Multiple formulas: Statisticians have multiple formulas for convenience—that is, convenience relative to the situation. The following statements will help you decide which formula to use:

1. When you are working on a computer and using statistical software, you will generally store all the data values first. The computer handles repeated operations easily and can "revisit" the stored data as often as necessary to complete a procedure. The computations for sample variance will be done using formula (2.6), following the process shown in Table 2.12.
2. When you are working on a calculator with built-in statistical functions, the calculator must perform all necessary operations on each data as the values are entered (most handheld nongraphing calculators do not have the ability to store data). Then after all data have been entered, the computations will be completed using the appropriate summations. The computations for sample variance will be done using formula (2.10), following the procedure shown in Table 2.14.
3. If you are doing the computations either by hand or with the aid of a calculator, but not using statistical functions, the most convenient formula to use will depend on how many data there are and how convenient the numerical values are to work with.

When a formula has multiple forms, look for one of these icons:

is used to identify the formula most likely to be used by a computer.

is used to identify the formula most likely to be used by a calculator.

is used to identify the formula most likely to be convenient for hand calculations.

D is used to identify the "definition" formula.

SECTION 2.5 EXERCISES

Statistics Now™

Skillbuilder Applet Exercises must be worked using an accompanying applet found on your Student's Suite CD-ROM or at the StatisticsNow website at **http://1pass.thomson.com.**

Datasets can be found on your Student's Suite CD-ROM or at the StatisticsNow website at **http://1pass.thomson.com.**

2.85 In 2004 the Tax Policy Center reported the following statistics about the average annual 2002 taxes and percent of personal income paid per person by state.

	Taxes per Capita	Rank	Percent of Personal Income	Rank
Hawaii	$2748	1	9.6	1
South Dakota	$1283	50	4.8	47
New Hampshire	$1478	45	4.4	50

Sources: Federation of Tax Administrators (2004) and U.S. Bureau of the Census and Bureau of Economic Analysis, http://taxpolicycenter.org/TaxFacts/TFDB/TFTemplate.cfm?Docid=309&Topic2id=90

a. Find the range for the amount of taxes paid per person.

b. Find the range for the percentage of personal income paid in taxes per person.

2.86 a. The data value $x = 45$ has a deviation value of 12. Explain the meaning of this.

b. The data value $x = 84$ has a deviation value of -20. Explain the meaning of this.

2.87 The summation $\Sigma(x - \bar{x})$ is always zero. Why? Think back to the definition of the mean (p. 73) and see if you can justify this statement.

2.88 All measures of variation are nonnegative in value for all sets of data.

a. What does it mean for a value to be "nonnegative"?

b. Describe the conditions necessary for a measure of variation to have the value zero.

c. Describe the conditions necessary for a measure of variation to have a positive value.

2.89 A sample contains the data {1, 3, 5, 6, 10}.

a. Use formula (2.6) to find the variance.

b. Use formula (2.10) to find the variance.

c. Compare the results from parts a and b.

2.90 Consider the sample 2, 4, 7, 8, 9. Find the following:

a. Range

b. Variance s^2, using formula (2.6)

c. Standard deviation, s

2.91 Consider the sample 6, 8, 7, 5, 3, 7. Find the following:

a. Range

b. Variance s^2, using formula (2.6)

c. Standard deviation, s

2.92 Given the sample 7, 6, 10, 7, 5, 9, 3, 7, 5, 13, find the following:

a. Variance s^2 using formula (2.6)

b. Variance s^2, using formula (2.10)

c. Standard deviation, s

2.93 Fifteen randomly selected college students were asked to state the number of hours they slept the previous night. The resulting data are 5, 6, 6, 8, 7, 7, 9, 5, 4, 8, 11, 6, 7, 8, 7. Find the following:

a. Variance s^2, using formula (2.6)

b. Variance s^2, using formula (2.10)

c. Standard deviation, s

2.94 [EX02-071] A random sample of 10 of the 2005 Nextel Cup NASCAR drivers produced the following ages: 33, 48, 41, 29, 40, 48, 44, 42, 49, 28.

a. Find the range.

b. Find the variance.

c. Find the standard deviation.

2.95 Adding (or subtracting) the same number from each value in a set of data does not affect the measures of variability for that set of data.

a. Find the variance of this set of annual heating-degree-day data: 6017, 6173, 6275, 6350, 6001, 6300.

b. Find the variance of this set of data (obtained by subtracting 6000 from each value in part a): 17, 173, 275, 350, 1, 300.

2.96 [EX02-096] One aspect of the beauty of scenic landscape is its variability. The elevations (feet above sea level) of 12 randomly selected towns in the Finger Lakes Regions of Upstate New York are recorded here.

559	815	767	668	651	895
1106	1375	861	1559	888	1106

Source: http://www.city-data.com

a. Find the mean.

b. Find the standard deviation.

2.97 [EX02-097] Recruits for a police academy were required to undergo a test that measures their exercise capacity. The exercise capacity (in minutes) was obtained for each of 20 recruits:

25	27	30	33	30	32	30	34	30	27
26	25	29	31	31	32	34	32	33	30

a. Draw a dotplot of the data.

b. Find the mean.

c. Find the range.

d. Find the variance.

e. Find the standard deviation.

f. Using the dotplot from part a, draw a line representing the range. Then draw a line starting at the mean with a length that represents the value of the standard deviation.

g. Describe how the distribution of data, the range, and the standard deviation are related.

2.98 [EX02-098] *Better Roads* magazine reported the percentage of interstate and state-owned bridges that were structurally deficient or functionally obsolete (%SD/FO) for each U.S. state in

2003. (Percentages are expressed in decimal form [e.g., 0.20 = 20%].)

State	SD/FO*	State	SD/FO*	State	SD/FO*
AK	0.20	AL	0.22	AR	0.20

••• Remainder of data on Student's Suite CD-ROM

Source: Better Roads, November 2003

*SD/FO = structurally deficient or functionally obsolete.

a. Construct a histogram.

b. Does the variable "%SD/FO" appear to have an approximately normal distribution?

c. Calculate the mean.

d. Find the median.

e. Find the range.

f. Find the standard deviation.

(Retain these solutions to use in Exercise 2.125 on p. 105.)

2.99 [EX02-099] One measure of airline performance is about overall flight on-time rates. For January 2005, the on-time arrival rates of domestic flights at the 31 largest U.S. airports were as follows.

ATL	69.1	BWI	74.0	BOS	62.1

••• Remainder of data on Student's Suite CD-ROM

Source: U.S. Department of Transportation, Bureau of Transportation Statistics

a. Find the range and the standard deviation for the on-time arrival rates.

b. Draw lines on the stem-and-leaf diagram drawn in answering Exercise 2.73 that represent the range and standard deviation. Remember: The standard deviation is a measure of the spread about the mean.

c. Describe the relationship among the distribution of the data, the range, and the standard deviation.

2.100 Consider these two sets of data:

Set 1	46	55	50	47	52
Set 2	30	55	65	47	53

Both sets have the same mean, 50. Compare these measures for both sets: $\sum(x-\bar{x})$, $\sum|x-\bar{x}|$, $SS(x)$, and range. Comment on the meaning of these comparisons.

