Management *U.S. wheat prices and production figures for a recent decade are given below.*

Year	*Price ($ per bushel)*	*Production (millions of bushels)*
1982	2.70	2200
1983	2.30	2000
1984	2.95	1750
1985	3.80	2200
1986	3.90	2400
1987	3.60	2800
1988	3.55	2800
1989	3.50	2450
1990	3.35	2600
1991	3.20	2750

Find the mean for each of the following.

41. Price per bushel of wheat

42. Wheat production

43. Social Science The table shows the median age at first marriage in the U.S. over the last 50 years.*

Year	*Male*	*Female*
1990	26.1	23.9
1980	24.7	22.0
1970	23.2	20.8
1960	22.8	20.3
1950	22.8	20.3
1940	24.3	21.5

(a) Find the mean of the median ages for males.
(b) Find the mean of the median ages for females.

*U.S. National Center for Health Statistics of the United States, *Vital Statistics of the United States*, annual, *Monthly Vital Statistics Report*.

44. Social Science The number of nations participating in the winter Olympic games, from 1968 to 1992, is given below.*

Year	*Nations Participating*
1968	37
1972	35
1976	37
1980	37
1984	49
1988	57
1992	64

Find the following measures for the data.
(a) Mean (b) Median (c) Mode
(d) Which of these measures best represents the data? Explain your reasoning.

45. Social Science Washington Post writer John Schwartz pointed out that if Microsoft Corp. cofounder Bill Gates, who was reportedly worth $10 billion in 1995, lived in a town with 10,000 totally penniless people, the average personal wealth in the town would make it seem as if everyone were a millionaire.†
(a) Verify Schwartz' statement.
(b) What would be the median personal wealth in this town?
(c) What would be the mode for the personal wealth in this town?
(d) In this example, which average is most representative: the mean, the median, or the mode?

46. According to an article in the *Chance* electronic newsletter, the mean salary for National and American League baseball players in 1994 was $1,183,416, while the median salary was $500,000.‡ What might explain the large discrepancy between the mean and the median?

**The Universal Almanac*, John C. Wright, General Editor, Andrews and McMeel, Kansas City, p. 682.

†John Schwartz, "Mean Statistics: When is Average Best?" *The Washington Post*, Jan. 11, 1995, p. H7.

‡*Chance*, Aug. 13, 1994.

11.2 MEASURES OF VARIATION

The mean gives a measure of central tendency of a list of numbers, but tells nothing about the *spread* of the numbers in the list. For example, look at the following three samples.

I	3	5	6	3	3
II	4	4	4	4	4
III	10	1	0	0	9

Each of these three samples has a mean of 4, and yet they are quite different; the amount of dispersion or variation within the samples is different. Therefore, in addition to a measure of central tendency, another kind of measure is needed that describes how much the numbers vary.

The largest number in sample I is 6, while the smallest is 3, a difference of 3. In sample II this difference is 0; in sample III it is 10. The difference between the largest and smallest number in a sample is called the **range,** one example of a measure of variation. The range of sample I is 3, of sample II is 0, and of sample III is 10. The range has the advantage of being very easy to compute and gives a rough estimate of the variation among the data in the sample. However, it depends only on the two extremes and tells nothing about how the other data are distributed between the extremes.

EXAMPLE 1 Find the range for each list of numbers.

(a) 12, 27, 6, 19, 38, 9, 42, 15

The highest number here is 42; the lowest is 6. The range is the difference of these numbers, or

$$42 - 6 = 36.$$

(b) 74, 112, 59, 88, 200, 73, 92, 175

$$\text{Range} = 200 - 59 = 141$$ ■ [1]

[1] Find the range for the numbers 159, 283, 490, 390, 375, 297.

Answer:
331

TECHNOLOGY TIP Many graphing calculators list the largest and smallest numbers in a list when displaying one-variable statistics, usually on the second screen of the display. ✔

The most useful measure of variation is the *standard deviation.* Before defining it, however, we must find the **deviations from the mean,** the differences found by subtracting the mean from each number in a distribution.

EXAMPLE 2 Find the deviations from the mean for the numbers

32, 41, 47, 53, 57.

