

## Lesson

## 1-6

# Measures of Spread: Variance and Standard Deviation

► **BIG IDEA** Variance and standard deviation depend on the mean of a set of numbers. Calculating these measures of spread depends on whether the set is a sample or population.

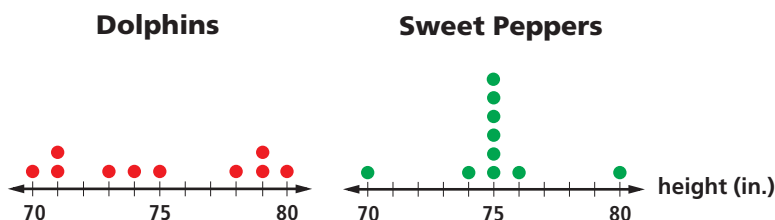
## Vocabulary

range  
 deviations  
 population variance  $\sigma^2$   
 population standard deviation  $\sigma$   
 sample variance  $s^2$   
 sample standard deviation  $s$

So far in this book you have studied the mean, the mode, the five-number summary, and the IQR. These statistics help to describe the distribution of numbers in a data set. In the Activity below you will use these statistics to compare two given data sets.

### Activity 1

The two dot plots below display frequency distributions of the height of the players on two hypothetical women's basketball teams.



- Step 1** Describe what you think is the main difference between the two dot plots from just looking at the graphs.
- Step 2** Find the mean, median, mode, five-number summary, and IQR for each data set.
- Step 3** Determine if any of the values from Step 2 are appropriate for distinguishing the main difference between the heights of the members of the two teams. Justify your conclusions.

In the Activity you may have concluded that the IQR helps to distinguish between the spreads of the two dot plots. It is one of the measures of *spread of a distribution*. Yet the IQR for the Sweet Peppers is 0 inches, which implies no spread. This number is not sensitive enough to indicate that there are Sweet Pepper players who are not 75" tall.

The simplest measure of the spread of a distribution is its *range*. The **range** is the difference of the maximum and minimum values of the variable. Each team has a mean height of 75" and a range of 80" – 70", or 10". In this way the distributions are quite similar. Yet the heights of the Dolphins seem more spread out than those of the Sweet Peppers.

### Mental Math

Lenny's average score after 3 tests is 88. What score on the 4th test would bring Lenny's average up to exactly 90?

So we need other statistical measures to better describe the spread of the heights of the players of each team. Two measures of spread that are influenced by every data point are *variance* and *standard deviation*.

## The Variance and Standard Deviation of a Population

The Dolphins can be viewed as a population of ten women. When the set of data is a population, Greek letters are used for mean, variance, and standard deviation. The mean is labeled  $\mu$  (mu), variance as  $\sigma^2$  (sigma-squared), and standard deviation as  $\sigma$  (sigma). The variance for a population is calculated from the squares of **deviations**, or differences of each data value  $x_i$  from the mean  $\mu$ . The shortest Dolphin player is 70" tall. The deviation  $x_i - \mu$  for that player is  $(70 - 75) = -5$  in. The square of her deviation is 25 square inches. Another of the Dolphin players is 78 inches tall. Her squared deviation is  $(78 - 75)^2 = 9$  in<sup>2</sup>.

The **population variance** is the mean of the squared deviations. That is, where  $n$  is the number of objects in a population, the variance is the sum of the squared deviations divided by  $n$ . The **population standard deviation** is the square root of the population variance. Example 1 shows how you can compute population variance and standard deviation by hand or by using a statistics utility.

### Example 1

Find the variance and standard deviation for the heights of the Dolphins (treating them as a population).

**Solution** To find the variance and standard deviation it helps to organize the work step-by-step.

- Step 1** Write the data  $x_i$  in a column. Find the mean by adding these numbers and dividing by  $n$ , the number of data points, which in this case is 10. Since the sum is 750,  $\mu = 75$ .
- Step 2** In the next column record the result of subtracting the mean from each score, yielding  $x_i - \mu$ , which in this case is  $x_i - 75$ . Deviations are either positive, zero, or negative.
- Step 3** Square each deviation and record each result in the next column.
- Step 4** Add the squares of the deviations. Divide the sum of the squared deviations by  $n$ , in this case 10, to obtain the variance  $\sigma^2$ .
- Step 5** Find the square root of the variance to get the standard deviation  $\sigma$ .

Results of these steps using technology are shown at the right and without using technology on the next page.