2.101 Comment on the statement: "The mean loss for customers at First State Bank (which was not insured) was $150. The standard deviation of the losses was −$125."

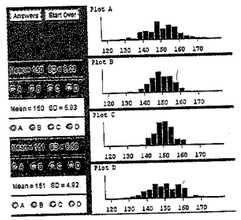
2.102 Start with $x = 100$ and add four x values to make a sample of five data such that:

a. $s = 0$ b. $0 < s < 1$

c. $5 < s < 10$ d. $20 < s < 30$

2.103 Each of two samples has a standard deviation of 5. If the two sets of data are made into one set of 10 data, will the new sample have a standard deviation that is less than, about the same as, or greater than the original standard deviation of 5? Make up two sets of five data, each with a standard deviation of 5, to justify your answer. Include the calculations.

2.104 Skillbuilder Applet Exercise matches means and standard deviations with corresponding histograms. After several practice rounds using "Start Over," explain your method of matching.

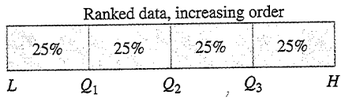


2.6 Measures of Position

Measures of position are used to describe the position a specific data value possesses in relation to the rest of the data when in ranked order. *Quartiles* and *percentiles* are two of the most popular measures of position.

Quartiles: Values of the variable that divide the ranked data into quarters; each set of data has three quartiles. The *first quartile*, Q_1, is a number such that at most 25% of the data are smaller in value than Q_1 and at most 75% are larger. The *second quartile* is the median. The *third quartile*, Q_3, is a number such that at most 75% of the data are smaller in value than Q_3 and at most 25% are larger. (See Figure 2.24.)

FIGURE 2.24
Quartiles



The procedure for determining the values of the quartiles is the same as that for percentiles and is shown in the following description of *percentiles*. Remember that your data must be ranked from low (L) to high (H).

Percentiles: Values of the variable that divide a set of ranked data into 100 equal subsets; each set of data has 99 percentiles (see Figure 2.25). The kth percentile, P_k, is a value such that at most k% of the data are smaller in value than P_k and at most (100 − k)% of the data are larger (see Figure 2.26).

FIGURE 2.25 Percentiles

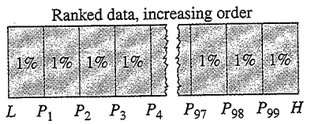
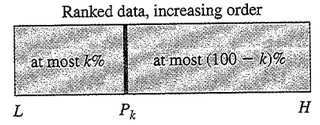


FIGURE 2.26 *k*th Percentile

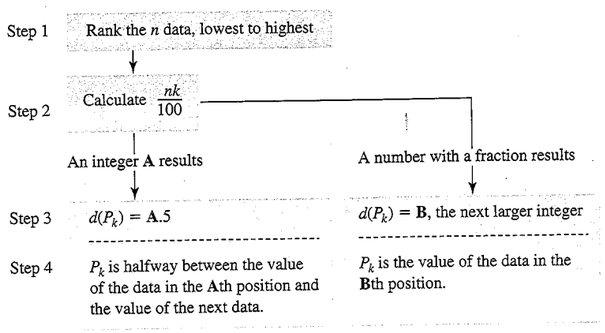


Notes:

1. The first quartile and the 25th percentile are the same; that is, $Q_1 = P_{25}$. Also, $Q_3 = P_{75}$.
2. The median, the second quartile, and the 50th percentile are all the same: $\tilde{x} = Q_2 = P_{50}$. Therefore, when asked to find P_{50} or Q_2, use the procedure for finding the median.

The procedure for determining the value of any kth percentile (or quartile) involves four basic steps as outlined on the diagram in Figure 2.27. Example 2.12 demonstrates the procedure.

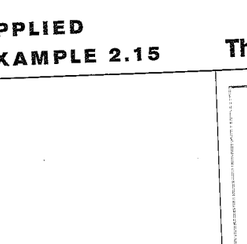
FIGURE 2.27
Finding P_k Procedure



APPLIED EXAMPLE 2.15 **The 85th Percentile Speed Limit**

DID YOU KNOW

The drivers in New York State drove a total of 135,046,000,000 miles on New York State roads in 2003. That's more than 5.4 million trips around the earth at the equator. That's a lot of driving!



GOING WITH 85% OF THE FLOW

To the uninitiated, the "85th percentile rule" seems bizarre, unorthodox, and maybe even scary, but this speed limit benchmark has guided traffic engineers for decades and is even recognized as official policy in many government jurisdictions. The idea is that maximum speed limits should be set so that 85% of the vehicles on a particular stretch of road are at or below the limit. Under California policies, traffic engineers routinely measure how fast motorists drive and then often set the limit at the 85th percentile of traffic speed.

"The reasoning is that 85% of people drive reasonably and 15% do not," said David Roseman, city traffic engineer for Long Beach. "So we should be designing our speeds to accommodate reasonable drivers." Adds Tom Jones, principal traffic engineer for the city of Los Angeles, "The 85th percentile rule was established many years ago. It is a design criteria [sic], but it doesn't mean that it is necessarily OK."

Safety advocacy groups hate the 85th percentile rule, because they believe that speeding is a serious and growing highway hazard. Indeed, police are performing fewer routine traffic patrols and speeds are creeping up, according to studies published by safety groups. Barbara Harsha, executive director of the Governors Highway Safety Assn. in Washington, for example, is concerned that the 85th percentile rule can be used to legitimize unsafe speeding. When congestion is not limiting speeds, many sections of Southern California freeways have average speeds of more than 80 mph, well above the legal limit. Posted limits of 25 mph on residential streets are routinely ignored, according to neighborhood traffic studies.

"That just shows that legal speed limits are too low," says Chad Dornsife of the National Motorists Assn., a group that represents people generally unhappy with and often indignant about traffic laws and police enforcement. He says improperly set low speed limits actually increase accidents and cost lives, because they encourage unequal speeds, creating a hazard. He claims, for example, that when Montana imposed speed limits for the first time, fatal accidents doubled. A secondary problem Dornsife cites regarding artificially low speed limits is that yellow-light intervals are sometimes based on the posted limits, which leaves too little time for faster-moving cars to stop for a changing light before reaching an intersection. That, Dornsife claims, creates intersection collisions. "Every generation that has gone through this doesn't believe in the 85th percentile rule," he adds. "The law-enforcement community doesn't like the 85th percentile rule because they write fewer tickets. New traffic engineers aren't even taught the 85th percentile rule."

Source: Ralph Vartabedian, *Los Angeles Times* Staff Writer, March 9, 2005, http://www.latimes.com/classified/automotive/highway1/la-hy-wheels9mar09,1,6721856.story?ctrack=2&cset=true

SECTION 2.6 EXERCISES

StatisticsNow™

Datasets can be found on your Student's Suite CD-ROM or at the StatisticsNow website at **http://1pass.thomson.com.**

2.105 Refer to the table of exam scores in Table 2.16 on page 94 for the following.

a. Using the concept of depth, describe the position of 91 in the set of 50 exam scores in two different ways.

b. Find P_{20} and P_{35} for the exam scores in Table 2.16 on page 94.

c. Find P_{80} and P_{95} for the exam scores in Table 2.16.