Adding these numbers and dividing by 5 gives a mean of 46. To find the deviations from the mean, subtract 46 from each number in the list. For example, the first deviation from the mean is $32 - 46 = -14$; the last is $57 - 46 = 11$.

Number	*Deviation from Mean*
32	−14
41	−5
47	1
53	7
57	11

To check your work, find the sum of these deviations. It should always equal 0. (The answer is always 0 because the positive and negative numbers cancel each other.) ■ [2]

[2] Find the deviations from the mean for each set of numbers.

(a) 19, 25, 36, 41, 52, 61

(b) 6, 9, 5, 11, 3, 2

Answers:

(a) Mean is 39; deviations are −20, −14, −3, 2, 13, 22.

(b) Mean is 6; deviations from the mean are 0, 3, −1, 5, −3, −4.

To find a measure of variation, we might be tempted to use the mean of the deviations. However, as mentioned above, this number is always 0, no matter how widely the data are dispersed. To avoid the problem of the positive and negative deviations averaging 0, statisticians square each deviation (producing a list of nonnegative numbers) and then find the mean.

Number	*Deviation from Mean*	*Square of Deviation*
32	−14	196
41	−5	25
47	1	1
53	7	49
57	11	121

In this case the mean of the squared deviations is

$$\frac{196 + 25 + 1 + 49 + 121}{5} = \frac{392}{5} = 78.4.$$

This number is called the **population variance,** because the sum was divided by $n = 5$, the number of items in the original list.

In most applications, however, it isn't practical to use the entire population, so a sample is used instead to estimate the variance and other measures. Since a random sample may not include extreme entries from the list, statisticians prefer to divide the sum of the squared deviations in a sample by $n - 1$, rather than n, to create what is called an *unbiased estimator.* The informal idea behind this is that using $n - 1$ increases the value of the variance, as would be the case if extreme entries were used. Using $n - 1$ in the distribution above gives

$$\frac{196 + 25 + 1 + 49 + 121}{5 - 1} = \frac{392}{4} = 98.$$

This number 98 is called the **sample variance** of the distribution and is denoted s^2 because it is found by averaging a list of squares. In this case, the population and sample variances differ by quite a bit. But when n is relatively large, as is the case in real-life applications, the difference between them is rather small.

SAMPLE VARIANCE

The variance of a sample of n numbers $x_1, x_2, x_3, \ldots, x_n$, with mean $\bar{x}$, is

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}.$$

When computing the sample variance by hand, it is often convenient to use the following shortcut formula, which can be derived algebraically from the definition in the box above.

$$s^2 = \frac{\Sigma x^2 - n\bar{x}^2}{n - 1}$$

To find the sample variance, we square the deviations from the mean, so the variance is in squared units. To return to the same units as the data, we use the *square root* of the variance, called the **sample standard deviation,** denoted s.

SAMPLE STANDARD DEVIATION

The standard deviation of a sample of n numbers $x_1, x_2, x_3, \ldots, x_n$, with mean $\bar{x}$, is

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}.$$

Similarly, the **population standard deviation** is the square root of the population variance (just replace $n - 1$ by n in the denominator in the box above).

TECHNOLOGY TIP When a graphing calculator computes one-variable statistics for a list of data, it usually displays the following information (not necessarily in this order and sometimes on two screens), and possibly other information as well.

Information	*Notation*
Number of data entries	n or $N\Sigma$
Mean	$\bar{x}$ or mean Σ
Sum of all data entries	Σx or TOT Σ
Sum of the squares of all data entries	Σx^2
Sample standard deviation	Sx or sx or $x\sigma_{n-1}$ or SSDEV
Population standard deviation	σx or $x\sigma_n$ or PSDEV
Largest / smallest data entries	maxX / minX or MAXΣ/MINΣ
Median	Med or MEDIAN ✔

NOTE Hereafter we shall deal exclusively with the sample variance and the sample standard deviation. So whenever standard deviation is mentioned, it means "sample standard deviation." ◆

As its name indicates, the standard deviation is the most commonly used measure of variation. The standard deviation is a measure of the variation from the mean. The size of the standard deviation indicates how spread out the data are from the mean.

