2.4 Measures of Variation

What You SHOULD LEARN

- How to find the range of a data set
- How to find the variance and standard deviation of a population and of a sample
- How to use the Empirical Rule and Chebychev's Theorem to interpret standard deviation
- How to approximate the sample standard deviation for grouped data

Range ▸ Deviation, Variance, and Standard Deviation ▸ Interpreting Standard Deviation ▸ Standard Deviation for Grouped Data

▸ Range

In this section, you will learn different ways to measure the variation of a data set. The simplest measure is the range of the set.

DEFINITION

The **range** of a data set is the difference between the maximum and minimum data entries in the set. To find the range, the data must be quantitative.

Range = (Maximum data entry) − (Minimum data entry)

EXAMPLE 1

Finding the Range of a Data Set

Two corporations each hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries for Corporation A.

Starting Salaries for Corporation A (1000s of dollars)

Salary	41	38	39	45	47	41	44	41	37	42

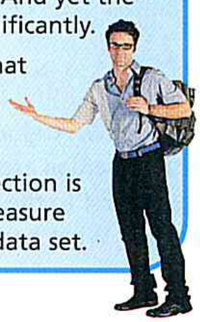
Starting Salaries for Corporation B (1000s of dollars)

Salary	40	23	41	50	49	32	41	29	52	58

Insight

Both data sets in Example 1 have a mean of 41.5, a median of 41, and a mode of 41. And yet the two sets differ significantly.

The difference is that the entries in the second set have greater variation. Your goal in this section is to learn how to measure the variation of a data set.



Solution Ordering the data helps to find the least and greatest salaries.

37 38 39 41 41 41 42 44 45 47

Minimum (37) Maximum (47)

Range = (Maximum salary) − (Minimum salary)
= 47 − 37
= 10

So, the range of the starting salaries for Corporation A is 10, or $10,000.

▸ Try It Yourself 1

Find the range of the starting salaries for Corporation B.

a. Identify the *minimum* and *maximum* salaries.
b. Find the *range*.
c. Compare your answer with that for Example 1.

Answer: Page A35

Deviation, Variance, and Standard Deviation

As a measure of variation, the range has the advantage of being easy to compute. Its disadvantage, however, is that it uses only two entries from the data set. Two measures of variation that use all the entries in a data set are the variance and the standard deviation. However, before you learn about these measures of variation, you need to know what is meant by the deviation of an entry in a data set.

DEFINITION

The **deviation** of an entry x in a population data set is the difference between the entry and the mean μ of the data set.

Deviation of $x = x - \mu$

Deviations of Starting Salaries for Corporation A

Salary (1000s of dollars) x	Deviation (1000s of dollars) $x - \mu$
41	−0.5
38	−3.5
39	−2.5
45	3.5
47	5.5
41	−0.5
44	2.5
41	−0.5
37	−4.5
42	0.5
$\Sigma x = 415$	$\Sigma(x - \mu) = 0$

EXAMPLE 2

Finding the Deviations of a Data Set

Find the deviation of each starting salary for Corporation A given in Example 1.

Solution The mean starting salary is $\mu = 415/10 = 41.5$. To find out how much each salary deviates from the mean, subtract 41.5 from the salary. For instance, the deviation of 41 (or $41,000) is

$$41 - 41.5 = -0.5 \text{ (or } -\$500\text{)}.$$

Deviation of $x = x - \mu$

x (41) $\quad\mu$ (41.5)

The table at the left lists the deviations of each of the 10 starting salaries.

Try It Yourself 2

Find the deviation of each starting salary for Corporation B given in Example 1.

a. Find the *mean* of the data set.

b. *Subtract* the mean from each salary.

Answer: Page A35

In Example 2, notice that the sum of the deviations is zero. Because this is true for any data set, it doesn't make sense to find the average of the deviations. To overcome this problem, you can square each deviation. When you add the squares of the deviations, you compute a quantity called the **sum of squares,** denoted SS_x. In a population data set, the mean of the squares of the deviations is called the **population variance.**

DEFINITION

The **population variance** of a population data set of N entries is

$$\text{Population variance} = \sigma^2 = \frac{\Sigma(x - \mu)^2}{N}$$

The symbol σ is the lowercase Greek letter sigma.

DEFINITION

The **population standard deviation** of a population data set of N entries is the square root of the population variance.

$$\text{Population standard deviation} = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

GUIDELINES

Finding the Population Variance and Standard Deviation

In Words	*In Symbols*
1. Find the mean of the population data set.	$\mu = \frac{\Sigma x}{N}$
2. Find the deviation of each entry.	$x - \mu$
3. Square each deviation.	$(x-\mu)^2$
4. Add to get the **sum of squares.**	$SS_x = \Sigma(x-\mu)^2$
5. Divide by N to get the **population variance.**	$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$
6. Find the square root of the variance to get the **population standard deviation.**	$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

Sum of Squares of Starting Salaries for Corporation A

Salary x	Deviation $x - \mu$	Squares $(x-\mu)^2$
41	−0.5	0.25
38	−3.5	12.25
39	−2.5	6.25
45	3.5	12.25
47	5.5	30.25
41	−0.5	0.25
44	2.5	6.25
41	−0.5	0.25
37	−4.5	20.25
42	0.5	0.25
	$\Sigma = 0$	$SS_x = 88.5$

EXAMPLE 3

Finding the Population Standard Deviation

Find the population variance and standard deviation of the starting salaries for Corporation A given in Example 1.

Solution The table at the left summarizes the steps used to find SS_x.

$$SS_x = 88.5, \quad N = 10, \quad \sigma^2 = \frac{88.5}{10} \approx 8.9, \quad \sigma = \sqrt{8.85} \approx 3.0$$

So, the population variance is about 8.9, and the population standard deviation is about 3.0, or $3000.