	A he	B av	C de	D sq
1	70	75	-5	25
2	71	75	-4	16
3	71	75	-4	25
4	73	75	-2	16
5	74	75	-1	25

varPop(height)	64
stDevPop(height)	8

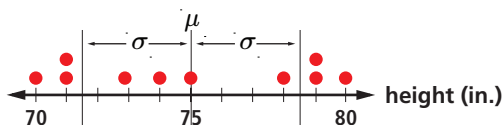
	Height (in.)	Deviation (in.)	Square of Deviation (in <sup>2</sup> )
	$x_i$	$x_i - \mu$	$(x_i - \mu)^2$
	70	$70 - 75 = -5$	25
	71	$71 - 75 = -4$	16
	71	-4	16
	73	-2	4
	74	-1	1
	75	0	0
	78	3	9
	79	4	16
	79	4	16
Sum	750	0	128

The mean  $\mu$  is  $\frac{750}{10} = 75$  in., the variance  $\sigma^2 = \frac{128}{10} = \frac{64}{5} = 12.8$  in<sup>2</sup>, and the standard deviation  $\sigma = \sqrt{\frac{128}{10}} = \frac{8\sqrt{5}}{5} = \sqrt{12.8} \approx 3.58$  in.

Notice that the sum of the deviations equals 0. This is a great way to check your work. Also, notice that when the deviations are squared, values farther from the mean contribute more to the variance than values close to the mean. For instance, a height of 80 contributes  $(80 - 75)^2 = 5^2 = 25$  to the sum of squared deviations, but 74 contributes only  $(74 - 75)^2 = (-1)^2 = 1$ . Because of this, groups with more data close to the mean generally have smaller standard deviations than groups with more data far from the mean.

Below is a picture of the Dolphins' data showing how the standard deviation of 3.58 relates to the distribution.

### Dolphins



You are asked to calculate the variance and standard deviation for the Sweet Peppers in Question 3.

Formulas for the variance and standard deviation are usually written using  $\Sigma$ -notation. For a set of  $n$  numbers  $x_1, x_2, x_3, \dots, x_n$  each deviation from the mean can be written as  $x_i - \mu$ , and the square of the deviation as  $(x_i - \mu)^2$ . Because the definition of the variance and standard deviation are based on the mean, they can be used only when it makes sense to calculate a mean.



### Basketball Beginnings

Basketball was introduced to women at Smith College in 1892, just one year after the game was invented.

### Definition of Variance and Standard Deviation of a Population

Let  $\mu$  be the mean of the population data set  $x_1, x_2, \dots, x_n$ . Then the **variance**  $\sigma^2$  and **standard deviation**  $\sigma$  of the population are

$$\sigma^2 = \frac{\text{sum of squared deviations}}{n} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

and  $\sigma = \sqrt{\text{variance}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}.$

### Variance and Standard Deviation of a Sample

For many years statisticians only used the population formulas. Over time, mathematicians and statisticians established that dividing by  $n$  for a sample variance did not produce the best estimate of the population variance. They showed that when using a sample, dividing by  $n - 1$  rather than by  $n$  provided a better estimate of population variance.

When the data is a sample, Roman letters are used for mean, variance, and standard deviation. The mean is labeled  $\bar{x}$  (read “x-bar”), variance is  $s^2$ , and standard deviation is  $s$ . The only difference in the formula is that  $n - 1$  is used in place of  $n$ .

### Definition of Variance and Standard Deviation of a Sample

Let  $\bar{x}$  be the mean of the sample data set  $x_1, x_2, \dots, x_n$ .

Then the **variance**  $s^2$  and **standard deviation**  $s$  of the sample are

$$s^2 = \frac{\text{sum of squared deviations}}{n - 1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

and  $s = \sqrt{\text{variance}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}.$

**CAUTION:** In this book most of the data come from samples, so unless directed otherwise, use the variance and standard deviation formulas for samples.

#### GUIDED

#### Example 2

According to the U.S. Department of Agriculture, ten to twenty earthworms per cubic foot is a sign of healthy soil. Mr. Green checked the soil in his garden by digging 7 one-cubic-foot holes and counting the earthworms. He found the following counts: 4, 23, 15, 10, 8, 12, 18.

Calculate the sample variance and sample standard deviation of the numbers of earthworms per cubic foot.

**Solution** Follow the same steps used in Example 1, but since this data represents a sample, use the variance and standard deviation formulas for samples. The symbols  $x_i$ ,  $x_i - \bar{x}$ , and  $(x_i - \bar{x})^2$  represent the earthworm count, deviation from the mean, and squared deviation, respectively.

Count (worms) $x_i$	Deviation (worms) $x_i - \bar{x}$	Square of Deviation (worms squared) $(x_i - \bar{x})^2$
4	-8.86	78.5
23	?	102.82
15	2.14	?
10	-2.86	8.18
8	?	23.62
?	-0.86	0.74
?	5.14	?
sum $\sum_{i=1}^7 x_i = ?$ $? \approx 0$ $\sum_{i=1}^7 (x_i - \bar{x})^2 = ?$		

The mean  $\bar{x} = \frac{\sum_{i=1}^7 x_i}{7} = \frac{90}{7} \approx 12.86$  worms.