EXAMPLE 3 Find the standard deviation of the numbers

7, 9, 18, 22, 27, 29, 32, 40

by hand, using the shortcut variance formula at the bottom of page 518.

Arrange the work in columns, as shown in the table.

Number	*Square of the Number*
7	49
9	81
18	324
22	484
27	729
29	841
32	1024
40	1600
184	5132

Now find the mean.

$$\bar{x} = \frac{\Sigma x}{8} = \frac{184}{8} = 23$$

The total of the second column gives $\Sigma x^2 = 5132$. The variance is

$$s^2 = \frac{\Sigma x^2 - n\bar{x}^2}{n - 1} = \frac{5132 - 8(23)^2}{8 - 1} = 128.6 \quad \text{(rounded)},$$

and the standard deviation is

$$s \approx \sqrt{128.6} \approx 11.3.$$

Use your calculator to verify this result. ■ [3]

[3] Find the standard deviation of each set of numbers. The deviations from the mean were found in Problem 2 at the side.

(a) 19, 25, 36, 41, 52, 61

(b) 6, 9, 5, 11, 3, 2

Answers:

(a) 15.9

(b) 3.5

One way to interpret the standard deviation uses the fact that, for many populations, most of the data are within three standard deviations of the mean. (See Section 11.3.) This implies that, in Example 3, most of the population from which this sample is taken is between

$$\bar{x} - 3s = 23 - 3(11.3) = -10.9$$

and

$$\bar{x} + 3s = 23 + 3(11.3) = 56.9.$$

This has important implications for quality control. If the sample in Example 3 represents measurements of a product that the manufacturer wants to be between 5 and 45, the standard deviation is too large, even though all the numbers are within these bounds.

For data in a grouped frequency distribution, a slightly different formula for the standard deviation is used.

STANDARD DEVIATION FOR A GROUPED DISTRIBUTION

The standard deviation for a distribution with mean $\bar{x}$, where x is an interval midpoint with frequency f, and $n = \Sigma f$, is

$$s = \sqrt{\frac{\Sigma fx^2 - n\bar{x}^2}{n-1}}.$$

The formula indicates that the product fx^2 is to be found for each interval. Then these products are summed, n times the square of the mean is subtracted, and the difference is divided by one less than the total frequency; that is, by $n - 1$. The square root of this result is s, the standard deviation. The standard deviation found by this formula may (probably will) differ somewhat from the standard deviation found from the original data.

CAUTION In calculating the standard deviation for either a grouped or ungrouped distribution, using a rounded value for the mean or variance may produce an inaccurate value. ◆

EXAMPLE 4 Find s for the grouped data of Example 5, Section 11.1. Begin by including columns for x^2 and fx^2 in the table.

Interval	f	x	x^2	fx^2
40–49	2	44.5	1980.25	3,960.50
50–59	4	54.5	2970.25	11,881.00
60–69	7	64.5	4160.25	29,121.75
70–79	9	74.5	5550.25	49,952.25
80–89	5	84.5	7140.25	35,701.25
90–99	3	94.5	8930.25	26,790.75
	Total: 30			Total: 157,407.50

[4] Find the standard deviation for the grouped data that follows. (*Hint:* $\bar{x} = 28.5$)

Value	*Frequency*
20–24	3
25–29	2
30–34	4
35–39	1

Answer:
5.3

Recall from Section 11.1 that $\bar{x} = 71.2$. Use the formula above with $n = 30$ to find s.

$$s = \sqrt{\frac{\Sigma fx^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{157{,}407.50 - 30(71.2)^2}{30 - 1}}$$

$$\approx 13.5$$ ■ [4]

NOTE A calculator is almost a necessity for finding a standard deviation. With a non-graphing calculator, a good procedure to follow is to first calculate $\bar{x}$. Then for each x, square that number, then multiply the result by the appropriate frequency. If your calculator has a key that accumulates a sum, use it to accumulate the total in the last column of the table. With a graphing calculator, simply enter the midpoints and the frequencies, then ask for the 1-variable statistics. ◆

11.2 EXERCISES

1. How are the variance and the standard deviation related?
2. Why can't we use the sum of the deviations from the mean as a measure of dispersion of a distribution?