▸Try It Yourself 3

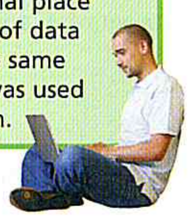
Find the population standard deviation of the starting salaries for Corporation B given in Example 1.

a. Find the *mean* and each *deviation,* as you did in Try It Yourself 2.
b. *Square* each deviation and *add* to get the sum of squares.
c. *Divide* by N to get the population variance.
d. Find the *square root* of the population variance.
e. *Interpret* the results by giving the population standard deviation in dollars.

Answer: Page A35

Study Tip

Notice that the variance and standard deviation in Example 3 have one more decimal place than the original set of data values has. This is the same *round-off rule* that was used to calculate the mean.



Study Tip

Note that when you find the *population* variance, you divide by N, the number of entries, but, for technical reasons, when you find the *sample* variance, you divide by $n - 1$, one less than the number of entries.

Symbols in Variance and Standard Deviation Formulas

	Population	Sample
Variance	σ^2	s^2
Standard deviation	σ	s
Mean	μ	$\bar{x}$
Number of entries	N	n
Deviation	$x - \mu$	$x - \bar{x}$
Sum of squares	$\Sigma(x - \mu)^2$	$\Sigma(x - \bar{x})^2$

See MINITAB and TI-83/84 steps on pages 124 and 125.

DEFINITION

The **sample variance** and **sample standard deviation** of a sample data set of n entries are listed below.

$$\text{Sample variance} = s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$$

$$\text{Sample standard deviation} = s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

GUIDELINES

Finding the Sample Variance and Standard Deviation

In Words	*In Symbols*
1. Find the mean of the sample data set.	$\bar{x} = \frac{\Sigma x}{n}$
2. Find the deviation of each entry.	$x - \bar{x}$
3. Square each deviation.	$(x - \bar{x})^2$
4. Add to get the **sum of squares.**	$SS_x = \Sigma(x - \bar{x})^2$
5. Divide by $n - 1$ to get the **sample variance.**	$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$
6. Find the square root of the variance to get the **sample standard deviation.**	$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$

EXAMPLE 4

Finding the Sample Standard Deviation

The starting salaries given in Example 1 are for the Chicago branches of Corporations A and B. Each corporation has several other branches, and you plan to use the starting salaries of the Chicago branches to estimate the starting salaries for the larger populations. Find the *sample* standard deviation of the starting salaries for the Chicago branch of Corporation A.

Solution

$$SS_x = 88.5, \qquad n = 10, \qquad s^2 = \frac{88.5}{9} \approx 9.8, \qquad s = \sqrt{\frac{88.5}{9}} \approx 3.1$$

So, the sample variance is about 9.8, and the sample standard deviation is about 3.1, or \$3100.

▸ Try It Yourself 4

Find the sample standard deviation of the starting salaries for the Chicago branch of Corporation B.

a. Find the *sum of squares*, as you did in Try It Yourself 3.
b. *Divide* by $n - 1$ to get the sample variance.
c. Find the *square root* of the sample variance.

Answer: Page A35

Office Rental Rates		
35.00	33.50	37.00
23.75	26.50	31.25
36.50	40.00	32.00
39.25	37.50	34.75
37.75	37.25	36.75
27.00	35.75	26.00
37.00	29.00	40.50
24.50	33.00	38.00

Study Tip

Here are instructions for calculating the sample mean and sample standard deviation on a TI-83/84 for Example 5.

STAT

Choose the EDIT menu.

1: Edit

Enter the sample office rental rates into L1.

STAT

Choose the CALC menu.

1: 1–Var Stats

ENTER

2nd L1 ENTER

EXAMPLE 5

Using Technology to Find the Standard Deviation

Sample office rental rates (in dollars per square foot per year) for Miami's central business district are shown in the table. Use a calculator or a computer to find the mean rental rate and the sample standard deviation. *(Adapted from Cushman & Wakefield Inc.)*

Solution MINITAB, Excel, and the TI-83/84 each have features that automatically calculate the mean and the standard deviation of data sets. Try using this technology to find the mean and the standard deviation of the office rental rates. From the displays, you can see that $\bar{x} \approx 33.73$ and $s \approx 5.09$.

MINITAB

Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev
Rental Rates	24	33.73	35.38	33.88	5.09

Variable	SE Mean	Minimum	Maximum	Q1	Q3
Rental Rates	1.04	23.75	40.50	29.56	37.44

EXCEL

	A	B
1	Mean	33.72917
2	Standard Error	1.038864
3	Median	35.375
4	Mode	37
5	Standard Deviation	5.089373
6	Sample Variance	25.90172
7	Kurtosis	-0.74282
8	Skewness	-0.70345
9	Range	16.75
10	Minimum	23.75
11	Maximum	40.5
12	Sum	809.5
13	Count	24

TI-83/84

1-Var Stats
$\bar{x}$=33.72916667
Σx=809.5
Σx^2=27899.5
Sx=5.089373342
σx=4.982216639
n=24

Sample Mean
Sample Standard Deviation

▸ Try It Yourself 5

Sample office rental rates (in dollars per square foot per year) for Seattle's central business district are listed. Use a calculator or a computer to find the mean rental rate and the sample standard deviation. *(Adapted from Cushman & Wakefield Inc.)*

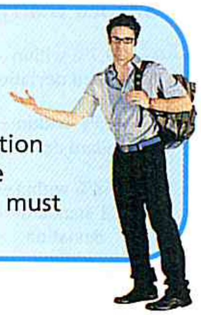
40.00	43.00	46.00	40.50	35.75	39.75	32.75
36.75	35.75	38.75	38.75	36.75	38.75	39.00
29.00	35.00	42.75	32.75	40.75	35.25	

a. *Enter* the data.
b. *Calculate* the sample mean and the sample standard deviation.

Answer: Page A35

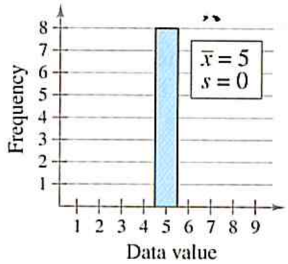
Insight

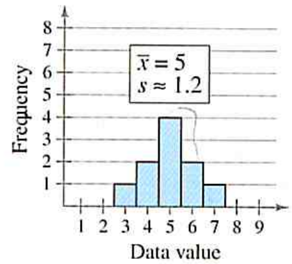
When all data values are equal, the standard deviation is 0. Otherwise, the standard deviation must be positive.

