

Name: \_\_\_\_\_  
 Algebra 1

Date: \_\_\_\_\_  
 Band: \_\_\_\_\_

## SAT Practice: Data Analysis (Statistics)

### Frequency and Histograms

**LT#1:** Make and interpret frequency tables and histograms.

#### Warm Up

As part of an environmental science project, you measure the trees in a park to the nearest foot. Your data are shown at the right. You want to estimate the number of trees between 30 ft and 60 ft tall. Choose a method for estimating and state any assumptions you make. What is your estimate? Explain your reasoning.

Tree Height	
Height (ft)	Number of Trees
0-25	 
26-50	                  
51-75	 
76-100	

#### Vocabulary & Examples

**Frequency of an interval:**

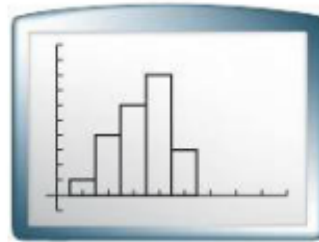
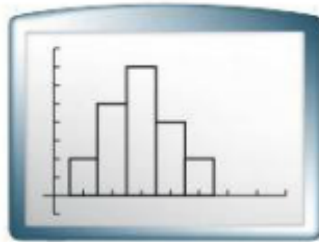
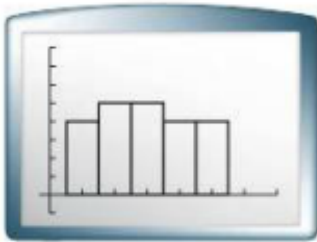
**Frequency table:**

#### Problem 1: Making a Frequency Table

The numbers of home runs by the batters in a local home run derby are listed below. What is a frequency table that represents the data?

7, 17, 14, 2, 7, 9, 5, 12, 3, 10, 4, 12, 7, 15

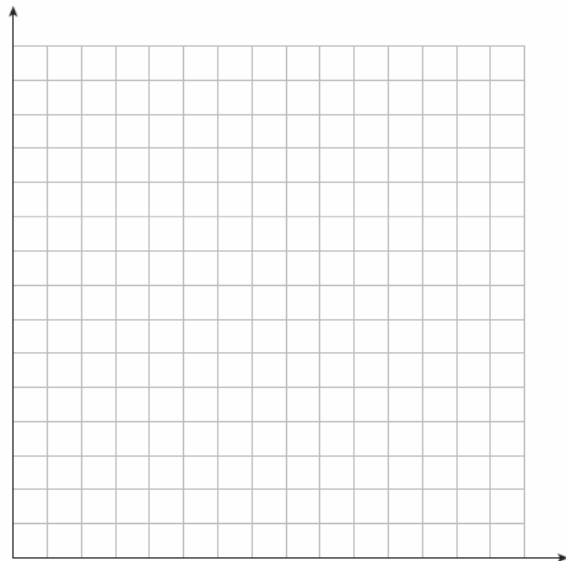
**Histogram:**



**Problem 2: Making a Histogram**

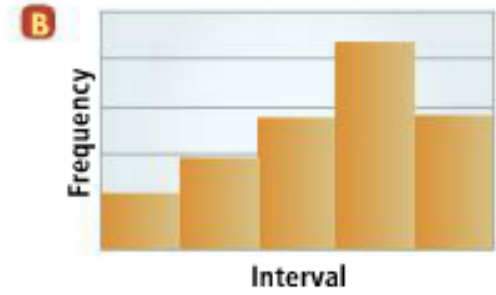
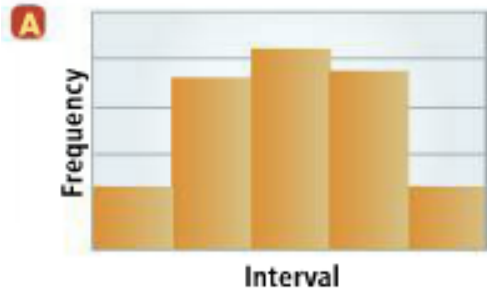
The data below are the numbers of hours per week a group of students spent watching television. What is a histogram that represents the data?