2.106 [EX02-106] Following are the American College Test (ACT) scores attained by the 25 members of a local high school graduating class:

21 24 23 17 31 19 19 20 19 25 17 23 16
21 20 28 25 25 21 14 19 17 18 28 20

a. Draw a dotplot of the ACT scores.

b. Using the concept of depth, describe the position of 24 in the set of 25 ACT scores in two different ways.

c. Find P_5, P_{10}, and P_{20} for the ACT scores.

d. Find P_{99}, P_{90}, and P_{80} for the ACT scores.

2.107 [EX02-107] The annual salaries (in \$100) of the kindergarten and elementary school teachers employed at one of the elementary schools in the local school district are listed here:

574 434 455 413 391 471 458 269 501
326 367 433 367 495 376 371 295 317

a. Draw a dotplot of the salaries.

b. Using the concept of depth, describe the position of 295 in the set of 18 salaries in two different ways.

c. Find Q_1 for these salaries.

d. Find Q_3 for these salaries.

2.108 [EX02-108] Fifteen countries were randomly selected from the *World Factbook 2004* list of world countries, and the infant mortality per 1000 live births rate was recorded:

6.38 101.68 9.48 69.18 64.19 3.73 21.31 52.71
13.43 29.64 15.24 5.85 11.74 9.67 8.68

Source: The World Factbook 2004

a. Find the first and third quartile for the infant mortality per 1000 rate.

b. Find the midquartile.

2.109 [EX02-109] The following data are the yields (in pounds) of hops:

3.9 3.4 5.1 2.7 4.4 7.0 5.6 2.6 4.8 5.6
7.0 4.8 5.0 6.8 4.8 3.7 5.8 3.6 4.0 5.6

a. Find the first and the third quartiles of the yields.

b. Find the midquartile.

c. Find and explain the percentiles P_{15}, P_{33}, and P_{90}.

2.110 [EX02-110] A research study of manual dexterity involved determining the time required to complete a task. The time required for each of 40 individuals with disabilities is shown here (data are ranked):

7.1 7.2 7.2 7.6 7.6 7.9 8.1 8.1 8.1 8.3 8.3 8.4 8.4 8.9
9.0 9.0 9.1 9.1 9.1 9.1 9.4 9.6 9.9 10.1 10.1 10.1 10.2
10.3 10.5 10.7 11.0 11.1 11.2 11.2 11.2 12.0 13.6 14.7 14.9 15.5

a. Find Q_1. b. Find Q_2.

c. Find Q_3. d. Find P_{95}.

e. Find the 5-number summary.

f. Draw the box-and-whisker display.

2.111 Draw a box-and-whiskers display for the set of data with the 5-number summary 42–62–72–82–97.

2.112 [EX02-112] The U.S. Geological Survey collected atmospheric deposition data in the Rocky Mountains. Part of the sampling process was to determine the concentration of ammonium ions (in percentages). Here are the results from the 52 samples:

2.9 4.1 2.7 3.5 1.4 5.6 13.3 3.9 4.0
2.9 7.0 4.2 4.9 4.6 3.5 3.7 3.3 5.7
3.2 4.2 4.4 6.5 3.1 5.2 2.6 2.4 5.2
4.8 4.8 3.9 3.7 2.8 4.8 2.7 4.2 2.9
2.8 3.4 4.0 4.6 3.0 2.3 4.4 3.1 5.5
4.1 4.5 4.6 4.7 3.6 2.6 4.0

a. Find Q_1. b. Find Q_2.

c. Find Q_3. d. Find the midquartile.

e. Find P_{30}. f. Find the 5-number summary.

g. Draw the box-and-whiskers display.

2.113 [EX02-113] The NCAA men's basketball "Big Dance" kicks into full gear every March. If you look at the graduation rate of these athletes, however, you will find that many teams do not make the grade, according to a study released in

March 2005. Following are the graduation rates for 64 of the 2005 tournament teams.

Graduation Rates (%), 2005 Men's Teams, NCAA Division I Basketball Tournament

40	38	100	55	44	58	30	11	40	19	43	27	0
64	75	58	54	44	40	11	50	30	40	71	92	53
33	29	40	25	33	43	25	27	58	47	55	60	8
17	17	0	40	25	14	67	45	33	20	15	50	27
29	57	36	45	45	73	15	100	67	44	57	55	

Source: 2004 NCAA Graduation-Rates Report

a. Draw a dotplot of the graduation rate data.

b. Draw a stem-and-leaf display of these data.

c. Find the 5-number summary and draw a box-and-whiskers display.

d. Find P_5 and P_{95}.

e. Describe the distribution of graduation rates, being sure to include information learned in parts a through d.

f. Are there teams whose graduation rates appear to be quite different from the rest? How many? Which ones? Explain.

2.114 [EX02-114] The fatality rate on the nation's highways in 2003 was the lowest since record keeping began 29 years ago, but these numbers are still mind-boggling. The number of persons killed in motor vehicle traffic crashes, by state, including the District of Columbia, in 2003 is listed here:

1001	294	135	471	462	262	439	668	203	943
95	142	293	928	1283	293	1491	512	1193	600
1120	67	1453	894	657	368	1531	1577	3675	394
627	3169	834	207	871	127	105	104	309	848
4215	1603	441	649	1232	747	1277	968	69	165
632									

Source: Road & Travel Magazine, 2004

a. Draw a dotplot of the fatality data.

b. Draw a stem-and-leaf display of these data.

c. Find the 5-number summary and draw a box-and-whiskers display. Describe how the three large-valued data are handled.

d. Find P_{10} and P_{90}.

e. Describe the distribution of number of fatalities per state, being sure to include information learned in parts a through d.

f. Why might it be unfair to draw conclusions about the relative safety level of highways in the 50 states and District of Columbia based on these data?

2.115 [EX02-115] Are airline flight arrivals ever on time? The general public thinks they are always late—but are they? The U.S. Bureau of Transportation Statistics keeps records and periodically reports the findings. Listed here are the percentages of on-time arrivals at the 31 major U.S. airports for the period January 1, 2004, to October 31, 2004.

ATL	73.55	BOS	78.38	BWI	80.91

••• Remainder of data on Student's Suite CD-ROM

Source: U.S. Department of Transportation, Bureau of Transportation Statistics

a. Draw a dotplot of on-time performance data.

b. Draw a stem-and-leaf display of these data.

c. Find the 5-number summary and draw a box-and-whiskers display.

d. Find P_{10} and P_{20}.

e. Describe the distribution of on-time percentage, being sure to include information learned in parts a through d.

f. Why would you be more likely to talk about the top 80% or 90% of the performance percentages than the middle 80% or 90%?

g. Are there airports whose on-time percentages appear to be quite different from the rest? How many? Which ones? Explain.

2.116 [EX02-116] Major league baseball stadiums vary in age, style, number of seats, and many other ways, but to the baseball players, the size of the field is of the utmost importance. Suppose we agree to measure the size of the field by using the distance from home plate to the centerfield fence. Following is the distance (in feet) to the centerfield fence in the 30 major league stadiums.

422	405	400	400	400	402	404	435	399	410
400	400	400	400	408	401	395	410	410	401
420	408	405	410	402	415	400	404	405	400

Source: http://mlb.mlb.com

a. Construct a histogram.

b. The interquartile range is described by the bounds of the middle 50% of the data, Q_1 and Q_3. Find the interquartile range.

c. Are there any fields that appear to be considerably smaller or larger than the others?

d. Is there a great deal of difference in the size of these 30 fields as measured by the distance to centerfield? Justify your answer with statistical evidence.