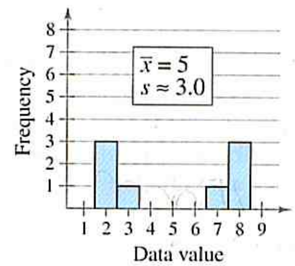


Interpreting Standard Deviation

When interpreting the standard deviation, remember that it is a measure of the typical amount an entry deviates from the mean. The more the entries are spread out, the greater the standard deviation.







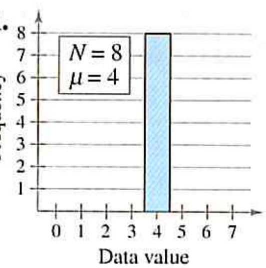
To explore this topic further, see Activity 2.4 on page 100.

EXAMPLE 6

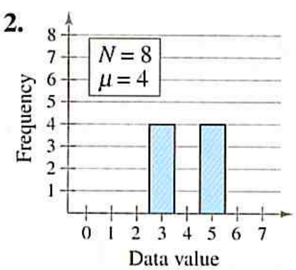
Estimating Standard Deviation

Without calculating, estimate the population standard deviation of each data set.

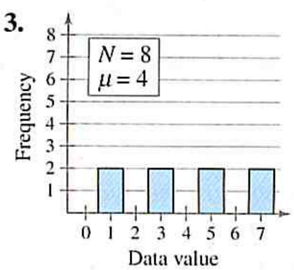
1.



2.



3.



Solution

1. Each of the eight entries is 4. So, each deviation is 0, which implies that

$$\sigma = 0.$$

2. Each of the eight entries has a deviation of ± 1. So, the population standard deviation should be 1. By calculating, you can see that

$$\sigma = 1.$$

3. Each of the eight entries has a deviation of ± 1 or ± 3. So, the population standard deviation should be about 2. By calculating, you can see that

$$\sigma \approx 2.24.$$

Try It Yourself 6

Write a data set that has 10 entries, a mean of 10, and a population standard deviation that is approximately 3. (There are many correct answers.)

a. *Write* a data set that has five entries that are three units less than 10 and five entries that are three units more than 10.

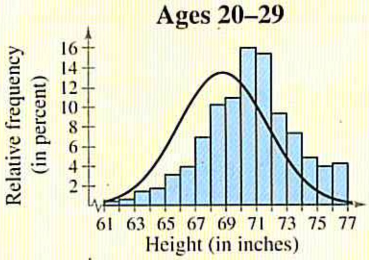
b. *Calculate* the population standard deviation to check that σ is approximately 3.

Answer: Page A35

PICTURING the WORLD

A survey was conducted by the National Center for Health Statistics to find the mean height of males in the U.S. The histogram shows the distribution of heights for the 724 men examined in the 20–29 age group. In this group, the mean was 69.6 inches and the standard deviation was 3.0 inches. (Source: National Center for Health Statistics)

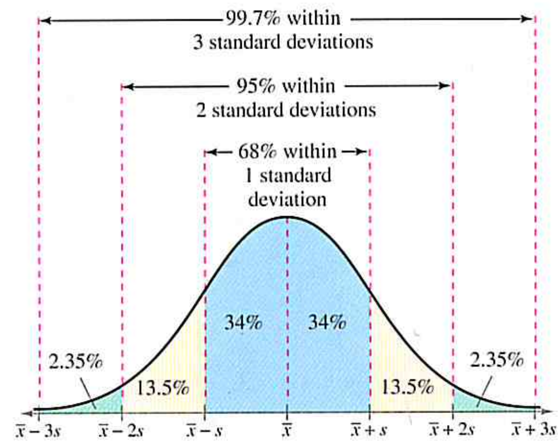
Heights of Men in the U.S. Ages 20–29



Roughly which two heights contain the middle 95% of the data?

Many real-life data sets have distributions that are approximately symmetric and bell shaped. Later in the text, you will study this type of distribution in detail. For now, however, the following *Empirical Rule* can help you see how valuable the standard deviation can be as a measure of variation.

Bell-Shaped Distribution



EMPIRICAL RULE (OR 68-95-99.7 RULE)

For data with a (symmetric) bell-shaped distribution, the standard deviation has the following characteristics.

1. About 68% of the data lie within one standard deviation of the mean.
2. About 95% of the data lie within two standard deviations of the mean.
3. About 99.7% of the data lie within three standard deviations of the mean.

EXAMPLE 7

Using the Empirical Rule

In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20–29) was 64 inches, with a sample standard deviation of 2.71 inches. Estimate the percent of the women whose heights are between 64 inches and 69.42 inches.

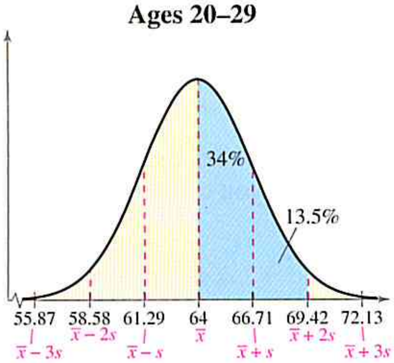
Solution The distribution of the women's heights is shown. Because the distribution is bell shaped, you can use the Empirical Rule. The mean height is 64, so when you add two standard deviations to the mean height, you get

$$\bar{x} + 2s = 64 + 2(2.71) = 69.42.$$

Because 69.42 is two standard deviations above the mean height, the percent of the heights between 64 inches and 69.42 inches is 34% + 13.5% = 47.5%.

Interpretation So, 47.5% of women are between 64 and 69.42 inches tall.