7, 10, 1, 5, 14, 22, 6, 8, 0, 11, 13, 3, 4, 14, 5



**Problem 3: Interpreting Histograms**

Is each histogram *uniform*, *symmetric*, or *skewed*?



**Cumulative frequency table:**

**Problem 4: Making a Cumulative Frequency Table**

The numbers of text messages sent on one day by different students are shown below. What is the cumulative frequency table that represents the data?

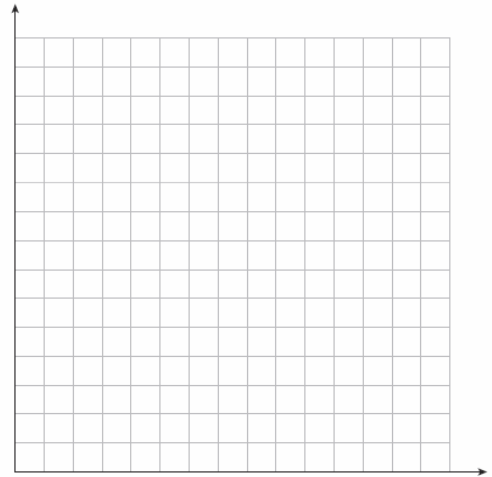
17, 3, 1, 30, 11, 7, 1, 5, 2, 39, 22, 13, 2, 0, 21, 1, 49, 41, 27, 2, 0

### Frequency & Histogram Practice

1. What is the frequency table for the data in Problem 1 that uses intervals of 5?

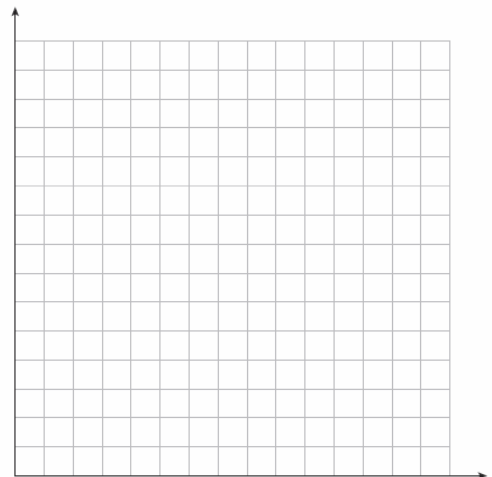
2. The finishing times, in seconds, for a race are shown below. What is a histogram that represents the data?

95, 105, 83, 80, 93, 98, 102, 99, 82, 89, 90, 82, 89



3. A. The following set of data shows the numbers of dollars Jay spent on lunch over the last two weeks. Make a histogram of the data. Is the histogram *uniform*, *symmetric*, or *skewed*?

17, 1, 4, 11, 14, 14, 5, 16, 6, 5, 9, 10, 13, 9



**B.** How much money should Jay play to bring for lunch next week? Explain your reasoning.

**4.** What is a cumulative frequency table that represents the data below?

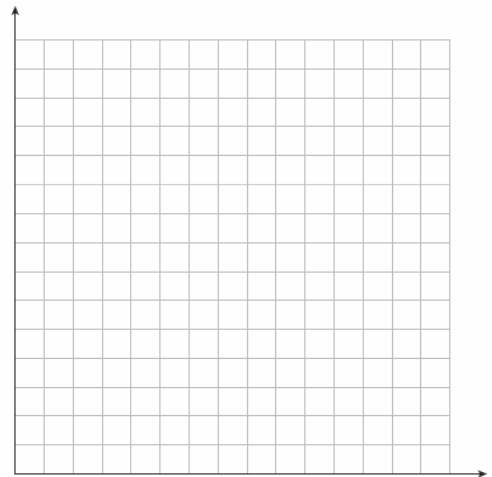
12, 13, 15, 1, 5, 7, 10, 9, 2, 2, 7, 11, 2, 1, 0, 15