2.117 What property does the distribution need for the median, the midrange, and the midquartile to all be the same value?

2.118 [EX02-118] Henry Cavendish, an English chemist and physicist (1731–1810), approached many of his experiments using quantitative measurements. He was the first to accurately measure the density of the Earth. Following are 29 measurements (ranked for your convenience) of the density of the Earth done by Cavendish in 1798 using a torsion balance. Density is presented as a multiple of the density of water. (Measurements are in g/cm³.)

4.88	5.07	5.10	5.26	5.27	5.29	5.29	5.30	5.34	5.34
5.36	5.39	5.42	5.44	5.46	5.47	5.50	5.53	5.55	5.57
5.58	5.61	5.62	5.63	5.65	5.68	5.75	5.79	5.85	

Source: The data and descriptive information are based on material from "Do robust estimators work with real data?" by Stephen M. Stigler, *Annals of Statistics* 5 (1977), 1055–1098.

a. Describe the data set by calculating the mean, median, and standard deviation.

b. Construct a histogram and explain how it demonstrates the values of the descriptive statistics in part a.

c. Find the 5-number summary.

d. Construct a box-and-whiskers display and explain how it demonstrates the values of the descriptive statistics in part c.

e. Based on the two graphs, what "shape" is this distribution of measurements?

f. Assuming that Earth density measurements have an approximately normal distribution, approximately 95% of the data should fall within 2 standard deviations of the mean. Is this true?

2.119 Find the z-score for test scores of 92 and 63 on a test that has a mean of 72 and a standard deviation of 12.

2.120 A sample has a mean of 50 and a standard deviation of 4.0. Find the z-score for each value of x:

a. $x = 54$ b. $x = 50$

c. $x = 59$ d. $x = 45$

2.121 An exam produced grades with a mean score of 74.2 and a standard deviation of 11.5. Find the z-score for each test score x:

a. $x = 54$ b. $x = 68$

c. $x = 79$ d. $x = 93$

2.122 A nationally administered test has a mean of 500 and a standard deviation of 100. If your standard score on this test was 1.8, what was your test score?

2.123 A sample has a mean of 120 and a standard deviation of 20.0. Find the value of x that corresponds to each of these standard scores:

a. $z = 0.0$ b. $z = 1.2$

c. $z = -1.4$ d. $z = 2.05$

2.124

a. What does it mean to say that $x = 152$ has a standard score of $+1.5$?

b. What does it mean to say that a particular value of x has a z-score of -2.1?

c. In general, the standard score is a measure of what?

2.125 [EX02-098] Consider the percentage of interstate and state-owned bridges that were structurally deficient or functionally obsolete (SD/FO) listed in Exercise 2.98 on page 92.

a. Omit the names of the states and rank the SD/FO values in ascending order, reading horizontally in each row.

b. Construct a 5-number summary table and the corresponding box-and-whiskers display.

c. Find the midquartile percentage and the interquartile range.

This theorem says that within 2 standard deviations of the mean ($k = 2$), you will always find at least 75% (that is, 75% or more) of the data:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75, \textbf{ at least 75\%}$$

Figure 2.34 shows a mounded distribution that illustrates at least 75%.

If we consider the interval enclosed by 3 standard deviations on either side of the mean ($k = 3$), the theorem says that we will always find at least 89% (that is, 89% or more) of the data:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 0.89, \textbf{ at least 89\%}$$

Figure 2.35 shows a mounded distribution that illustrates at least 89%.

FIGURE 2.34 Chebyshev's Theorem with $k = 2$

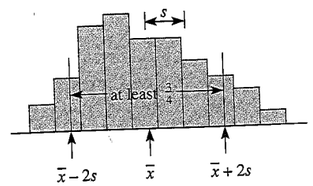
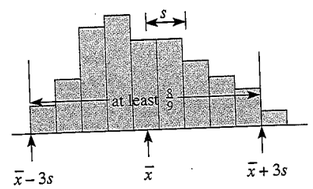
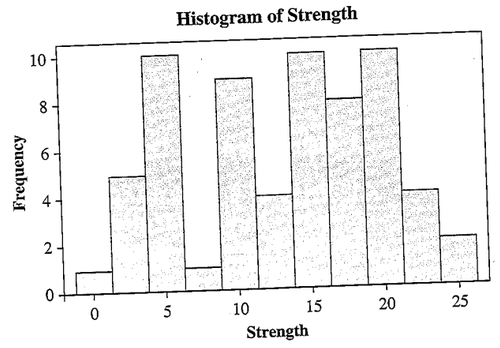


FIGURE 2.35 Chebyshev's Theorem with $k = 3$



Let's revisit the results of the physical-fitness strength test given to the third-graders in Exercise 2.45 on page 70. Their test results are listed here in rank order and shown on the histogram. [EX02-045]

1	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6
8	9	9	9	9	9	9	10	10	11	12	12	12	13	14	14
14	15	15	15	15	16	16	16	17	17	17	17	18	18	18	18
19	19	19	19	20	20	20	21	21	21	22	22	22	23	24	24



Some questions of interest are: Does this distribution satisfy the empirical rule? Does Chebyshev's theorem hold true? Is this distribution approximately normal?

To answer the first two questions, we need to find the percent of data in each of the three intervals about the mean. The mean is 13.0, and the standard deviation is 6.6.

mean ± *k* (Std. Dev.)	Interval	Percentage Found	Empirical	Chebyshev
13.0 ± 1(6.6)	6.4 to 19.6	39/64 = 60.9%	68%	–
13.0 ± 2(6.6)	−0.2 to 26.2	64/64 = 100%	95%	At least 75%
13.0 ± 3(6.6)	−6.8 to 32.8	64/64 = 100%	99.70%	At least 89%

It is left to you to verify the values of the mean, standard deviation, the intervals, and the percentages.

The three percentages found (60.9, 100, and 100) do not approximate the 68, 95, and 99.7 percentages stated in the empirical rule. The two percentages found (100 and 100) do agree with Chebyshev's theorem in that they are greater than 75% and 89%. Remember, Chebyshev's theorem holds for all distributions.

The normality test, introduced on pages 108 and 109, yields a *p*-value of 0.009, and along with the distribution seen on the histogram and the three percentages found, it is reasonable to conclude that these test results are not normally distributed.

SECTION 2.7 EXERCISES

Statistics Now™

Datasets can be found on your Student's Suite CD-ROM or at the StatisticsNow website at **http://1pass.thomson.com.**

2.129 Instructions for an essay assignment include the statement "The length is to be within 25 words of 200." What values of *x*, number of words, satisfy these instructions?

2.130 The empirical rule indicates that we can expect to find what proportion of the sample included between the following:

a. $\bar{x} - s$ and $\bar{x} + s$ b. $\bar{x} - 2s$ and $\bar{x} + 2s$

c. $\bar{x} - 3s$ and $\bar{x} + 3s$

2.131 Why is it that the *z*-score for a value that belongs to a normal distribution usually lies between −3 and +3?

2.132 The mean lifetime of a certain tire is 30,000 miles and the standard deviation is 2500 miles.

a. If we assume the mileages are normally distributed, approximately what percentage of all such tires will last between 22,500 and 37,500 miles?

b. If we assume nothing about the shape of the distribution, approximately what percentage of all such tires will last between 22,500 and 37,500 miles?

