3–3

Measures of Position

Objective 3

Identify the position of a data value in a data set, using various measures of position, such as percentiles, deciles, and quartiles.

In addition to measures of central tendency and measures of variation, there are measures of position or location. These measures include standard scores, percentiles, deciles, and quartiles. They are used to locate the relative position of a data value in the data set. For example, if a value is located at the 80th percentile, it means that 80% of the values fall below it in the distribution and 20% of the values fall above it. The *median* is the value that corresponds to the 50th percentile, since one-half of the values fall below it and one-half of the values fall above it. This section discusses these measures of position.

Standard Scores

There is an old saying, "You can't compare apples and oranges." But with the use of statistics, it can be done to some extent. Suppose that a student scored 90 on a music test and 45 on an English exam. Direct comparison of raw scores is impossible, since the exams might not be equivalent in terms of number of questions, value of each question, and so on. However, a comparison of a relative standard similar to both can be made. This comparison uses the mean and standard deviation and is called a standard score or z score. (We also use z scores in later chapters.)

A standard score or z score tells how many standard deviations a data value is above or below the mean for a specific distribution of values. If a standard score is zero, then the data value is the same as the mean.

A ***z* score** or **standard score** for a value is obtained by subtracting the mean from the value and dividing the result by the standard deviation. The symbol for a standard score is z. The formula is

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

For samples, the formula is

$$z = \frac{X - \overline{X}}{s}$$

For populations, the formula is

$$z = \frac{X - \mu}{\sigma}$$

The z score represents the number of standard deviations that a data value falls above or below the mean.

For the purpose of this section, it will be assumed that when we find z scores, the data were obtained from samples.

Example 3–29

Test Scores

A student scored 65 on a calculus test that had a mean of 50 and a standard deviation of 10; she scored 30 on a history test with a mean of 25 and a standard deviation of 5. Compare her relative positions on the two tests.

Interesting Fact

The average number of faces that a person learns to recognize and remember during his or her lifetime is 10,000.

Solution

First, find the z scores. For calculus the z score is

$$z = \frac{X - \overline{X}}{s} = \frac{65 - 50}{10} = 1.5$$

For history the z score is

$$z = \frac{30 - 25}{5} = 1.0$$

Since the z score for calculus is larger, her relative position in the calculus class is higher than her relative position in the history class.

Note that if the z score is positive, the score is above the mean. If the z score is 0, the score is the same as the mean. And if the z score is negative, the score is below the mean.

Example 3–30

Test Scores

Find the z score for each test, and state which is higher.

Test A	$X = 38$	$\overline{X} = 40$	$s = 5$
Test B	$X = 94$	$\overline{X} = 100$	$s = 10$

Solution

For test A,

$$z = \frac{X - \overline{X}}{s} = \frac{38 - 40}{5} = -0.4$$

For test B,

$$z = \frac{94 - 100}{10} = -0.6$$

The score for test A is relatively higher than the score for test B.

When all data for a variable are transformed into z scores, the resulting distribution will have a mean of 0 and a standard deviation of 1. A z score, then, is actually the number of standard deviations each value is from the mean for a specific distribution. In Example 3–29, the calculus score of 65 was actually 1.5 standard deviations above the mean of 50. This will be explained in greater detail in Chapter 6.

Percentiles

Percentiles are position measures used in educational and health-related fields to indicate the position of an individual in a group.

Percentiles divide the data set into 100 equal groups.

In many situations, the graphs and tables showing the percentiles for various measures such as test scores, heights, or weights have already been completed. Table 3–3 shows the percentile ranks for scaled scores on the Test of English as a Foreign Language. If a student had a scaled score of 58 for section 1 (listening and comprehension), that student would have a percentile rank of 81. Hence, that student did better than 81% of the students who took section 1 of the exam.

Interesting Facts

The highest recorded temperature on earth was 136°F in Libya in 1922. The lowest recorded temperature on earth was −129°F in Antarctica in 1983.