Heights of Women in the U.S. Ages 20–29



▸ Try It Yourself 7

Estimate the percent of the heights that are between 61.29 and 64 inches.

a. How many standard deviations is 61.29 to the left of 64?

b. Use the Empirical Rule to estimate the percent of the data between $\bar{x} - s$ and $\bar{x}$.

c. *Interpret* the result in the context of the data.

Answer: Page A35

Insight

Data values that lie more than two standard deviations from the mean are considered unusual. Data values that lie more than three standard deviations from the mean are very unusual.

The Empirical Rule applies only to (symmetric) bell-shaped distributions. What if the distribution is not bell-shaped, or what if the shape of the distribution is not known? The following theorem gives an inequality statement that applies to *all* distributions. It is named after the Russian statistician Pafnuti Chebychev (1821–1894).

CHEBYCHEV'S THEOREM

The portion of any data set lying within k standard deviations ($k > 1$) of the mean is at least

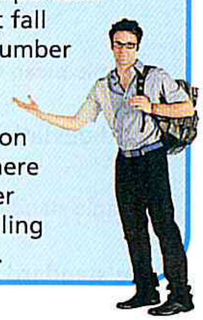
$$1 - \frac{1}{k^2}.$$

- $k = 2$: In any data set, at least $1 - \frac{1}{2^2} = \frac{3}{4}$, or 75%, of the data lie within 2 standard deviations of the mean.
- $k = 3$: In any data set, at least $1 - \frac{1}{3^2} = \frac{8}{9}$, or 88.9%, of the data lie within 3 standard deviations of the mean.

Insight

In Example 8, Chebychev's Theorem gives you an inequality statement that says that at least 75% of the population of Florida is under the age of 88.8. This is a true statement, but it is not nearly as strong a statement as could be made from reading the histogram.

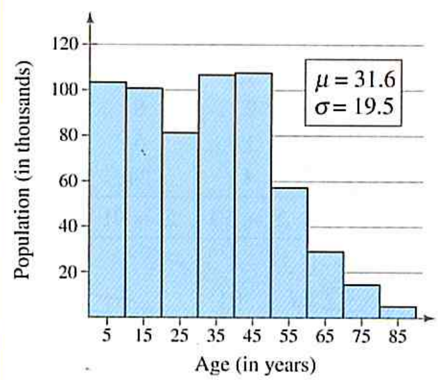
In general, Chebychev's Theorem gives the minimum percent of data values that fall within the given number of standard deviations of the mean. Depending on the distribution, there is probably a higher percent of data falling in the given range.

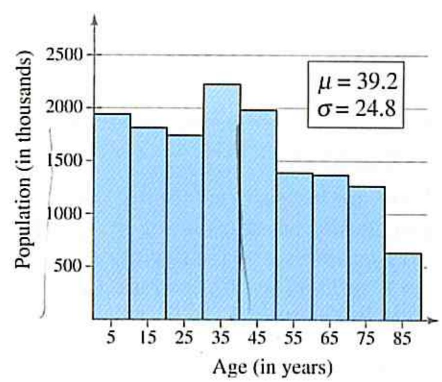


EXAMPLE 8

Using Chebychev's Theorem

The age distributions for Alaska and Florida are shown in the histograms. Decide which is which. Apply Chebychev's Theorem to the data for Florida using $k = 2$. What can you conclude?





Solution The histogram on the right shows Florida's age distribution. You can tell because the population is greater and older. Moving two standard deviations to the left of the mean puts you below 0, because $\mu - 2\sigma = 39.2 - 2(24.8) = -10.4$. Moving two standard deviations to the right of the mean puts you at $\mu + 2\sigma = 39.2 + 2(24.8) = 88.8$. By Chebychev's Theorem, you can say that at least 75% of the population of Florida is between 0 and 88.8 years old.

▸ Try It Yourself 8

Apply Chebychev's Theorem to the data for Alaska using $k = 2$.

a. *Subtract* two standard deviations from the mean.
b. *Add* two standard deviations to the mean.
c. *Apply* Chebychev's Theorem for $k = 2$ and *interpret* the results.

Answer: Page A35

Study Tip

Remember that formulas for grouped data require you to multiply by the frequencies.

Standard Deviation for Grouped Data

In Section 2.1, you learned that large data sets are usually best represented by a frequency distribution. The formula for the sample standard deviation for a frequency distribution is

$$\text{Sample standard deviation} = s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}}$$

where $n = \Sigma f$ is the number of entries in the data set.

EXAMPLE 9

Finding the Standard Deviation for Grouped Data

You collect a random sample of the number of children per household in a region. The results are shown at the left. Find the sample mean and the sample standard deviation of the data set.

Number of Children in 50 Households				
1	3	1	1	1
1	2	2	1	0
1	1	0	0	0
1	5	0	3	6
3	0	3	1	1
1	1	6	0	1
3	6	6	1	2
2	3	0	1	1
4	1	1	2	2
0	3	0	2	4

Solution These data could be treated as 50 individual entries, and you could use the formulas for mean and standard deviation. Because there are so many repeated numbers, however, it is easier to use a frequency distribution.

x	f	xf
0	10	0
1	19	19
2	7	14
3	7	21
4	2	8
5	1	5
6	4	24
	$\Sigma = 50$	$\Sigma = 91$

$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
−1.8	3.24	32.40
−0.8	0.64	12.16
0.2	0.04	0.28
1.2	1.44	10.08
2.2	4.84	9.68
3.2	10.24	10.24
4.2	17.64	70.56
		$\Sigma = 145.40$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{91}{50} \approx 1.8$$

Sample mean

Use the sum of squares to find the sample standard deviation.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{145.4}{49}} \approx 1.7$$

Sample standard deviation

So, the sample mean is about 1.8 children, and the standard deviation is about 1.7 children.

Try It Yourself 9

Change three of the 6s in the data set to 4s. How does this change affect the sample mean and sample standard deviation?

a. Write the first three columns of a *frequency distribution.*
b. Find the *sample mean.*
c. Complete the *last three columns* of the frequency distribution.
d. Find the *sample standard deviation.*

Answer: Page A36

Study Tip

Here are instructions for calculating the sample mean and sample standard deviation on a TI-83/84 for the grouped data in Example 9.