**The data below show battery life, in hours, for different brands of batteries.**

**12, 9, 10, 14, 10, 11, 10, 18, 21, 10, 14, 22**

**5.** Make a frequency table of the data.

**6.** Make a histogram of the data.



**7.** Make a cumulative frequency table of the data.

8. How might a frequency table help a store owner determine the busiest business hours?
  
9. What is the different between a symmetric histogram and a skewed histogram?
  
10. How can you use a frequency table of a data set to construct a cumulative frequency table?

### **Measures of Central Tendency and Dispersion**

**LT#2:** Find mean, median, mode, and range.

**Warm Up**

Refer to the data at the right. Which basketball player would you rather have on your team? Justify your answer.

Game	Player 1		Player 2	
	Points	Assists	Points	Assists
1	12	8	23	5
2	10	4	10	4
3	15	4	12	2
4	11	5	25	3
5	12	3	5	2

**Vocabulary & Examples**

**Measures of central tendency:**

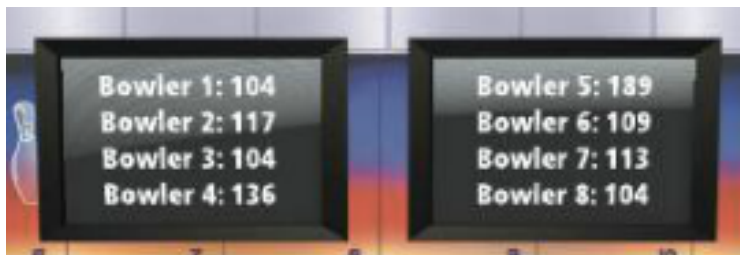
**Outlier:**

**Mean, Median, and Mode:**

Measure	When to Use
Mean	
Median	
Mode	

**Problem 1: Finding Measures of Central Tendency**

What are the mean, median, and mode of the bowling scores below? Which measure of central tendency best describes the scores?



**Problem 2: Finding a Data Value**

Your grades on three exams are 80, 93, and 91. What grade do you need on the next exam to have an average of 90 on the four exams?

**Problem 3: Finding the Range**

The closing prices, in dollars, of two stocks for the first five days in February are shown below. What are the range and mean of each set of data? Use the results to compare the data sets.

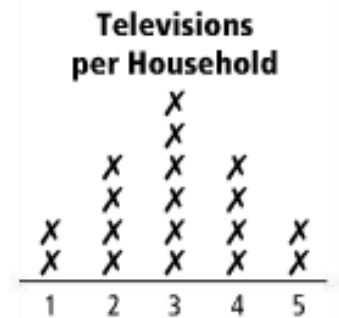
**Stock A: 25, 30, 30, 47, 28**

**Stock B: 34, 28, 31, 36, 31**

**Problem 4: Finding Measures of Central Tendency and Ranges**

The results of a survey on the number of televisions in students' households are shown in the line plot.

**A.** Calculate the mean, median, and range of the data.

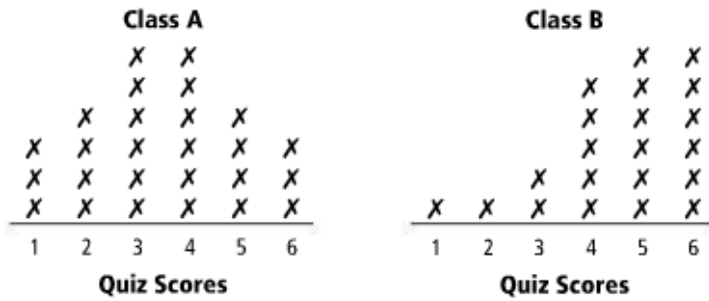


**B.** How can you tell from the graph that the mean and median are equal?



**Problem 5: Comparing Measures of Central Tendency**

The results from the same quiz given to two different classes are shown in the line plots.