Table 3–3 Percentile Ranks and Scaled Scores on the Test of English as a Foreign Language*

Scaled score	Section 1: Listening comprehension	Section 2: Structure and written expression	Section 3: Vocabulary and reading comprehension	Total scaled score	Percentile rank
68	99	98			
66	98	96	98	660	99
64	96	94	96	640	97
62	92	90	93	620	94
60	87	84	88	600	89
→58	81	76	81	580	82
56	73	68	72	560	73
54	64	58	61	540	62
52	54	48	50	520	50
50	42	38	40	500	39
48	32	29	30	480	29
46	22	21	23	460	20
44	14	15	16	440	13
42	9	10	11	420	9
40	5	7	8	400	5
38	3	4	5	380	3
36	2	3	3	360	1
34	1	2	2	340	1
32		1	1	320	
30		1	1	300	
Mean	51.5	52.2	51.4	517	Mean
S.D.	7.1	7.9	7.5	68	S.D.

*Based on the total group of 1,178,193 examinees tested from July 1989 through June 1991.

Source: Reprinted by permission of Educational Testing Service, the copyright owner. However, the test question and any other testing information are provided in their entirety by McGraw-Hill Companies, Inc. No endorsement of this publication by Educational Testing Service should be inferred.

Figure 3–5 shows percentiles in graphical form of weights of girls from ages 2 to 18. To find the percentile rank of an 11-year-old who weighs 82 pounds, start at the 82-pound weight on the left axis and move horizontally to the right. Find 11 on the horizontal axis and move up vertically. The two lines meet at the 50th percentile curved line; hence, an 11-year-old girl who weighs 82 pounds is in the 50th percentile for her age group. If the lines do not meet exactly on one of the curved percentile lines, then the percentile rank must be approximated.

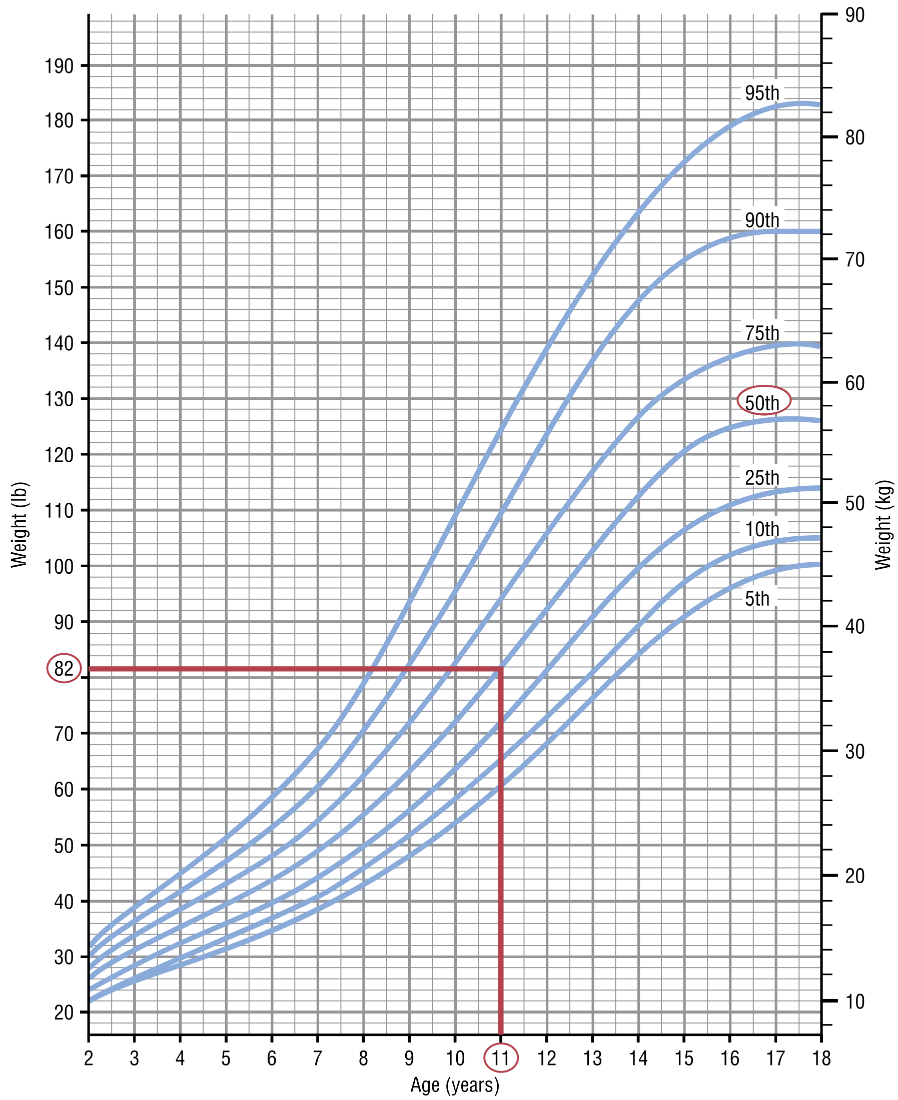
Percentiles are also used to compare an individual's test score with the national norm. For example, tests such as the National Educational Development Test (NEDT) are taken by students in ninth or tenth grade. A student's scores are compared with those of other students locally and nationally by using percentile ranks. A similar test for elementary school students is called the California Achievement Test.