The variance  $s^2 = \frac{\sum_{i=1}^7 (x_i - \bar{x})^2}{?} = \frac{?}{?} = ?$  worms squared.

The standard deviation  $s = \sqrt{\text{variance}} = \sqrt{?} \approx ?$  worms, to two decimal places.



**Wiggle, Squiggle, and Squirm** In one acre of land, you can often find more than one million worms.

Because of these different formulas, some statistics utilities have two sets of symbols:  $s^2$  and  $s$ , and  $\sigma$  and  $\sigma^2$ . Other calculators and programs use only one set of formulas for variance and standard deviation.

## Activity 2

Use your calculator or statistics software to find (to the nearest tenth) the standard deviation of the following data set.

89, 79, 74, 67, 99, 91, 84, 81

If more than one standard deviation is given, record both. (You should find that the mean is 83 and both the sample and population standard deviation are between 9 and 11.)

a:={89,79,74,67,99,91,84,81}	
{89,79,74,67,99,91,84,81}	
OneVar a,1: stat.results	
"Title"	"One-Variable Stat
" $\bar{x}$ "	83.
" $\Sigma x$ "	664.
" $\Sigma x^2$ "	55826.
"sX := s <sub>n-1</sub> X"	10.0995
" $\sigma_X := \sigma_n X$ "	9.44722
2/99	



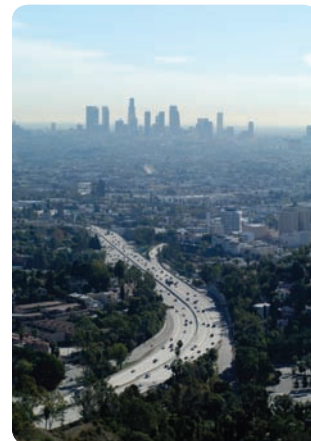
## Questions

### COVERING THE IDEAS

- State whether the statistic is a measure of center or a measure of spread.
  - mean
  - range
  - variance
  - interquartile range
  - standard deviation
  - median
- Multiple Choice** The standard deviation of a set of scores is
  - the sum of the deviations.
  - the difference between the highest and lowest scores.
  - the score that occurs with the greatest frequency.
  - none of the above.
- Use the heights of the Dolphins and Sweet Peppers.
  - Calculate the variance and standard deviation for the heights of the Sweet Peppers using the formulas for populations.
  - Find the difference of the means.
  - Find the difference of the ranges.
  - Find the difference of the standard deviations.
  - Explain in your own words what the differences in Parts b–d tell you about the two data sets.
- What statistics change if the Dolphins and Sweet Peppers are considered samples of all women basketball players?
- If the standard deviation  $s = 4.5$  cm, find the sample variance.

In 6 and 7, the measurements refer to pulse rates of two students while jogging, in beats per minute. Use a calculator or statistical software to find the mean and sample standard deviation for each situation.

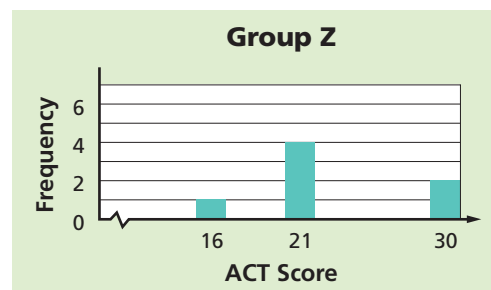
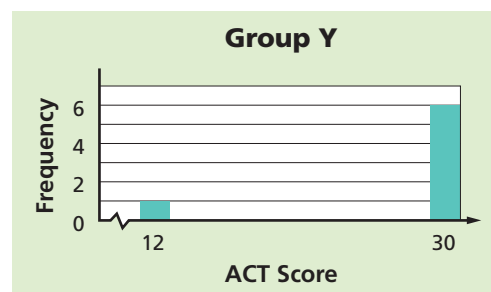
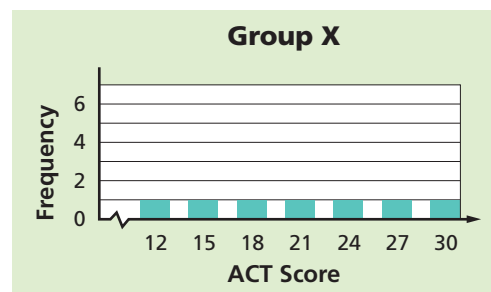
- student A with four measurements: 100, 120, 115, 133
- student B with five measurements: 110, 120, 124, 116, 120
- Suppose you used the formula for the sample standard deviation in Example 1. Would your answer be greater than, equal to, or less than the population standard deviation shown there?
- Each of the following situations produces data that can be summarized with mean and standard deviation. Which would require population formulas for variance and standard deviation rather than sample formulas?
  - The Environmental Protection Agency measures carbon monoxide content of air at 15 locations of a metropolitan area.
  - An algebra teacher has scores from his students' final exams.
  - A consumer magazine tests four cars from each of three brands of hybrid vehicles to evaluate operating cost per mile.
  - Number of home runs per game for the Boston Red Sox in the 2008 season.