Find the range and standard deviation for each of the following sets of numbers. (See Examples 1 and 3.)

3. 6, 8, 9, 10, 12
4. 12, 15, 19, 23, 26
5. 7, 6, 12, 14, 18, 15
6. 4, 3, 8, 9, 7, 10, 1
7. 42, 38, 29, 74, 82, 71, 35
8. 122, 132, 141, 158, 162, 169, 180
9. 241, 248, 251, 257, 252, 287
10. 51, 58, 62, 64, 67, 71, 74, 78, 82, 93

Find the standard deviation for the grouped data in Exercises 11 and 12. (See Example 4.)

11. (From Exercise 1, Section 11.1)

College Units	*Frequency*
0–24	4
25–49	3
50–74	6
75–99	3
100–124	5
125–149	9

12. (From Exercise 2, Section 11.1)

Scores	*Frequency*
30–39	1
40–49	6
50–59	13
60–69	22
70–79	17
80–89	13
90–99	8

13. **Natural Science** Twenty-five laboratory rats, used in an experiment to test the food value of a new product, made the following weight gains in grams.

5.25	5.03	4.90	4.97	5.03
5.12	5.08	5.15	5.20	4.95
4.90	5.00	5.13	5.18	5.18
5.22	5.04	5.09	5.10	5.11
5.23	5.22	5.19	4.99	4.93

Find the mean gain and the standard deviation of the gains.

14. **Management** An assembly-line machine turns out washers with the following thicknesses (in millimeters).

1.20	1.01	1.25	2.20	2.58	2.19	1.29	1.15
2.05	1.46	1.90	2.03	2.13	1.86	1.65	2.27
1.64	2.19	2.25	2.08	1.96	1.83	1.17	2.24

Find the mean and standard deviation of these thicknesses.

An application of standard deviation is given by Chebyshev's theorem. *(P. L. Chebyshev was a Russian mathematician who lived from 1821 to 1894.) This theorem applies to any distribution of numerical data. It states:*

> *For any distribution of numerical data, at least $1 - 1/k^2$ of the numbers lie within k standard deviations of the mean.*

Example *For any distribution, at least*

$$1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9}$$

of the numbers lie within 3 standard deviations of the mean. Find the fraction of all the numbers of a data set lying within the following numbers of standard deviations from the mean.

15. 2
16. 4
17. 5

In a certain distribution of numbers, the mean is 50 with a standard deviation of 6. Use Chebyshev's theorem to tell what percent of the numbers are

18. between 32 and 68;
19. between 26 and 74;
20. less than 38 or more than 62;
21. less than 32 or more than 68;
22. less than 26 or more than 74.
23. **Management** The Britelite Company conducted tests on the life of its light bulbs and those of a competitor (Brand X) with the following results for samples of 10 bulbs of each brand.

	Hours of Use (in 100s)									
Britelite	20	22	22	25	26	27	27	28	30	35
Brand X	15	18	19	23	25	25	28	30	34	38

Compute the mean and standard deviation for each sample. Compare the means and standard deviations of the two brands and then answer the questions below.

(a) Which bulbs have a more uniform life in hours?

(b) Which bulbs have the highest average life in hours?

24. **Management** The weekly wages of the six employees of Harold's Hardware Store are \$300, \$320, \$380, \$420, \$500, and \$2000.

(a) Find the mean and standard deviation of this distribution.

(b) How many of the employees earn within one standard deviation of the mean? How many earn within two standard deviations of the mean?

25. **Social Science** The number of unemployed workers in the United States in 1988–1993 (in millions) is given below.*

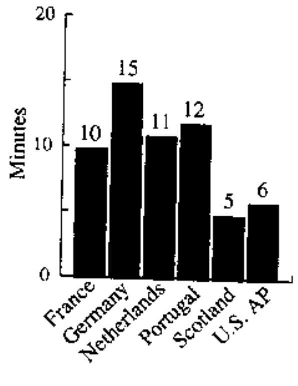
Year	*Number Unemployed*
1988	6.70
1989	6.53
1990	6.87
1991	8.43
1992	9.38
1993	8.73

(a) Find the mean number unemployed (in millions) in this period. Which year has unemployment closest to the mean?

