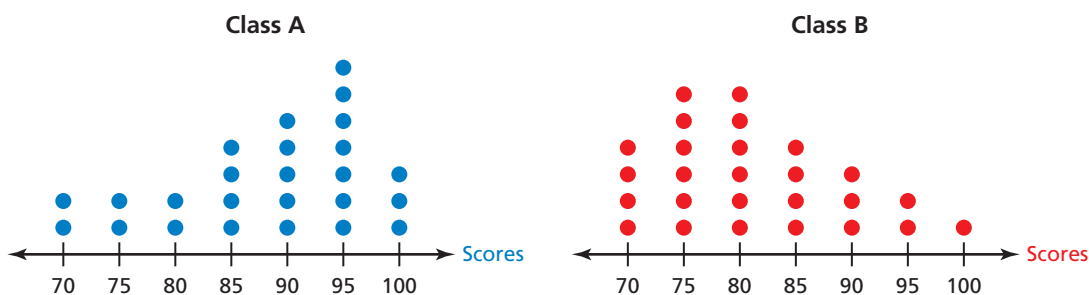
Volleyball Game Attendance: Mean = 80, MAD = 20
 Basketball Game Attendance: Mean = 160, MAD = 20
 The difference in means is four times the MAD,
 so attendance at basketball games is significantly
 greater than attendance at volleyball games.



12. **MODELING REAL LIFE** The double box-and-whisker plot represents the goals scored per game by two hockey teams during a 20-game season. Is the number of goals scored per game significantly greater for one team than the other? Explain.



13. **MODELING REAL LIFE** The dot plots show the test scores for two classes taught by the same teacher. Are the test scores significantly greater for one class than the other? Explain.



14. **MP PROBLEM SOLVING** A scientist experiments with mold colonies of equal area. She adds a treatment to half of the colonies. After a week, she measures the area of each colony. If the areas are significantly different, the scientist will repeat the experiment. The results are shown. Should the scientist repeat the experiment? Justify your answer.