STAT

Choose the EDIT menu.

1: Edit

Enter the values of x into L1.
Enter the frequencies f into L2.

STAT

Choose the CALC menu.

1: 1–Var Stats

ENTER

2nd L1 , 2nd L2

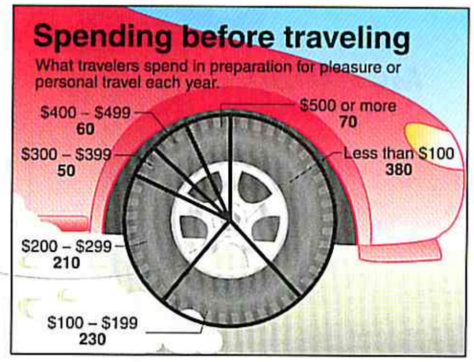
ENTER

When a frequency distribution has classes, you can estimate the sample mean and standard deviation by using the midpoint of each class.

EXAMPLE 10

Using Midpoints of Classes

The circle graph at the right shows the results of a survey in which 1000 adults were asked how much they spend in preparation for personal travel each year. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set. *(Adapted from Travel Industry Association of America)*



Solution Begin by using a frequency distribution to organize the data.

Class	x	f	xf
0–99	49.5	380	18,810
100–199	149.5	230	34,385
200–299	249.5	210	52,395
300–399	349.5	50	17,475
400–499	449.5	60	26,970
500+	599.5	70	41,965
		$\Sigma = 1{,}000$	$\Sigma = 192{,}000$

$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
−142.5	20,306.25	7,716,375.0
−42.5	1,806.25	415,437.5
57.5	3,306.25	694,312.5
157.5	24,806.25	1,240,312.5
257.5	66,306.25	3,978,375.0
407.5	166,056.25	11,623,937.5
		$\Sigma = 25{,}668{,}750.0$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{192{,}000}{1{,}000} = 192$$ Sample mean

Use the sum of squares to find the sample standard deviation.

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{25{,}668{,}750}{999}} \approx 160.3$$ Sample standard deviation

So, the sample mean is \$192 per year, and the sample standard deviation is about \$160.3 per year.

Study Tip

When a class is open, as in the last class, you must assign a single value to represent the midpoint. For this example, we selected 599.5.

▸ Try It Yourself 10

In the frequency distribution, 599.5 was chosen to represent the class of \$500 or more. How would the sample mean and standard deviation change if you used 650 to represent this class?

a. Write the first four columns of a *frequency distribution.*
b. Find the *sample mean.*
c. Complete the *last three columns* of the frequency distribution.
d. Find the *sample standard deviation.*

Answer: Page A36

2.4 EXERCISES

Building Basic Skills and Vocabulary

In Exercises 1 and 2, find the range, mean, variance, and standard deviation of the population data set.

1. 12 9 7 5 7 8 10 4 11 6

2. 15 24 17 19 20 18 20
16 21 23 17 18 22 14

In Exercises 3 and 4, find the range, mean, variance, and standard deviation of the sample data set.

3. 17 8 13 18 15 9 10 11 6

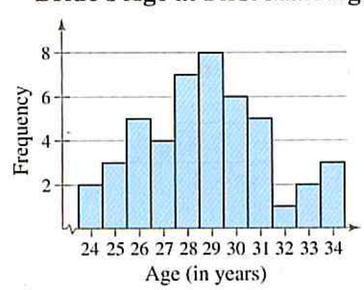
4. 28 25 21 15 7 14 9
27 21 24 14 17 16

Graphical Reasoning *In Exercises 5 and 6, find the range of the data set represented by the display or graph.*

5. Key: 2|3 = 23

2	3 9
3	0 0 2 3 6 7
4	0 1 2 3 3 8
5	0 1 1 9
6	1 2 9 9
7	5 9
8	4 8
9	0 2 5 6

6. **Bride's Age at First Marriage**



7. Explain how to find the range of a data set. What is an advantage of using the range as a measure of variation? What is a disadvantage?

8. Explain how to find the deviation of an entry in a data set. What is the sum of all the deviations in any data set?

9. Why is the standard deviation used more frequently than the variance? (*Hint:* Consider the units of the variance.)

10. Explain the relationship between variance and standard deviation. Can either of these measures be negative? Explain. Find a data set for which $n = 5$, $\bar{x} = 7$, and $s = 0$.

11. Marriage Ages The ages of 10 brides at their first marriage are given below.

31.8 24.5 26.7 21.3 45.6 35.9 22.5 33.1 42.3 30.6

(a) Find the range of the data set.

(b) Change 45.6 to 65.6 and find the range of the new data set.

(c) Compare your answer to part (a) with your answer to part (b).

12. Find a population data set that contains six entries, has a mean of 5, and has a standard deviation of 2.

Using and Interpreting Concepts

13. Graphical Reasoning Both data sets have a mean of 165. One has a standard deviation of 16, and the other has a standard deviation of 24. Which is which? Explain your reasoning.

(a) Key: 12|8 = 128

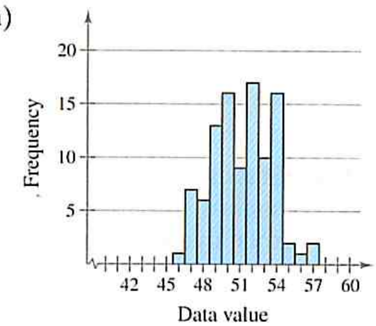
12	8 9
13	5 5 8
14	1 2
15	0 0 6 7
16	4 5 9
17	1 3 6 8
18	0 8 9
19	6
20	3 5 7

(b)

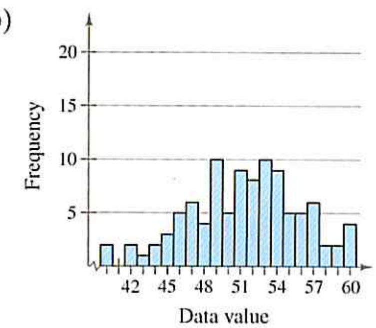
12	
13	1
14	2 3 5
15	0 4 5 6 8
16	1 1 2 3 3 3
17	1 5 8 8
18	2 3 4 5
19	0 2
20	

14. Graphical Reasoning Both data sets represented below have a mean of 50. One has a standard deviation of 2.4, and the other has a standard deviation of 5. Which is which? Explain your reasoning.