(b) Find the standard deviation for the data.

(c) In how many of these years is unemployment within 1 standard deviation of the mean?

26. **Social Science** In an article comparing national mathematics examinations in the U.S. and some European countries, a researcher found the results shown in the histogram for the number of minutes allowed for each open-ended question.†

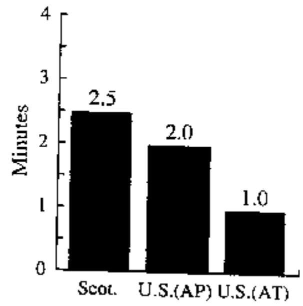


(a) Find the mean and standard deviation. (*Hint:* Do not use the rounded value of the mean to find the standard deviation.)

(b) How many standard deviations from the mean is the largest number of minutes?

(c) How many standard deviations from the mean is the U.S. AP examination?

27. **Social Science** For all questions, the researcher in Exercise 26 found the following results for Scotland and the United States, the countries with the least rigorous examinations.

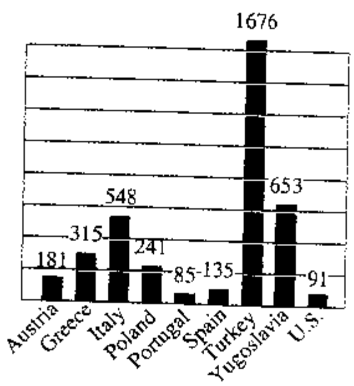


(a) Find the mean and standard deviation.

(b) How many standard deviations from the mean is the AT (achievement test)?

(c) How many standard deviations from the mean is the Scottish test?

28. **Social Science** Recently Germany has endured rioting and disruption because of the increasing numbers of immigrants. The nine top countries of origin of German immigrants are shown in the histogram.



(a) Find the mean and standard deviation of these data.

(b) How many standard deviations from the mean is the largest number of immigrants? the smallest? Which country of origin is closest to the mean?

*U.S. Bureau of Labor Statistics, Bulletin 2307; and *Employment and Earnings*, Monthly.

†Information from article comparing national mathematics examinations in the U.S. and some European countries as appeared in FOCUS, June 1993. Reprinted by permission of the Mathematical Association of America.

29. **Social Science** The numbers of immigrants to the United States (in thousands) from selected parts of the world in 1990 are shown in the table.*

Region	*Immigrants*
Europe	145.4
Asia	357.0
Canada	15.2
Mexico	213.8
Other America	210.3

(a) Find the mean number of immigrants from these regions. Which region produces the number of immigrants closest to the mean?

(b) Find the standard deviation. Are any of the data more than 2 standard deviations from the mean? If so, which ones? Are any of the data more than 1 standard deviation from the mean? If so, which ones?

30. **Management** The Quaker Oaks Company conducted a survey to determine if a proposed premium, to be included in their cereal, was appealing enough to generate new sales.† Four cities were used as test markets, where the cereal was distributed with the premium, and four cities as control markets, where the cereal was distributed without the premium. The eight cities were chosen on the basis of their similarity in terms of population, per capita income, and total cereal purchase volume. The results were as follows.

		Percent Change in Average Market Shares per Month
Test Cities	1	+18
	2	+15
	3	+7
	4	+10
Control Cities	1	+1
	2	−8
	3	−5
	4	0

(a) Find the mean of the change in market share for the four test cities.

(b) Find the mean of the change in market share for the four control cities.

(c) Find the standard deviation of the change in market share for the test cities.

(d) Find the standard deviation of the change in market share for the control cities.

(e) Find the difference between the mean of part (a) and the mean of part (b). This represents the estimate of the percent change in sales due to the premium.

(f) The two standard deviations from part (c) and part (d) were used to calculate an "error" of ± 7.95 for the estimate in part (e). With this amount of error what is the smallest and largest estimate of the increase in sales? (*Hint:* Use the answer to part (e).)

On the basis of the results of Exercise 30, the company decided to mass produce the premium and distribute it nationally.