Percentiles are not the same as percentages. That is, if a student gets 72 correct answers out of a possible 100, she obtains a percentage score of 72. There is no indication of her position with respect to the rest of the class. She could have scored the highest, the lowest, or somewhere in between. On the other hand, if a raw score of 72 corresponds to the 64th percentile, then she did better than 64% of the students in her class.

Figure 3–5

Weights of Girls by Age and Percentile Rankings

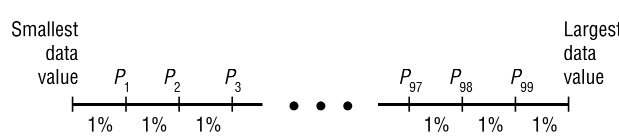
Source: Distributed by Mead Johnson Nutritional Division. Reprinted with permission.



Percentiles are symbolized by

$$P_1, P_2, P_3, \ldots, P_{99}$$

and divide the distribution into 100 groups.



Percentile graphs can be constructed as shown in Example 3–31. Percentile graphs use the same values as the cumulative relative frequency graphs described in Section 2–2, except that the proportions have been converted to percents.

Example 3–31 Systolic Blood Pressure

The frequency distribution for the systolic blood pressure readings (in millimeters of mercury, mm Hg) of 200 randomly selected college students is shown here. Construct a percentile graph.

A Class boundaries	B Frequency	C Cumulative frequency	D Cumulative percent
89.5–104.5	24		
104.5–119.5	62		
119.5–134.5	72		
134.5–149.5	26		
149.5–164.5	12		
164.5–179.5	4		
	200		

Solution

Step 1 Find the cumulative frequencies and place them in column C.

Step 2 Find the cumulative percentages and place them in column D. To do this step, use the formula

$$\text{Cumulative \%} = \frac{\text{cumulative frequency}}{n} \cdot 100$$

For the first class,

$$\text{Cumulative \%} = \frac{24}{200} \cdot 100 = 12\%$$

The completed table is shown here.

A Class boundaries	B Frequency	C Cumulative frequency	D Cumulative percent
89.5–104.5	24	24	12
104.5–119.5	62	86	43
119.5–134.5	72	158	79
134.5–149.5	26	184	92
149.5–164.5	12	196	98
164.5–179.5	4	200	100
	200		

Step 3 Graph the data, using class boundaries for the x axis and the percentages for the y axis, as shown in Figure 3–6.