2.133 The average clean-up time for a crew of a medium-size firm is 84.0 hours and the standard deviation is 6.8 hours. Assume the empirical rule is appropriate.

a. What proportion of the time will it take the clean-up crew 97.6 hours or more to clean the plant?

b. Within what interval will the total clean-up time fall 95% of the time?

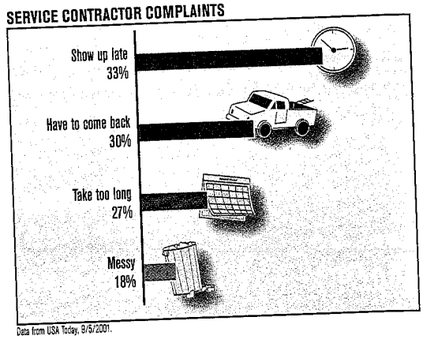
2.134 a. What proportion of a normal distribution is greater than the mean?

b. What proportion is within 1 standard deviation of the mean?

Truncated Scale

APPLIED EXAMPLE 2.17 **Simple Is Not Always Best**

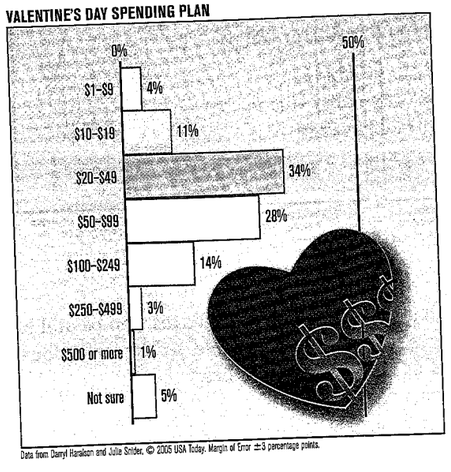
This graphic is neat and very readable, but does it represent the information being shown? Truncating scales on graphs often leads to misleading visual impressions. For example, in "Service contractor complaints," it appears that "Take too long" is twice as likely to be the complaint as "Messy." Look for other visual misrepresentations.



SECTION 2.8 EXERCISES

2.147 Is it possible for eight employees to earn between \$300 and \$350, a ninth to earn \$1250 per week, and the mean to be \$430? Verify your answer.

2.148 The graphic "Valentine's Day spending plan" shows a relative frequency distribution. This graph qualifies as a "tricky graph."



a. Is the graph a bar graph or a histogram? Explain.

b. How does this graph violate the guidelines for drawing histograms?

2.149 "What's wrong with this picture?" That's the question one should be asking when viewing the graphs in Applied Example 2.16 on page 115.

a. Find and describe at least four features about the *Ithaca Times* front-page graph that are incorrectly used.

b. Find and describe at least two features about the "Pecking Order" graph that are misrepresenting.

2.150 a. Find and describe at least four incorrect impressions created by truncating the horizontal axis on the Applied Example 2.17's "Service contractor complaints" graphic on page 116.

b. Redraw the bar graph starting the horizontal scale at "zero."

c. Comment on the effect that your graph has on the impression presented.

2.151 The best value for price was most often reported by consumers as one of the draws when deciding where to do their holiday shopping. When asked, "What draws holiday shoppers to stores?" at Christmas time in 2004, they responded as follows.

What	Percent
Value	76
Convenient location	68
Quality	62
Selection	60

Source: USA Today and NPD Group

Prepare two bar graphs to depict the percentage data. Scale the vertical axis on the first graph from 50 to 80. Scale the second graph from 0 to 80. What is your conclusion concerning how the percentages of the four responses stack up based on the two bar graphs, and what would you recommend, if anything, to improve the presentations?

2.152 Find an article or an advertisement containing a graph that in some way misrepresents the information of statistics. Describe how this graph misrepresents the facts.

2.9 Mean and Standard Deviation of Frequency Distribution (optional)

When the sample data are in the form of a frequency distribution, we need to make a slight adaptation to formulas (2.1) and (2.10) in order to find the mean, the variance, and the standard deviation.

EXAMPLE 2.18 **Calculations Using a Frequency Distribution**

TABLE 2.18
Ungrouped Frequency Distribution

x	f
1	5
2	9
3	8
4	6
	$\sum f = 28$

Find the mean, the variance, and the standard deviation for the sample data represented by the frequency distribution in Table 2.18.

Note: This frequency distribution represents a sample of 28 values: five 1s, nine 2s, eight 3s, and six 4s.

In order to calculate the sample mean $\bar{x}$ and the sample variance s^2 using formulas (2.1) and (2.10), we need the sum of the 28 x values, $\sum x$, and the sum of the 28 x-squared values, $\sum x^2$.

The summations, $\sum x$ and $\sum x^2$, could be found as follows:

$$\sum x = \underbrace{1 + 1 + \cdots + 1}_{5 \text{ of them}} + \underbrace{2 + 2 + \cdots + 2}_{9 \text{ of them}} + \underbrace{3 + 3 + \cdots + 3}_{8 \text{ of them}} + \underbrace{4 + 4 + \cdots + 4}_{6 \text{ of them}}$$

$$= (5)(1) + (9)(2) + (8)(3) + (6)(4)$$

$$= 5 + 18 + 24 + 24 = \mathbf{71}$$

$$\sum x^2 = \underbrace{1^2 + \cdots + 1^2}_{5 \text{ of them}} + \underbrace{2^2 + \cdots + 2^2}_{9 \text{ of them}} + \underbrace{3^2 + \cdots + 3^2}_{8 \text{ of them}} + \underbrace{4^2 + \cdots + 4^2}_{6 \text{ of them}}$$

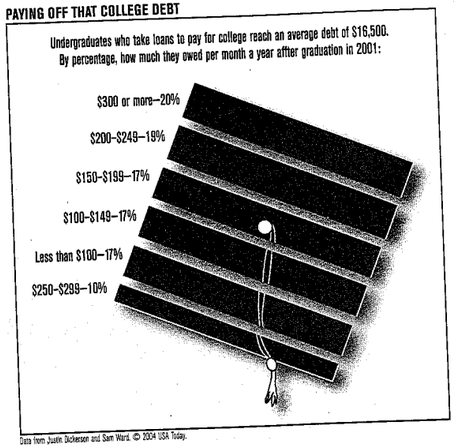
$$= (5)(1) + (9)(4) + (8)(9) + (6)(16)$$

$$= 5 + 36 + 72 + 96 = \mathbf{209}$$

a. Construct a circle graph showing this same information.

b. Compare the appearance of the divided bar graph given and the circle graph drawn in part a. Which one is easier to read? Which one gives a more accurate representation of the information being presented?

2.174 One of the ways that students pay for college is to borrow money. These loans eventually must be paid back, and the accompanying divided bar graph shows the monthly debt that many face after graduation.



a. Construct a circle graph showing this same information.

b. Construct a bar graph showing this same information.

c. Compare the appearance of the divided bar graph given with the circle graph drawn in part a and the bar graph drawn in part b. Which one best represents the relationship between various monthly debt amounts?