(a)



(b)



15. Writing Describe the difference between the calculation of population standard deviation and that of sample standard deviation.

16. Writing Given a data set, how do you know whether to calculate σ or s?

17. Salary Offers You are applying for a job at two companies. Company A offers starting salaries with μ = \$31,000 and σ = \$1000. Company B offers starting salaries with μ = \$31,000 and σ = \$5000. From which company are you more likely to get an offer of \$33,000 or more?

18. Golf Strokes An Internet site compares the strokes per round of two professional golfers. Which golfer is more consistent: Player A with $\mu = 71.5$ strokes and $\sigma = 2.3$ strokes, or Player B with $\mu = 70.1$ strokes and $\sigma = 1.2$ strokes?

Comparing Two Data Sets *In Exercises 19–22, you are asked to compare two data sets and interpret the results.*

19. Annual Salaries Sample annual salaries (in thousands of dollars) for municipal employees in Los Angeles and Long Beach are listed.

Los Angeles:	20.2	26.1	20.9	32.1	35.9	23.0	28.2	31.6	18.3
Long Beach:	20.9	18.2	20.8	21.1	26.5	26.9	24.2	25.1	22.2

(a) Find the range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

20. Annual Salaries Sample annual salaries (in thousands of dollars) for municipal employees in Dallas and Houston are listed.

Dallas:	34.9	25.7	17.3	16.8	26.8	24.7	29.4	32.7	25.5
Houston:	25.6	23.2	26.7	27.7	25.4	26.4	18.3	26.1	31.3

(a) Find the range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

21. SAT Scores Sample SAT scores for eight males and eight females are listed.

Male SAT scores:	1059	1328	1175	1123	923	1017	1214	1042
Female SAT scores:	1226	965	841	1053	1056	1393	1312	1222

(a) Find the range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

22. Annual Salaries Sample annual salaries (in thousands of dollars) for public and private elementary school teachers are listed.

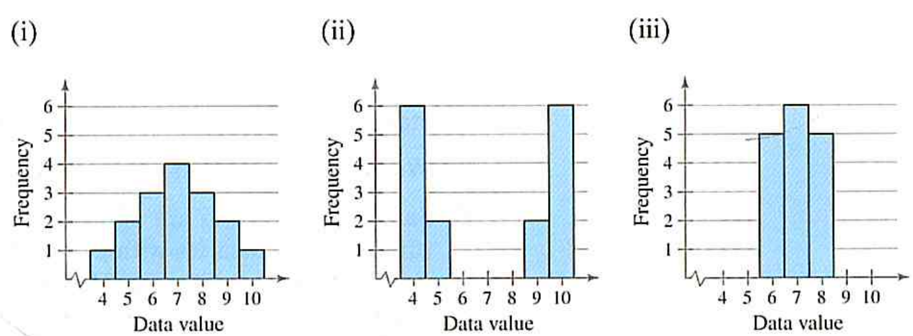
Public teachers:	38.6	38.1	38.7	36.8	34.8	35.9	39.9	36.2
Private teachers:	21.8	18.4	20.3	17.6	19.7	18.3	19.4	20.8

(a) Find the range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

Reasoning with Graphs *In Exercises 23–26, you are asked to compare three data sets.*

23. (a) Without calculating, determine which data set has the greatest sample standard deviation. Which has the least sample standard deviation? Explain your reasoning.



(b) How are the data sets the same? How do they differ?

24. (a) Without calculating, determine which data set has the greatest sample standard deviation. Which has the least sample standard deviation? Explain your reasoning.

(i)

0	9
1	5 8
2	3 3 7 7
3	2 5
4	1

Key: $4|1 = 41$

(ii)

0	9
1	5
2	3 3 3 7 7 7
3	5
4	1

Key: $4|1 = 41$

(iii)

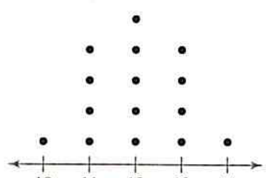
0	
1	5
2	3 3 3 3 7 7 7 7
3	5
4	

Key: $4|1 = 41$

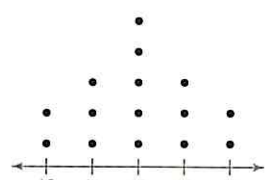
(b) How are the data sets the same? How do they differ?

25. (a) Without calculating, determine which data set has the greatest sample standard deviation. Which has the least sample standard deviation? Explain your reasoning.

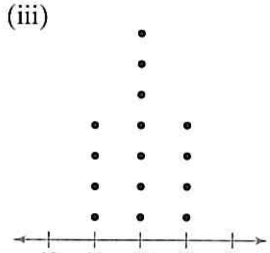
(i)



(ii)



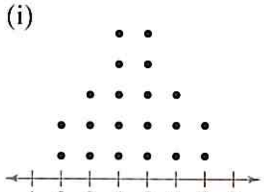
(iii)



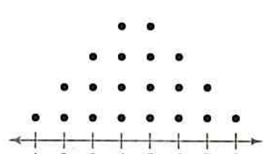
(b) How are the data sets the same? How do they differ?

26. (a) Without calculating, determine which data set has the greatest sample standard deviation. Which has the least sample standard deviation? Explain your reasoning.

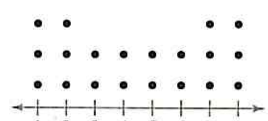
(i)



(ii)



(iii)



(b) How are the data sets the same? How do they differ?