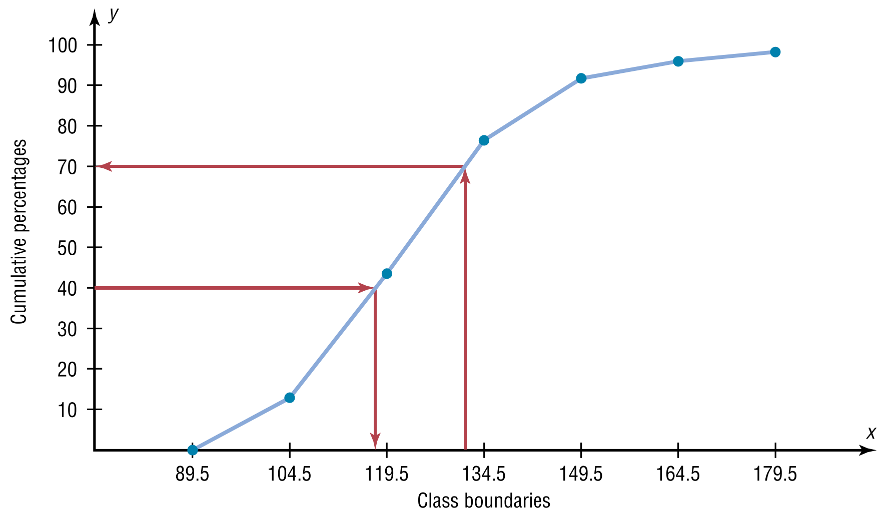
Once a percentile graph has been constructed, one can find the approximate corresponding percentile ranks for given blood pressure values and find approximate blood pressure values for given percentile ranks.

For example, to find the percentile rank of a blood pressure reading of 130, find 130 on the x axis of Figure 3–6, and draw a vertical line to the graph. Then move horizontally to the value on the y axis. Note that a blood pressure of 130 corresponds to approximately the 70th percentile.

If the value that corresponds to the 40th percentile is desired, start on the y axis at 40 and draw a horizontal line to the graph. Then draw a vertical line to the x axis and read

Figure 3–6

Percentile Graph for Example 3–31



the value. In Figure 3–6, the 40th percentile corresponds to a value of approximately 118. Thus, if a person has a blood pressure of 118, he or she is at the 40th percentile.

Finding values and the corresponding percentile ranks by using a graph yields only approximate answers. Several mathematical methods exist for computing percentiles for data. These methods can be used to find the approximate percentile rank of a data value or to find a data value corresponding to a given percentile. When the data set is large (100 or more), these methods yield better results. Examples 3–32 through 3–35 show these methods.

Percentile Formula

The percentile corresponding to a given value X is computed by using the following formula:

$$\text{Percentile} = \frac{(\text{number of values below } X) + 0.5}{\text{total number of values}} \cdot 100$$

Example 3–32

Test Scores

A teacher gives a 20-point test to 10 students. The scores are shown here. Find the percentile rank of a score of 12.

18, 15, 12, 6, 8, 2, 3, 5, 20, 10

Solution

Arrange the data in order from lowest to highest.

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Then substitute into the formula.

$$\text{Percentile} = \frac{(\text{number of values below } X) + 0.5}{\text{total number of values}} \cdot 100$$

Since there are six values below a score of 12, the solution is

$$\text{Percentile} = \frac{6 + 0.5}{10} \cdot 100 = 65\text{th percentile}$$

Thus, a student whose score was 12 did better than 65% of the class.

Note: One assumes that a score of 12 in Example 3–32, for instance, means theoretically any value between 11.5 and 12.5.

Example 3–33 Test Scores

Using the data in Example 3–32, find the percentile rank for a score of 6.

Solution

There are three values below 6. Thus

$$\text{Percentile} = \frac{3 + 0.5}{10} \cdot 100 = 35\text{th percentile}$$

A student who scored 6 did better than 35% of the class.

Examples 3–34 and 3–35 show a procedure for finding a value corresponding to a given percentile.

Example 3–34 Test Scores

Using the scores in Example 3–32, find the value corresponding to the 25th percentile.

Solution

Step 1 Arrange the data in order from lowest to highest.

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Step 2 Compute

$$c = \frac{n \cdot p}{100}$$

where
n = total number of values
p = percentile

Thus,

$$c = \frac{10 \cdot 25}{100} = 2.5$$

Step 3 If c is not a whole number, round it up to the next whole number; in this case, $c = 3$. (If c is a whole number, see Example 3–35.) Start at the lowest value and count over to the third value, which is 5. Hence, the value 5 corresponds to the 25th percentile.