2.175 [EX02-175] The 10 leading causes of death in the United States during 2002 were listed by the National Center for Statistics and Analysis in a January 2005 report. There were a total of 2,443,387 deaths recorded.

Cause of Death	Number (10,000s)
Alzheimer's disease	5.9
Chronic respiratory disease	12.5
Diabetes	7.3
Heart disease	69.7
Influenza/pneumonia	6.6
Malignant neoplasms	55.7
Motor vehicle traffic crashes	4.4
Nephritis/nephrosis	4.1
Septicemia	3.4
Stroke	16.3

Source: NHTSA's National Center for Statistics and Analysis

a. Construct a Pareto diagram of this information.

b. Write a paragraph describing what the Pareto diagram dramatically shows to its reader.

2.176 [EX02-176] The U.S. Census Bureau posted the following 2003 age distribution for the people of New York State. The 2003 American Community Survey is limited to the household population and excludes the population living in institutions, college dormitories, and other group quarters.

Sex and Age Distribution

Male	8,953,019	Younger than 5 years	1,205,816
Female	9,647,508	5–14 years	2,537,813
		15–24 years	2,353,665
		25–34 years	2,587,995
		35–44 years	2,991,609
		45–54 years	2,682,845
		55–64 years	1,897,521
		65–74 years	1,218,850
		75–84 years	857,177
		85 years and older	267,236

Source: U.S. Census Bureau

a. Construct a relative frequency distribution of both the gender and the age data.

b. Construct a bar graph of the gender data.

c. Construct a histogram of the age data.

d. Explain why the graph drawn in part b is not a histogram and the graph drawn in part c is a histogram.

2.177 Identify each of the following as examples of (1) attribute (qualitative) or (2) numerical (quantitative) variables.

a. Scores registered by people taking their written state automobile driver's license examination

b. Whether or not a motorcycle operator possesses a valid motorcycle operator's license

c. The number of television sets installed in a house

d. The brand of bar soap being used in a bathroom

e. The value of a cents-off coupon used with the purchase of a box of cereal

2.178 Identify each of the following as examples of (1) attribute (qualitative) or (2) numerical (quantitative) variables.

a. The amount of weight lost in the past month by a person following a strict diet

b. Batting averages of major league baseball players

c. Decisions by the jury in felony trials

d. Sunscreen usage before going in the sun (always, often, sometimes, seldom, never)

e. Reason a manager failed to act against an employee's poor performance

2.179 Consider samples A and B. Notice that the two samples are the same except that the 8 in A has been replaced by a 9 in B.

A:	2	4	5	5	7	8
B:	2	4	5	5	7	9

What effect does changing the 8 to a 9 have on each of the following statistics?

a. Mean b. Median c. Mode d. Midrange

e. Vange f. Variance g. Std. dev.

2.180 Consider samples C and D. Notice that the two samples are the same except for two values.

C:	20	60	60	70	90
D:	20	30	70	90	90

What effect does changing the two 60s to 30 and 90 have on each of the following statistics?

a. Mean b. Median c. Mode d. Midrange

e. Range f. Variance g. Std. dev.

2.181 The addition of a new accelerator is claimed to decrease the drying time of latex paint by more than 4%. Several test samples were conducted with the following percentage decreases in drying time:

5.2	6.4	3.8	6.3	4.1	2.8	3.2	4.7

a. Find the sample mean.

b. Find the sample standard deviation.

c. Do you think these percentages average 4 or more? Explain.

(Retain these solutions to use in Exercise 9.28, p. 490.)

2.182 [EX02-182] Gasoline pumped from a supplier's pipeline is supposed to have an octane rating of 87.5. On 13 consecutive days, a sample of octane ratings was taken and analyzed, with the following results:

88.6	86.4	87.2	88.4	87.2	87.6	86.8
86.1	87.4	87.3	86.4	86.6	87.1	

a. Find the sample mean.

b. Find the sample standard deviation.

c. Do you think these readings average 87.5? Explain.

(Retain these solutions to use in Exercise 9.56, p. 494.)

2.183 [EX02-183] These data are the ages of 118 known offenders who committed an auto theft last year in Garden City, Michigan:

11	14	15	15	16	16	17	18	19	21	25	36
12	14	15	15	16	16	17	18	19	21	25	39
13	14	15	15	16	17	17	18	20	22	26	43
13	14	15	15	16	17	17	18	20	22	26	46
13	14	15	16	16	17	17	18	20	22	27	50
13	14	15	16	16	17	17	19	20	23	27	54
13	14	15	16	16	17	18	19	20	23	29	59
13	15	15	16	16	17	18	19	20	23	30	67
14	15	15	16	16	17	18	19	21	24	31	
14	15	15	16	16	17	18	19	21	24	34	

a. Find the mean. b. Find the median.

c. Find the mode. d. Find Q_1 and Q_3.

e. Find P_{10} and P_{95}.

2.184 [EX02-184] A survey of 32 workers at building 815 of Eastman Kodak Company was

taken last May. Each worker was asked: "How many hours of television did you watch yesterday?" The results were as follows:

0	0	½	1	2	0	3	2½	0	0	1
1½	5	2½	0	2	2½	1	0	2	0	
2½	4	06	2½	0	½	1	1½	0	2	

a. Construct a stem-and-leaf display.
b. Find the mean.
c. Find the median.
d. Find the mode.
e. Find the midrange.
f. Which measure of central tendency would best represent the average viewer if you were trying to portray the typical television viewer? Explain.
g. Which measure of central tendency would best describe the amount of television watched? Explain.
h. Find the range.
i. Find the variance.
j. Find the standard deviation.

2.185 [EX02-185] The stopping distance on a wet surface was determined for 25 cars, each traveling at 30 miles per hour. The data (in feet) are shown on the following stem-and-leaf display:

```
 6 | 3 7 6 3 9
 7 | 4 2 0 1 1 2 0 5
 8 | 5 4 5 5 6
 9 | 4 1 0 0 5
10 | 5 4
```

Find the mean and the standard deviation of these stopping distances.

2.186 [EX02-186] Forbes.com posted the 2004 EPS (earnings per share) in dollars for 17 banking industry companies.

Name	EPS ($)	Name	EPS ($)
Astoria Financial	2.92	Popular	1.71
Banknorth Group	2.20	State Street	3.13
Bank of America	3.67	Synovus Finl	1.36
BB & T	2.61	UnionBanCal	4.70
Compass Bancshares	2.86	Wachovia	3.68
Golden West Finl	3.97	Wells Fargo	4.00
M & T Bank	5.74	Westcorp	3.71
National City	3.75	Zions Bancorp	4.36
North Fork Bancorp	1.83		

Source: http://www.forbes.com/lists/results.jhtml

a. Find the mean EPS for the banks.
b. Find the median EPS for the banks.
c. Find the midrange of the EPS for the banks.
d. Write a discussion comparing the results from parts a, b, and c.
e. Find the standard deviation of the EPS for the banks.
f. Find the percentage of the data that are within 1 standard deviation of the mean.
g. Find the percentage of the data that are within 2 standard deviations of the mean.
h. Based on the preceding results, discuss whether you think the data are normally distributed, and why.