A. Which class has a higher standard for being in the top half of the quiz scores?

B. By comparing line plots, how can you tell which mean is greater?

**Measures of Central Tendency and Dispersion Practice**

1. Consider the scores from Problem 1 that do not include the outlier, 189. What are the mean, median, and mode of the scores? Which measure of central tendency best describes the data?

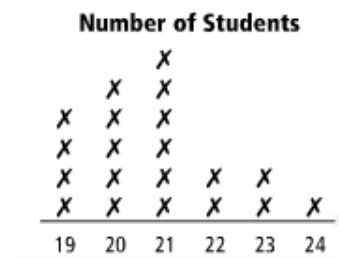
**2. A.** The grades in Problem 2 were 80, 93, and 91. What grade would you need on your next exam to have an average of 88 on the four exams?

**B.** If 100 is the highest possible score on the fourth exam, is it possible to raise your average of 92? Explain.

**3.** For the same days, the closing prices, in dollars, of Stock C were 7, 4, 3, 6, and 1. The closing prices, in dollars, of Stock D were 24, 15, 2, 10, and 5. What are the range and mean of each set of data? Use your results to compare Stock C with Stock D.

**4.** The line plot shows the number of students in each homeroom at Jefferson High School.

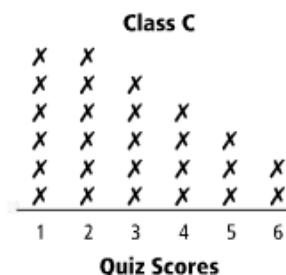
**A.** Find the mean, median, and range of the data.



**B.** How can you use the line plot to determine whether the mean and median are equal?

5. Results from a third class are shown here.

A. Is the mean from Class C greater than or less than the mean for Class A?



B. Is the median for Class C greater than or less than the median for Class A?

C. How can you tell which data set has a greater mean by comparing the graphs?

**Find the mean, median, and mode of each data set. Explain what measure best describes the data.**

6. 1, 29, 33, 31, 30, 33

7. 8.2, 9.3, 8.5, 8.8, 9.0

8. A student has gotten the following grades on his tests: 87, 95, 86, and 88. He needs to have an average of 90 to receive an A for the class. What is the minimum grade he must get on the last test in order to have an average of 90?

9. How do mean, median, and mode described the central tendency of a data set? Why are three different measures needed?

10. One student said 10 was the range of the data set 2, 10, 8, and 3. Another student said the range was 8. Which student is correct? Explain.

11. How is the range of a data set affected by an outlier?

### Box-and-Whisker Plots

**LT#3:** Make and interpret box-and-whisker plots.

**LT#4:** Find quartile and percentiles.

#### Warm Up

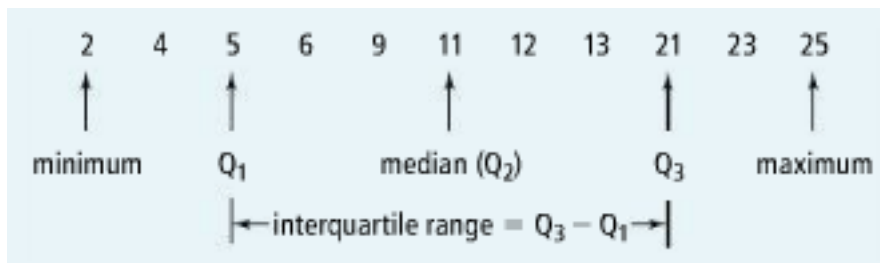
Suppose you are moving to a new town and are looking for a warm climate. You can choose between Morrell or Glenville. Based on the average monthly temperatures below, which town would you choose? How did you decide?