Example 3–35

Using the data set in Example 3–32, find the value that corresponds to the 60th percentile.

Solution

Step 1 Arrange the data in order from smallest to largest.

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Step 2 Substitute in the formula.

$$c = \frac{n \cdot p}{100} = \frac{10 \cdot 60}{100} = 6$$

Step 3 If c is a whole number, use the value halfway between the c and $c + 1$ values when counting up from the lowest value—in this case, the 6th and 7th values.

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

↗ ↖

6th value 7th value

The value halfway between 10 and 12 is 11. Find it by adding the two values and dividing by 2.

$$\frac{10 + 12}{2} = 11$$

Hence, 11 corresponds to the 60th percentile. Anyone scoring 11 would have done better than 60% of the class.

The steps for finding a value corresponding to a given percentile are summarized in this Procedure Table.

Procedure Table

Finding a Data Value Corresponding to a Given Percentile

Step 1 Arrange the data in order from lowest to highest.

Step 2 Substitute into the formula

$$c = \frac{n \cdot p}{100}$$

where

n = total number of values

p = percentile

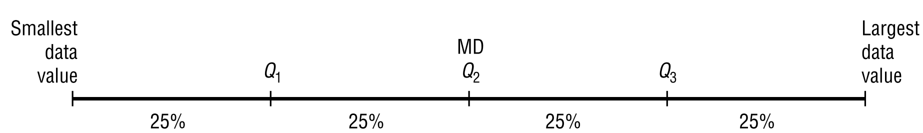
Step 3A If c is not a whole number, round up to the next whole number. Starting at the lowest value, count over to the number that corresponds to the rounded-up value.

Step 3B If c is a whole number, use the value halfway between the cth and $(c + 1)$st values when counting up from the lowest value.

Quartiles and Deciles

Quartiles divide the distribution into four groups, separated by Q_1, Q_2, Q_3.

Note that Q_1 is the same as the 25th percentile; Q_2 is the same as the 50th percentile, or the median; Q_3 corresponds to the 75th percentile, as shown:



Quartiles can be computed by using the formula given for computing percentiles on page 147. For Q_1 use $p = 25$. For Q_2 use $p = 50$. For Q_3 use $p = 75$. However, an easier method for finding quartiles is found in this Procedure Table.

Procedure Table

Finding Data Values Corresponding to Q_1, Q_2, and Q_3

Step 1 Arrange the data in order from lowest to highest.

Step 2 Find the median of the data values. This is the value for Q_2.

Step 3 Find the median of the data values that fall below Q_2. This is the value for Q_1.

Step 4 Find the median of the data values that fall above Q_2. This is the value for Q_3.

Example 3–36 shows how to find the values of Q_1, Q_2, and Q_3.

Example 3–36

Find Q_1, Q_2, and Q_3 for the data set 15, 13, 6, 5, 12, 50, 22, 18.

Solution

Step 1 Arrange the data in order.

5, 6, 12, 13, 15, 18, 22, 50

Step 2 Find the median (Q_2).

5, 6, 12, 13, 15, 18, 22, 50
↑
MD

$$\text{MD} = \frac{13 + 15}{2} = 14$$

Step 3 Find the median of the data values less than 14.

5, 6, 12, 13
↑
Q_1

$$Q_1 = \frac{6 + 12}{2} = 9$$

So Q_1 is 9.

Step 4 Find the median of the data values greater than 14.

15, 18, 22, 50
↑
Q_3

$$Q_3 = \frac{18 + 22}{2} = 20$$

Here Q_3 is 20. Hence, $Q_1 = 9$, $Q_2 = 14$, and $Q_3 = 20$.

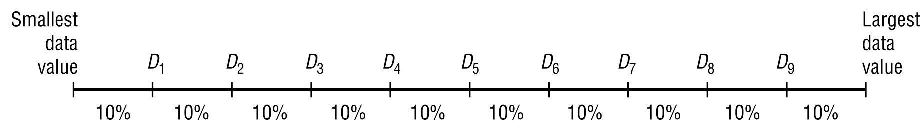
Unusual Stat

Of the alcoholic beverages consumed in the United States, 85% is beer.