2.187 [EX02-187] The Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation, reported the number of mishandled baggage reports filed per 1000 airline passengers during October 2004. The industry average was 4.02.

Airline	Reports	Passengers	Reports/1000
AirTran	2084	1,148,779	1.81
JetBlue	2295	1,057,510	2.17

••• Remainder of data on Student's Suite CD-ROM

Source: Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation

a. Define the terms *population* and *variable* with regard to this information.
b. Are the numbers reported (1.81, 2.17, . . . , 12.21) data or statistics? Explain.
c. Is the average, 4.02, a data value, a statistic, or a parameter value? Explain why.
d. Is the "industry average" the mean of the airline rates of reports per 1000? If not, explain in detail how the 19 airline values are related to the industry average.

2.188 [EX02-188] One of the first scientists to study the density of nitrogen was Lord Raleigh. He noticed that the density of nitrogen produced from the air seemed to be greater than the density of nitrogen produced from chemical compounds. Do his conclusions seem to be justified even though he has so little data?

Lord Raleigh's measurements, which first appeared in *Proceedings, Royal Society* (London, 55, 1894, pp. 340–344) are listed here. The data are the mass of nitrogen filling a certain flask under specified pressure and temperature.

Atmospheric		Chemical	
2.31017	2.31010	2.30143	2.29940
2.30986	2.31028	2.29890	2.29849
2.31010	2.31163	2.29816	2.29889
2.31001	2.30956	2.30182	2.30074
2.31024		2.29869	2.30054

Source: http://exploringdata.cqu.edu.au/datasets/nitrogen.xls

a. Construct side-by-side dotplots of the two sets of data, using a common scale.
b. Calculate mean, median, standard deviation, and first and third quartiles for each set of data.
c. Construct side-by-side boxplots of the two sets of data, using a common scale.
d. Discuss how these two sets of data compare. Do these two very small sets of data show convincing evidence of a difference?

FYI The differences between these sets of data helped lead to the discovery of argon.

2.189 [EX02-189] The top 2004 Nationwide Tour money leaders, together with their total earnings, are listed here:

Player	Money ($)	Player	Money ($)
Jimmy Walker	371,346	D. A. Points	332,815

••• Remainder of data on Student's Suite CD-ROM

Source: PGA Tour, Inc.

a. Calculate the mean and standard deviation of the earnings of the Nike Tour golf players.
b. Find the values of $\bar{x} - s$ and $\bar{x} + s$.
c. How many of the 50 pieces of data have values between $\bar{x} - s$ and $\bar{x} + s$? What percentage of the sample is this?
d. Find the values of $\bar{x} - 2s$ and $\bar{x} + 2s$.
e. How many of the 50 pieces of data have values between $\bar{x} - 2s$ and $\bar{x} + 2s$? What percentage of the sample is this?
f. Find the values of $\bar{x} - 3s$ and $\bar{x} + 3s$.
g. What percentage of the sample has values between $\bar{x} - 3s$ and $\bar{x} + 3s$?
h. Compare the answers found in parts e and g to the results predicted by Chebyshev's Theorem.
i. Compare the answers found in parts c, e, and g to the results predicted by the empirical rule. Does the result suggest an approximately normal distribution?
j. Verify your answer to part i using one of the sets of technology instructions.
k. Does your answer to part j make sense? Explain.

2.190 Ask one of your instructors for a list of exam grades (15 to 25 grades) from a class.
a. Find five measures of central tendency.
b. Find the three measures of dispersion.
c. Construct a stem-and-leaf display. Does this diagram suggest that the grades are normally distributed?
d. Find the following measures of location: (i) Q_1 and Q_3, (ii) P_{15} and P_{60}, and (iii) the standard score z for the highest grade.

2.191 [EX02-191] The lengths (in millimeters) of 100 brown trout in pond 2-B at Happy Acres Fish Hatchery on June 15 of last year were as follows:

15.0 15.3 14.4 10.4

••• Remainder of data on Student's Suite CD-ROM

a. Find the mean.
b. Find the median.
c. Find the mode.
d. Find the midrange.
e. Find the range.
f. Find Q_1 and Q_3.
g. Find the midquartile.
h. Find P_{35} and P_{64}.
i. Construct a grouped frequency distribution that uses 10.0–10.5 as the first class.
j. Construct a histogram of the frequency distribution.
k. Construct a cumulative relative frequency distribution.
l. Construct an ogive of the cumulative relative frequency distribution.
m. Find the mean of the frequency distribution. (Optional.)
n. Find the standard deviation of the frequency distribution. (Optional.)

TABLE 4.2

M&M Colors by Percentages

Color	Percent
Brown	13.2
Yellow	16.2
Red	14.7
Blue	21.8
Orange	19.8
Green	14.3
	100.0

large proportion of just one or two colors. Have you ever noticed either of these extremes when you opened a bag of M&M's?

The percents reported on Table 4.2 are the percentages for each color found in this sample of 692 M&M's. Percentages behave very much like probability numbers, but the question being asked in probability is quite different. In the preceding illustration, we are treating the information as sample data and describing the results found. If we now think in terms of a probability, we will turn the orientation around and treat the complete set of 692 M&M's as the complete list of possibilities and ask questions about the likeliness of certain events when one M&M is randomly selected from the entire collection of 692.

For example, suppose we were to dump all 692 M&M's into a large bowl and thoroughly mix them. Now consider the question, "If one M&M is selected at random from the bowl, what is the probability that it will be orange?" We hope that your thinking is along the following lines: selected randomly means each M&M has the same chance of being selected, and because there are 137 orange M&M's in the bowl, the probability of selecting an orange M&M is 137/692, or 0.198.

You have seen the number 0.198 before, only it was expressed as 19.8%. Percentages and probability numbers are "the same, but different." (You have probably heard that before, somewhere.) The numbers have the same value and behave with the same properties; however, the orientation of the situation and the questions asked are different, as you will see in Section 4.2.

After completing Chapter 4, you will have an opportunity to further investigate "Sweet Statistics" in the Chapter 4 Project section.

SECTION 4.1 EXERCISES

4.1 a. If you bought a bag of M&M's, what color M&M would you expect to see the most? What color the least? Why?

b. If you bought a bag of M&M's, would you expect to find the percentages listed previously in Table 4.2? If not, why and what would you expect?

4.2 a. Construct a bar graph showing the Table 4.2 percentages obtained from the 692 M&M's.

b. Based on your graph, which color M&M occurred most often? How does this show on your graph?

c. Based on your graph, which color M&M occurred the least? How does this show on your graph?

4.3 If you were given a small bag of M&M's with 40 candies in it, using the percentages in Table 4.2, how many of each color would you "expect" to find?

4.4 *Bad charts?* Just like there are bad graphs (as seen in Section 2.8), there are bad charts—misleading and hard-to-read charts. Mothers Against Drunk Driving (MADD) presented the following chart regarding the 6764 holiday traffic fatalities that occurred in 2002.