Average Monthly Temperatures												
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
Morrell	56	57	60	68	74	82	83	85	79	70	62	55
Glenville	58	62	66	70	76	78	81	84	77	73	68	63

**Vocabulary & Examples**

Quartiles:

Interquartile range:

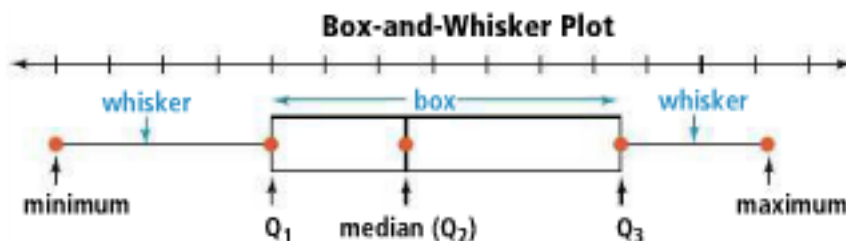


**Problem 1: Summarizing a Data Set**

What are the minimum, first quartile, median, third quartile, and maximum of the data set below?

125, 80, 140, 135, 126, 140, 350, 75

**Box-and-whisker plot:**



**Problem 2: Making a Box-and-Whisker Plot**

The table at the right shows the amount of crops harvested in the United States for a certain period. What box-and-whisker plot represents the data?

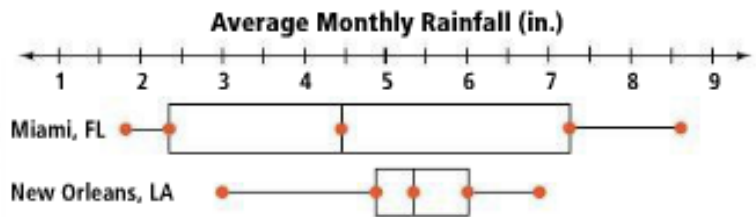
Crops Harvested			
Year	Acres (millions)	Year	Acres (millions)
0	314	6	307
1	321	7	316
2	315	8	312
3	316	9	314
4	314	10	303
5	311		

SOURCE: U.S. Department of Agriculture



**Problem 3: Interpreting Box-and-Whisker Plots**

Use the box-and-whisker plots below. What do the interquartile ranges tell you about the average monthly rainfall for each city?



SOURCE: National Climatic Data Center

**Percentile:**

**Percentile rank:**

**Problem 4: Finding a Percentile Rank**

Of 25 test scores, eight are less than or equal to 75. What is the percentile rank of a test score of 75?

**Box-and-Whisker Plots Practice**

1. What are the minimum, first quartile, median, third quartile, and maximum of each data set?

A. 95, 85, 75, 85, 65, 60, 100, 105, 75, 85, 75

B. 11, 19, 7, 5, 21, 53

2. What box-and-whisker plot represents the following monthly sales, in millions of dollars, of audio devices: 15, 4, 9, 16, 10, 16, 8, 14, 25, 34



3. In Problem 2, what do the medians tell you about the average monthly rainfalls for Miami and New Orleans?

4. A. Of the 25 scores in Problem 4, there are 15 scores less than or equal to 85. What is the percentile rank of 85?

B. Is it possible to have a percentile rank of 0? Explain.

**Identify the minimum, first quartile, median, third quartile, and maximum of each data set. Then make a box-and-whisker plot of each data set.**

5. file sizes (megabytes): 54, 100, 84, 124, 188, 48, 256

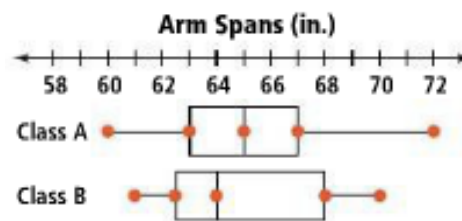


6. daily attendance: 29, 24, 28, 32, 30, 26, 33





7. In the box-and-whisker plots below, which class has the greater interquartile range of arm spans?



8. Which portion of a box-and-whisker plot represents the interquartile range?

9. Students taking a make-up test receive the following grades: 77, 89, 88, 67, 91, 95, 83, 79, 81, and 65. Which grade has a percentile rank of 70?