31. **Management** The following table gives 10 samples, of three measurements, made during a production run.

SAMPLE NUMBER									
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
2	3	−2	−3	−1	3	0	−1	2	0
−2	−1	0	1	2	2	1	2	3	0
1	4	1	2	4	2	2	3	2	2

(a) Find the mean $\bar{x}$ for each sample of three measurements.

(b) Find the standard deviation s for each sample of three measurements.

(c) Find the mean $\overline{X}$ of the sample means.

(d) Find the mean $\bar{s}$ of the sample standard deviations.

(e) The upper and lower control limits of the sample means here are $\overline{X} \pm 1.954\bar{s}$. Find these limits. If any of the measurements are outside these limits, the process is out of control. Decide if this production process is out of control.

32. Discuss what the standard deviation tells us about a distribution.

Social Science *The reading scores of a second-grade class given individualized instruction are shown below. The table also shows the reading scores of a second-grade class given traditional instruction in the same school.*

Scores	*Individualized Instruction*	*Traditional Instruction*
50–59	2	5
60–69	4	8
70–79	7	8
80–89	9	7
90–99	8	6

33. Find the mean and standard deviation for the individualized instruction scores.

*Source: U.S. Immigration and Naturalization Service, *Statistics Yearbook*, annual; and releases.

†This example was supplied by Jeffrey S. Berman, Senior Analyst, Marketing Information, Quaker Oats Company.

. Find the mean and standard deviation for the traditional instruction scores.

. Discuss a possible interpretation of the differences in the means and the standard deviations in Exercises 33 and 34.

. **Social Science** In 1998, nineteen state governors earned $100,000 or more (not counting expense allowances) as listed in the following table. (Salaries are given in thousands of dollars and are rounded to the nearest $1000.)*

(a) Find the mean salary of these governors. Which state has the governor with the salary closest to the mean?

(b) Find the standard deviation for the data.

(c) What percent of the governors have salaries within 1 standard deviation of the mean?

(d) What percent of the governors have salaries within 2 standard deviations of the mean?

State	*Salary*
California	131
Delaware	107
Florida	111
Georgia	103
Illinois	123
Maryland	120
Massachusetts	100
Michigan	127
Minnesota	115
Missouri	107
New York	130
North Carolina	107
Ohio	116
Pennsylvania	125
South Carolina	106
Texas	115
Virginia	110
Washington	121
Wisconsin	102

orld Almanac and Book of Facts, 1998.

NORMAL DISTRIBUTIONS

The general idea of a probability distribution was first discussed in Section 10.5. In this section, we consider an important specific type of probability distribution.

Figure 11.3(a) shows a histogram and frequency polygon for the bank transaction example in Section 10.5. The heights of the bars are the probabilities. The transaction times in the example were given to the nearest minute. Theoretically, at least, they could have been timed to the nearest tenth of a minute, or hundredth of a minute, or even more precisely. In each case, a histogram and frequency polygon could be drawn. If the times are measured with smaller and smaller units, there are more bars in the histogram, and the frequency polygon begins to look more and more like the curve in Figure 11.3(b) instead of a polygon. Actually it is possible for the transaction times to take on any real number value greater than 0. A distribution in which the outcomes can take any real number value within some interval is a **continuous distribution.** The graph of a continuous distribution is a curve.

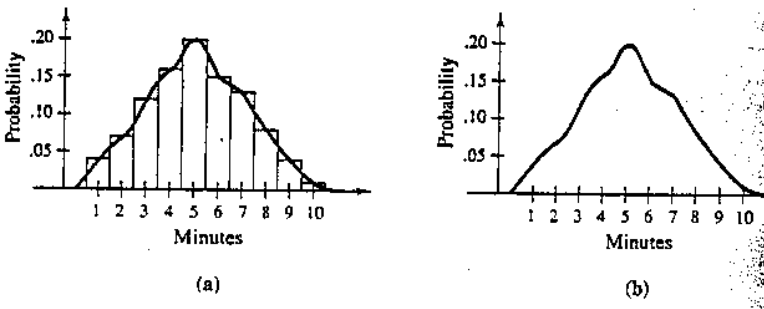


FIGURE 11.3