Los Angeles, CA

### APPLYING THE MATHEMATICS

10. Suppose you know the distance in miles each student in a class lives from school. For this data set, state the unit for each statistic.
- mean
  - range
  - variance
  - standard deviation
11. Beth found the variance of a data set to be  $-11$ . Why must her answer be wrong?
12. Suppose two samples have the same mean, but different standard deviations  $s_1$  and  $s_2$ , with  $s_1 < s_2$ . Which sample shows more variability?
13. a. Consider the weights (in kilograms) of a group of deer. If the standard deviation is  $7.8$  kg, what is the variance?  
 b. If the variance is  $19$  kg<sup>2</sup>, what is the standard deviation?
14. Use the hypothetical frequency distributions of ACT scores for groups X, Y, and Z at the right.
- Match each group with its best description.
    - consistently near the mean
    - very widely spread
    - evenly distributed.
  - Without calculating, tell which group's ACT scores have the greatest standard deviation and which have the smallest.
  - Verify your answer to Part b with calculations.
15. More than 1.3 million students in the class of 2007 took the ACT. On the mathematics section,  $\mu = 21.0$  and  $\sigma = 5.1$ . Students receive scores rounded to the nearest whole number. What is the interval of student scores that lie within one standard deviation of the mean?



For 16 and 17, use the following data that represent the times (rounded to the nearest 5 seconds) for 20 sixth-graders to run 400 meters.

70 80 80 85 90 100 100 100 100 100  
 100 105 105 105 120 130 130 130 140 150

16. Find a sample of five students out of the 20 whose standard deviation for running time is as small as possible.
17. Find a sample of four running times whose standard deviation is larger than 25 seconds. Compute the standard deviation.

18. **Multiple Choice** A class of students is said to be *homogeneous* if the students in the class are very much alike on some measure. Here are four classes of students who were tested on a 20-point chemistry test. Which class is the most homogeneous with respect to scores on the test? Explain your answer.

- A  $n = 20$                        $\bar{x} = 15.3$                        $s = 2.5$   
 B  $n = 25$                        $\bar{x} = 12.1$                        $s = 5.4$   
 C  $n = 18$                        $\bar{x} = 11.3$                        $s = 3.2$   
 D  $n = 30$                        $\bar{x} = 10.4$                        $s = 3.2$

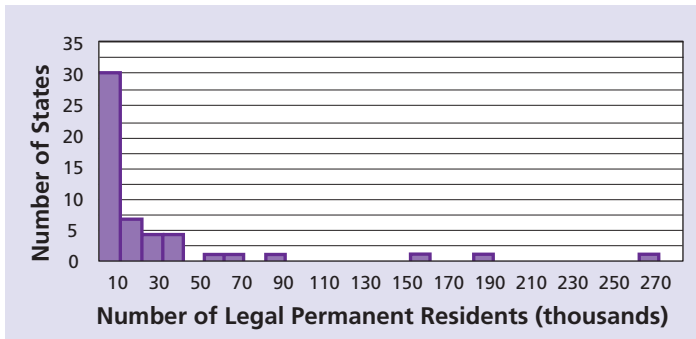
**REVIEW**

19. Use the table below on seasonally adjusted U.S. domestic imports for 2007.

U.S. Total Imports in Goods and Services (\$ billions)												
TOTAL	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Total (per month)	187	186	192	191	193	195	197	197	198	200	205	203
Cumulative Total	?	?	?	?	?	?	?	?	?	?	?	?

Source: U.S. Census Bureau

- a. Complete the table and make both a histogram and a cumulative data line graph for imports.  
 b. What was the total cost of U.S. domestic imports for 2007? (Lessons 1-5, 1-3)
20. Two data sets of heights of people each have minimum = 50", median = 67", and maximum = 80". One data set has IQR = 15"; the other has IQR = 10".  
 a. Draw possible box plots for each data set.  
 b. Which data set shows more spread? (Lessons 1-4)
21. The histogram below shows the number of states receiving the number of legal permanent residents specified in each interval in 2006. Write a paragraph describing immigration in 2006. Include both specific information such as maximum, minimum, mean, or median values (when possible), and general trends such as skewness. (Lessons 1-3, 1-2, 1-1)



Source: U.S. Department of Homeland Security