10. About what percent of the data in a data set falls between the minimum value and the third quartile? Explain.

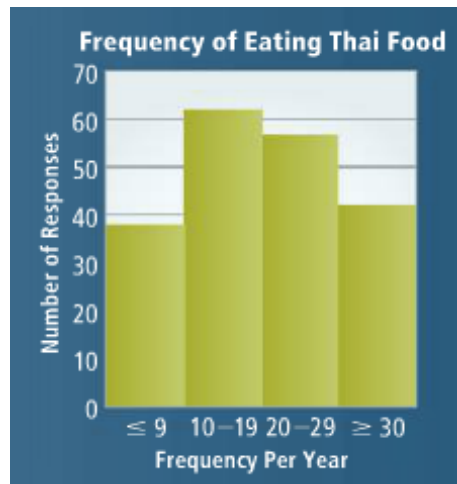
11. A test is graded on a scale from 0 to 100. Your friend says that if you score a 78, your percentile rank must be 78. Is your friend correct? Explain.

## Samples and Surveys

**LT#5:** Classify data and analyze samples and surveys.

### Warm Up

The manager of a Thai restaurant wants to know how often people in her city eat Thai food. She surveys 200 of her customers. The manager says the results in the histogram shown that most people in the city eat Thai food at least 10 times per year. Do you think there were any problems with the survey? Explain.



### Vocabulary & Examples

Types of Data	Description	Examples
Quantitative		
Qualitative		

#### Problem 1: Classifying Data

Is each data set *quantitative* or *qualitative*?

A. favorite movies

B. numbers of students in different schools who take Spanish

C. football jersey numbers

**Univariate:**

**Bivariate:**

**Problem 2: Identifying Types of Data**

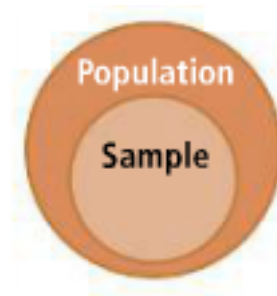
Is each data set *univariate* or *bivariate*?

A. the atomic weights of the elements in the periodic table

B. the edge lengths and volumes of cubes

**Population:**

**Sample:**



When designing a survey, you should choose a sample that reflects the population.

Name	Sampling Method	Example
Random		
Systematic		
Stratified		

**Problem 3: Choosing a Sample**

You want to find out how many DVDs students at your school rent in a month. You interview every tenth teenager you see at a mall. What sampling method are you using? Is this a good sample?

**Bias:**

**Problem 4: Determine Bias in a Survey Question**

A reporter wants to find out what kinds of movies are most popular with local residents. The reporter asks, “Do you prefer exciting action movies or boring documentaries?” Is the question biased? Explain.

**Problem 5: Determining Bias in a Sample**

You want to determine what percent of teens ages 14 to 18 watch wrestling on TV. At a high school wrestling match, you ask every third teenager whether he or she watches wrestling on TV. How might this cause bias in the results of your survey?

### Samples and Surveys Practice

1. Is each data set *qualitative* or *quantitative*? Explain.

A. costs of CDs

B. eye colors

2. Is each data set *univariate* or *bivariate*? Explain.

A. heights and weights of mammals

B. the cost of Internet service from several different providers

3. You revise your plan and interview all students leaving a school assembly who are wearing the school colors. Will this plan give a good sample? Explain.

4. How can the question in Problem 4 be reworded so that it is not biased?

5. You want to know how many of your classmates have cell phones. To determine this, you send every classmate an e-mail asking, "Do you own a cell phone?" How might this method of gathering data affect the result of your survey?