Holiday 2002	Total Traffic Fatalities	Total Fatalities Alcohol-Related
New Year's Eve (2001)	118	45
New Year's Day	165	94
New Year's Holiday	575	301
Super Bowl Sunday	147	86
St. Patrick's Day	158	72
Memorial Day	491	237
Fourth of July	683	330
Labor Day weekend	541	300
Halloween	268	109
Thanksgiving	543	255
Thanksgiving–New Year's	4019	1561
Christmas	130	68
New Year's Eve (2002)	123	57

Source: Mothers Against Drunk Driving (MADD), http://www.infoplease.com/ipa/A0777960.html

a. The column totals are not included because they would be meaningless values. Examine the table and explain why.

b. Select the appropriate nonoverlapping holidays (column 1) and verify the 6764 total number of traffic fatalities for 2002.

c. Using the holidays selected in part b, find the total number of holiday alcohol-related traffic fatalities for 2002?

d. Describe how you would organize this chart to make it more meaningful.

4.5 Use either the random-number table (Appendix B), a calculator, or a computer (see p. 101) to simulate the following:

a. The rolling of a die 50 times; express your results as relative frequencies.

b. The tossing of a coin 100 times; express your results as relative frequencies.

4.6 Use either the random-number table (Appendix B), a calculator, or a computer (see p. 101) to simulate the random selection of 100 single-digit numbers, 0 through 9.

a. List the 100 digits.

b. Prepare a relative frequency distribution of the 100 digits.

c. Prepare a relative frequency histogram of the distribution in part b.

4.2 Probability of Events

We are now ready to define what is meant by *probability*. Specifically, we talk about "the probability that a certain event will occur."

Probability an event will occur: The relative frequency with which that event can be expected to occur.

The probability of an event may be obtained in three different ways: (1) *empirically*, (2) *theoretically*, and (3) *subjectively*.

The **empirical** method was illustrated by the M&M's and their percentages in Section 4.1 and might be called *experimental* or *empirical probability*. This probability is the *observed relative frequency* with which an event occurs. In our M&M example, we observed that 137 of the 692 M&M's were orange. The observed empirical probability for the occurrence of orange was 137/692, or 0.198.

The value assigned to the probability of event A as a result of experimentation can be found by means of the formula:

Empirical (Observed) Probability: $P'(A)$

In words: $$\text{empirical probability of } A = \frac{\text{number of times } A \text{ occurred}}{\text{number of trials}}$$

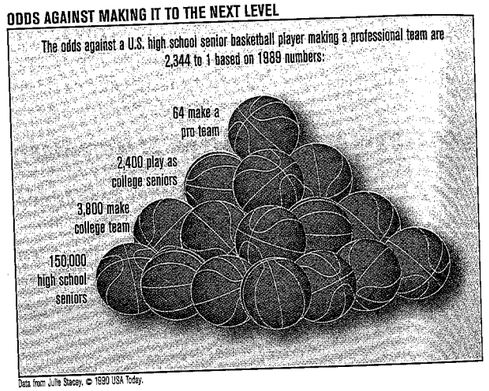
In algebra: $$P'(A) = \frac{n(A)}{n} \qquad (4.1)$$

Therefore, the probability of rain tomorrow is $\frac{4}{4+1}$, or $\frac{4}{5} = 0.8$. The odds against rain tomorrow are 1 to 4 (or 1:4), and the probability that there is no rain tomorrow is $\frac{1}{4+1}$, or $\frac{1}{5} = 0.2$.

APPLIED EXAMPLE 4.7 **Trying to Beat the Odds**

Many young men aspire to become professional athletes. Only a few make it to the big time, as indicated in the following graph. For every 2400 college senior basketball players, only 64 make a professional team; that translates to a probability of only 0.027 (64/2400).

There are many other interesting specifics hidden in this information. For example, many high school boys dream of becoming a professional basketball player, but according to these numbers, the probability of their dream being realized is only 0.000427 (64/150000).



Once a player has made a college basketball team, he might be very interested in the odds that he will play as a senior. Of the 3800 players making a college team, 2400 play as seniors, whereas 1400 do not. Thus, if a player has made a college team, the odds he will play as a senior are 2400 to 1400, which reduces to 12 to 7.

The college senior who is playing is interested in his chances of making it to the next level. We see that of the 2400 college seniors, only 64 make the pros, whereas 2336 do not; thus, the odds against him making it to the next level are 2336 to 64, which reduces to 73 to 2. Odds are strongly against him making it.

SECTION 4.2 EXERCISES

StatisticsNow™

Skillbuilder Applet Exercises must be worked using an accompanying applet found on your Student's Suite CD-ROM or at the StatisticsNow website at **http://1pass.thomson.com.**

4.7 If you roll a die 40 times and 9 of the rolls result in a "5," what empirical probability was observed for the event "5"?

4.8 Explain why an empirical probability, an observed proportion, and a relative frequency are actually three different names for the same thing.

4.9 Millions of people ride railroads every year. The National Association of Railroad Passengers provides the following figures for railroad ridership in 2004.

Rail System	Riders (millions)
Amtrak system	25.0
Northeast Corridor	14.2
Intercity + West	10.8

Source: National Association of Railroad Passengers, http://www.infoplease.com/ipa/A0855824.html

a. What percentage of the railroad riders rode the Amtrak system in 2004?

b. If one of these riders is to be interviewed, what is the probability that the rider rode the Amtrak system in 2004 if he or she is selected at random?

c. Explain the difference and the relationship between questions and answers in parts a and b.

4.10 Webster Aquatic Center offers various levels of swimming lessons year-round. The March 2005 Monday and Wednesday evening lessons included instructions from Water Babies through Adults. The number in each classification is given in the table that follows.

Swim Lesson Types	No. of Participants
Water Babies	15
Tiny Tots	12
Tadpoles	12
Level 2	15
Level 3	10
Level 4	6
Level 5	2
Level 6	1
Adults	4
Total	77

If one participant is selected at random, find the probability of the following:

a. The participant is in Tiny Tots.

b. The participant is in the Adults lesson.

c. The participant is in a Level 2 to Level 6 lesson.

4.11 In September 2004, the American Payroll Association reported the results of their National Payroll Week 2004 Survey. One of the questions asked about annual household income.

Annual Household Income	Number	Percentage
Less than $15,000	423	1.9%
$15,001–$30,000	2225	9.8%
$30,001–$50,000	5394	23.9%
$50,001–$75,000	5772	25.5%
$75,001–$100,000	4730	20.9%
$100,001–$150,000	3065	13.6%
More than $150,000	984	4.4%

Source: American Payroll Association, http://www.AmericanPayroll.org

Suppose one of the respondents from the survey is to be selected at random for a follow-up interview. Find the probability of the following events.

a. The respondent's family income is $50,000 or less.

b. The respondent's family income is $75,001 or more.

c. The respondent's family income is between $30,000 and $100,000.

d. The respondent's family income is at least $100,001.

4.12 The U.S. Department of Transportation annually reports the number of consumer complaints against the top U.S. airlines by category. Following are the figures for 2002.

Complaint Category	Number of Complaints	Complaint Category	Number of Complaints
Flight problems	2031	Oversales	454
Customer service	1715	Fares	523
Baggage	1421	Disability	477
Reservations/ticketing/ boarding	1159	Advertising	68
Refunds	1106	Other	322

Source: Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation, Air Travel Consumer Report, http://www.infoplease.com/ipa/A0198353.html

If one of these complaints is selected at random for follow-up evaluation, what is the probability that the complaint is:

a. About flight problems?