27. Writing Discuss the similarities and the differences between the Empirical Rule and Chebychev's Theorem.

28. Writing What must you know about a data set before you can use the Empirical Rule?

Using the Empirical Rule *In Exercises 29–34, you are asked to use the Empirical Rule.*

29. The mean value of land and buildings per acre from a sample of farms is \$1500, with a standard deviation of \$200. The data set has a bell-shaped distribution. Estimate the percent of farms whose land and building values per acre are between \$1300 and \$1700.

30. The mean value of land and buildings per acre from a sample of farms is \$2400, with a standard deviation of \$450. Between what two values do about 95% of the data lie? (Assume the data set has a bell-shaped distribution.)

31. Using the sample statistics from Exercise 29, do the following. (Assume the number of farms in the sample is 75.)

(a) Use the Empirical Rule to estimate the number of farms whose land and building values per acre are between \$1300 and \$1700.

(b) If 25 additional farms were sampled, about how many of these farms would you expect to have land and building values between \$1300 per acre and \$1700 per acre?

32. Using the sample statistics from Exercise 30, do the following. (Assume the number of farms in the sample is 40.)

(a) Use the Empirical Rule to estimate the number of farms whose land and building values per acre are between \$1500 and \$3300.

(b) If 20 additional farms were sampled, about how many of these farms would you expect to have land and building values between \$1500 per acre and \$3300 per acre?

33. Using the sample statistics from Exercise 29 and the Empirical Rule, determine which of the following farms, whose land and building values per acre are given, are outliers (more than two standard deviations from the mean).

\$1150, \$1775, \$1000, \$1475, \$2000, \$1850

34. Using the sample statistics from Exercise 30 and the Empirical Rule, determine which of the following farms, whose land and building values per acre are given, are outliers (more than two standard deviations from the mean).

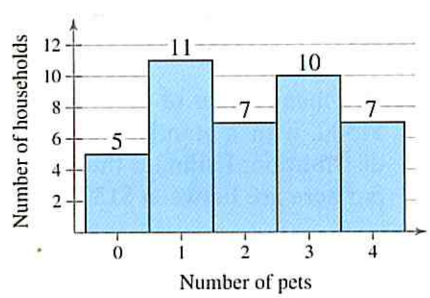
\$3325, \$2450, \$3200, \$1490, \$1675, \$2950

35. Chebychev's Theorem Old Faithful is a famous geyser at Yellowstone National Park. From a sample with $n = 32$, the mean duration of Old Faithful's eruptions is 3.32 minutes and the standard deviation is 1.09 minutes. Using Chebychev's Theorem, determine at least how many of the eruptions lasted between 1.14 minutes and 5.5 minutes. *(Source: Yellowstone National Park)*

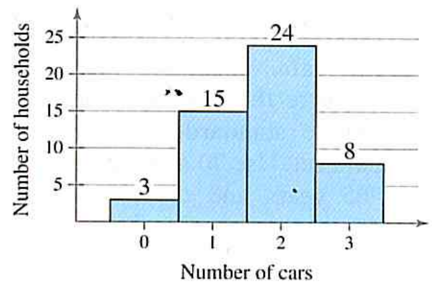
36. Chebychev's Theorem The mean time in a women's 400-meter dash is 57.07 seconds, with a standard deviation of 1.05. Apply Chebychev's Theorem to the data using $k = 2$. Interpret the results.

Calculating Using Grouped Data *In Exercises 37–44, use the grouped data formulas to find the indicated mean and standard deviation.*

37. Pets per Household The results of a random sample of the number of pets per household in a region are shown in the histogram. Estimate the sample mean and the sample standard deviation of the data set.



38. Cars per Household A random sample of households in a region and the number of cars per household are shown in the histogram. Estimate the sample mean and the sample deviation of the data set.



39. Football Wins The number of wins for each National Football League team in 2006 are listed. Make a frequency distribution (using five classes) for the data set. Then approximate the population mean and the population standard deviation of the data set. *(Source: National Football League)*

12	10	7	6	13	8	8	4	12	8	8	6
14	9	9	2	10	9	8	5	13	8	6	3
10	8	7	4	9	8	7	5				

40. Water Consumption The number of gallons of water consumed per day by a small village are listed. Make a frequency distribution (using five classes) for the data set. Then approximate the population mean and the population standard deviation of the data set.

167	180	192	173	145	151	174
175	178	160	195	224	244	146
162	146	177	163	149	188	

41. Amount of Caffeine The amount of caffeine in a sample of five-ounce servings of brewed coffee is shown in the histogram. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set.

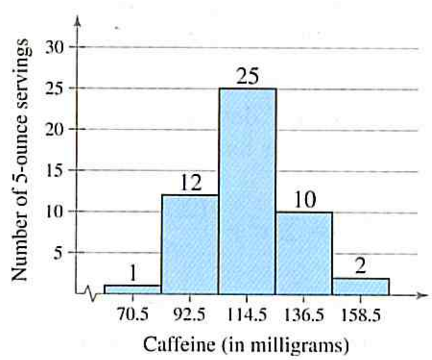


FIGURE FOR EXERCISE 41

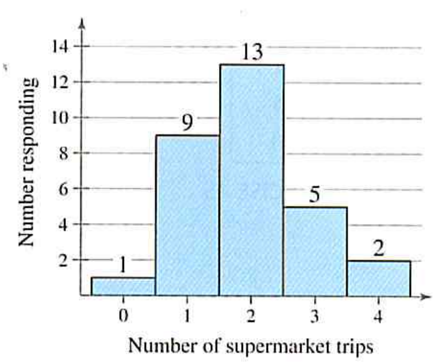
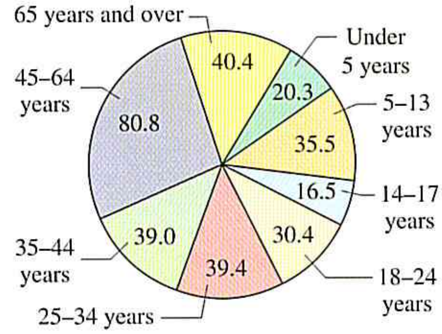


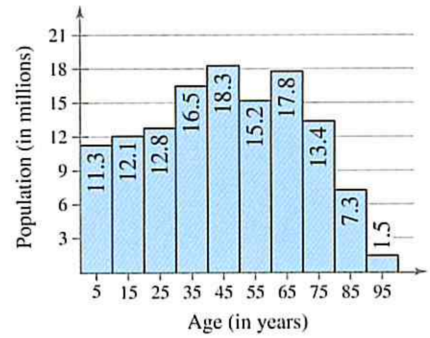
FIGURE FOR EXERCISE 42

42. Supermarket Trips Thirty people were randomly selected and asked how many trips to the supermarket they made in the past week. The responses are shown in the histogram. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set.