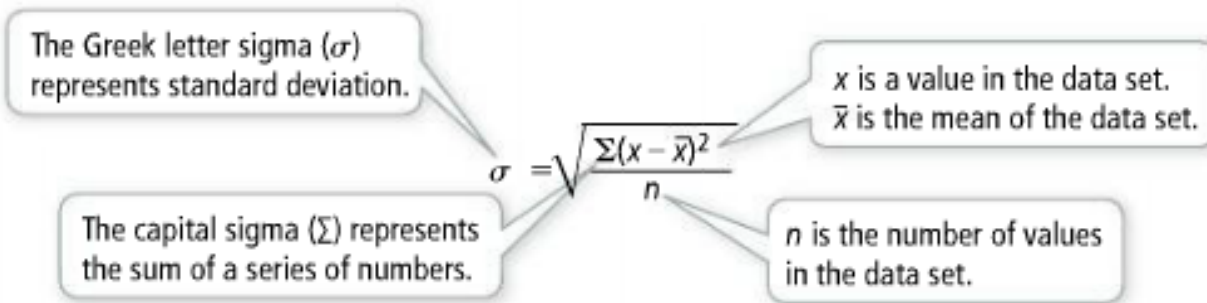
**Determine whether each sampling method is *random*, *systematic*, or *stratified*.**

1. You survey every tenth student who enters the cafeteria.
2. You draw student ID numbers out of a hat and survey those students.
3. You survey two students at random from each class.
4. Is a data set of your class's test scores *qualitative* or *quantitative* data?
5. Explain why "Do you prefer delicious fruit or plain vegetables for a snack food?" is a biased survey question.
6. What is the difference between univariate data and bivariate data? Give an example of each type of data.

### Standard Deviation Activity

You have learned about one measure of dispersion, range. Another measure of dispersion is *standard deviation*. **Standard deviation** is a measure of how the values in a data set vary, or deviate, from the mean.

Statisticians use several special symbols in the formula for standard deviation.



### Example

Find the mean and standard deviation of each data set. Which data set has a greater standard deviation? Use tables to help organize your work.

**Step 1** Find the mean,  $\bar{x}$ .

**Step 2** Find the difference between each data value and the mean,  $x - \bar{x}$ .

**Step 3** Square each difference,  $(x - \bar{x})^2$ .

**Step 4** Find the average (mean) of these squares,  $\frac{\sum(x - \bar{x})^2}{n}$ .

**Step 5** Take the square root to find the standard deviation,  $\sqrt{\frac{\sum(x - \bar{x})^2}{n}}$ .

Data Set 1				Data Set 2					
$x_1$	$\bar{x}_1$	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	$x_2$	$\bar{x}_2$	$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$		
12.6	15	-2.4	5.76	13.4	14.5	-1.1	1.21		
15.1	15	0.1	0.01	11.7	14.5	-2.8	7.84		
11.2	15	-3.8	14.44	18.3	14.5	3.8	14.44		
17.9	15	2.9	8.41	14.8	14.5	0.3	0.09		
18.2	15	3.2	10.24	14.3	14.5	-0.2	0.04		
			$\frac{\sum(x_1 - \bar{x}_1)^2}{n}$	7.772				$\frac{\sum(x_2 - \bar{x}_2)^2}{n}$	4.724
			$\sqrt{\frac{\sum(x_1 - \bar{x}_1)^2}{n}}$	$\approx 2.79$				$\sqrt{\frac{\sum(x_2 - \bar{x}_2)^2}{n}}$	$\approx 2.17$

Data set 1 has a greater standard deviation at 2.79.

### Practice

Find the mean and standard deviation of each data set. Round to the nearest hundredth. Which data set has a greater standard deviation?

1. Data set 1: 4, 8, 5, 12, 3, 9, 5, 2

Data set 2: 5, 9, 11, 4, 6, 11, 2, 7

2. Data set 1: 102, 98, 103, 86, 101, 110

Data set 2: 90, 89, 100, 97, 102, 97

3. Data set 1: 8.2, 11.6, 8.7, 10.6, 9.4, 10.1, 9.3

Data set 2: 9.3, 10.2, 8.1, 12.3, 8.7, 9.9, 10.1

4. Data set 1: 32, 40, 35, 28, 42, 32, 44

Data set 2: 40, 38, 51, 39, 46, 40, 52

### **Designing Your Own Survey Activity**

You have learned how to organize, display, and summarize data. In this activity you will explore methods of collecting data.

Suppose a statistician is trying to predict how a town will vote in an upcoming election. She could ask every person in town, but this method takes too much time and work. Instead, she might rely on an information-gathering survey that is sent to only some people in the town. She can then use the results to predict how other people in the town might vote.

When you design a survey, you need to make sure that the people you survey are representative of the group you want to study.

#### **Activity 1**

Suppose you want to find out how many hours of exercise the students at your school get each week. At the school gym you ask everybody you see, "How many hours of exercise do you get every week?"

1. Will the results of your survey be representative of your entire school? Explain.
2. Is there a better location to conduct your survey?
3. Suppose you asked, "Do you work out every day like a healthy person, or are you a lazy couch potato who only works out once in a while?" Do you think the results of your survey would change? Explain your reasoning.

#### **Activity 2**

In this activity, you will design and conduct a survey.

4. Select a topic for your survey. You could ask about a favorite sporting events, snacks, musical instruments, or another topic of your choice.
5. What question will you ask? Will your question influence the opinion of the people you are surveying?
6. What group of people do you want to study? Are you going to ask the entire group or just a portion of the whole group?
7. Complete your survey.
8. Summarize your results with a graph and a brief description.
9. Are the people you surveyed representative of the group you want to study? Explain.



### Two-Way Frequency Tables Activity

Two-way frequency tables are a convenient way to show data.

#### Activity 1

The table at the right gives information about ticket sales for two movies. The table also separates those sales into ticket purchases by men and ticket purchases by women.

Gender	Movie	
	<i>Story of Love</i>	<i>Martial Arts Champ</i>
Men	78	12
Women	25	86

Of the 103 tickets sold for *Story of Love*, 78 tickets, or 75.7%, were sold to men. Of the 90 men who purchased tickets, 78 of them, or 86.7%, purchased a a ticket for *Story of Love*.

**Calculate the ratio and percent for each of the following.**

- Men who purchased tickets to *Martial Arts Champ* to all men.
- Men who purchased tickets to *Martial Arts Champ* to all who purchased tickets to *Martial Arts Champ*.
- Women who purchased tickets to *Marial Arts Champ* to all who purchased tickets to *Martial Arts Champ*.
- Women who purchased tickets to *Story of Love* to all who purchased tickets to *Story of Love*.
- Which movie is more popular with women? With men?

#### Activity 2

You can use two-way frequency tables to make better predictions.

The citizens of Parkdale are preparing to vote on a bond issue to fund an expansion of the public library. The table at the right records data about support for the bond issue.

	Age		
	18–25	26–64	≥ 65
Support	79	55	33
Oppose	21	45	67

**For each of the following, calculate the ratio and percent in relation to all people in the given age group.**

- In people between 18 and 25 who support the bond issue.
- The people between 26 and 64 who support the bond issue.
- The people 65 or older who oppose the bond issue.
- The people between 26 and 64 who oppose the bond issue.
- An expert predicts that 10% of the voters will be between 18 and 25, 40% will be between 26 and 64, and 50% will be 65 or older. Based on her predictions, she can calculate the result of the election from the expression  $0.1a + 0.4b + 0.5c$  where  $a$ ,  $b$ , and  $c$  are the decimal forms of the ratio of people supporting the bond issue in the three groups. Predict the percent of votes in favor of the bond issue.