43. U.S. Population The estimated distribution (in millions) of the U.S. population by age for the year 2011 is shown in the circle graph. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set. Use 70 as the midpoint for "65 years and over." *(Source: U.S. Census Bureau)*



44. Japan's Population Japan's estimated population for the year 2014 is shown in the bar graph. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set. *(Source: U.S. Census Bureau, International Data Base)*



Extending Concepts

45. Coefficient of Variation The **coefficient of variation** CV describes the standard deviation as a percent of the mean. Because it has no units, you can use the coefficient of variation to compare data with different units.

$$CV = \frac{\text{Standard deviation}}{\text{Mean}} \times 100\%$$

The table to the left shows the heights (in inches) and weights (in pounds) of the members of a basketball team. Find the coefficient of variation for each data set. What can you conclude?

Heights	Weights
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

TABLE FOR EXERCISE 45

46. Shortcut Formula You used $SS_x = \Sigma(x - \bar{x})^2$ when calculating variance and standard deviation. An alternative formula that is sometimes more convenient for hand calculations is

$$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n}.$$

You can find the sample variance by dividing the sum of squares by $n - 1$ and the sample standard deviation by finding the square root of the sample variance.

(a) Use the shortcut formula to calculate the sample standard deviation for the data set given in Exercise 21.

(b) Compare your results with those obtained in Exercise 21.

47. Scaling Data Sample annual salaries (in thousands of dollars) for employees at a company are listed.

42 36 48 51 39 39 42 36 48 33 39 42 45

(a) Find the sample mean and sample standard deviation.

(b) Each employee in the sample is given a 5% raise. Find the sample mean and sample standard deviation for the revised data set.

(c) To calculate the monthly salary, divide each original salary by 12. Find the sample mean and sample standard deviation for the revised data set.

(d) What can you conclude from the results of (a), (b), and (c)?

48. Shifting Data Sample annual salaries (in thousands of dollars) for employees at a company are listed.

40 35 49 53 38 39 40 37 49 34 38 43 47

(a) Find the sample mean and sample standard deviation.

(b) Each employee in the sample is given a \$1000 raise. Find the sample mean and sample standard deviation for the revised data set.

(c) Each employee in the sample takes a pay cut of \$2000 from their original salary. Find the sample mean and sample standard deviation for the revised data set.

(d) What can you conclude from the results of (a), (b), and (c)?

49. Mean Absolute Deviation Another useful measure of variation for a data set is the **mean absolute deviation** MAD. It is calculated by the formula

$$\frac{\Sigma|x - \bar{x}|}{n}.$$

(a) Find the mean absolute deviations of the data sets in Exercise 21. Compare your results with the sample standard deviation.

(b) Find the mean absolute deviations of the data sets in Exercise 22. Compare your results with the sample standard deviation.

50. Chebychev's Theorem At least 99% of the data in any data set lie within how many standard deviations of the mean? Explain how you obtained your answer.

51. Pearson's Index of Skewness The English statistician Karl Pearson (1857–1936) introduced a formula for the skewness of a distribution.

$$P = \frac{3(\bar{x} - \text{median})}{s}$$ Pearson's index of skewness

Most distributions have an index of skewness between -3 and 3. When $P > 0$, the data are skewed right. When $P < 0$, the data are skewed left. When $P = 0$, the data are symmetric. Calculate the coefficient of skewness for each distribution. Describe the shape of each.

(a) $\bar{x} = 17, s = 2.3$, median $= 19$

(b) $\bar{x} = 32, s = 5.1$, median $= 25$

CASE STUDY

Earnings of Athletes

The earnings of professional athletes in different sports can vary. An athlete can be paid a base salary, earn signing bonuses upon signing a new contract, or even earn money by finishing in a certain position in a race or tournament. The data shown below are the earnings (for performance only, no endorsements) from Major League Baseball (MLB), Major League Soccer (MLS), the National Basketball Association (NBA), the National Football League (NFL), the National Hockey League (NHL), the National Association for Stock Car Auto Racing (NASCAR), and the Professional Golf Association Tour (PGA) for a recent year.

Organization	Number of players
MLB	824
MLS	321
NBA	444
NFL	1877
NHL	727
NASCAR	78
PGA	263

Number of Players Separated into Earnings Ranges

Organization	\$0–\$500,000	\$500,001–\$2,000,000	\$2,000,001–\$6,000,000	\$6,000,001–\$10,000,000	\$10,000,001+
MLB	299	207	189	73	56
MLS	316	5	0	0	0
NBA	31	166	147	58	42
NFL	760	758	274	70	15
NHL	85	448	177	17	0
NASCAR	27	13	33	5	0
PGA	115	117	29	2	0

Exercises

1. **Revenue** Which organization had the greatest total player earnings? Explain your reasoning.
2. **Mean Earnings** Estimate the mean earnings of a player in each organization. Use \$16,500,000 as the midpoint for \$10,000,001+.
3. **Revenue** Which organization had the greatest earnings per player? Explain your reasoning.
4. **Standard Deviation** Estimate the standard deviation for the earnings of a player in each organization. Use \$16,500,000 as the midpoint for \$10,000,001+.
5. **Standard Deviation** Which organization had the greatest standard deviation? Explain your reasoning.
6. **Bell-Shaped Distribution** Of the seven organizations, which is more bell shaped? Explain your reasoning.