In addition to dividing the data set into four groups, quartiles can be used as a rough measurement of variability. The **interquartile range (IQR)** is defined as the difference between Q_1 and Q_3 and is the range of the middle 50% of the data.

The interquartile range is used to identify outliers, and it is also used as a measure of variability in exploratory data analysis, as shown in Section 3–4.

Deciles divide the distribution into 10 groups, as shown. They are denoted by D_1, D_2, etc.



Note that D_1 corresponds to P_{10}; D_2 corresponds to P_{20}; etc. Deciles can be found by using the formulas given for percentiles. Taken altogether then, these are the relationships among percentiles, deciles, and quartiles.

Deciles are denoted by $D_1, D_2, D_3, \ldots, D_9$, and they correspond to $P_{10}, P_{20}, P_{30}, \ldots, P_{90}$.

Quartiles are denoted by Q_1, Q_2, Q_3 and they correspond to P_{25}, P_{50}, P_{75}.

The median is the same as P_{50} or Q_2 or D_5.

The position measures are summarized in Table 3–4.

Table 3–4 Summary of Position Measures

Measure	Definition	Symbol(s)
Standard score or z score	Number of standard deviations that a data value is above or below the mean	z
Percentile	Position in hundredths that a data value holds in the distribution	P_n
Decile	Position in tenths that a data value holds in the distribution	D_n
Quartile	Position in fourths that a data value holds in the distribution	Q_n

Outliers

A data set should be checked for extremely high or extremely low values. These values are called *outliers.*

An **outlier** is an extremely high or an extremely low data value when compared with the rest of the data values.

An outlier can strongly affect the mean and standard deviation of a variable. For example, suppose a researcher mistakenly recorded an extremely high data value. This value would then make the mean and standard deviation of the variable much larger than they really were. Outliers can have an effect on other statistics as well.

There are several ways to check a data set for outliers. One method is shown in this Procedure Table.

Procedure Table

Procedure for Identifying Outliers

Step 1 Arrange the data in order and find Q_1 and Q_3.

Step 2 Find the interquartile range: IQR $= Q_3 - Q_1$.

Step 3 Multiply the IQR by 1.5.

Step 4 Subtract the value obtained in step 3 from Q_1 and add the value to Q_3.

Step 5 Check the data set for any data value that is smaller than $Q_1 - 1.5(\text{IQR})$ or larger than $Q_3 + 1.5(\text{IQR})$.

This procedure is shown in Example 3–37.

Example 3–37

Check the following data set for outliers.

5, 6, 12, 13, 15, 18, 22, 50

Solution

The data value 50 is extremely suspect. These are the steps in checking for an outlier.

Step 1 Find Q_1 and Q_3. This was done in Example 3–36; Q_1 is 9 and Q_3 is 20.

Step 2 Find the interquartile range (IQR), which is $Q_3 - Q_1$.

$$\text{IQR} = Q_3 - Q_1 = 20 - 9 = 11$$

Step 3 Multiply this value by 1.5.

$$1.5(11) = 16.5$$

Step 4 Subtract the value obtained in step 3 from Q_1, and add the value obtained in step 3 to Q_3.

$$9 - 16.5 = -7.5 \qquad \text{and} \qquad 20 + 16.5 = 36.5$$

Step 5 Check the data set for any data values that fall outside the interval from -7.5 to 36.5. The value 50 is outside this interval; hence, it can be considered an outlier.

There are several reasons why outliers may occur. First, the data value may have resulted from a measurement or observational error. Perhaps the researcher measured the variable incorrectly. Second, the data value may have resulted from a recording error. That is, it may have been written or typed incorrectly. Third, the data value may have been obtained from a subject that is not in the defined population. For example, suppose test scores were obtained from a seventh-grade class, but a student in that class was actually in the sixth grade and had special permission to attend the class. This student might have scored extremely low on that particular exam on that day. Fourth, the data value might be a legitimate value that occurred by chance (although the probability is extremely small).

There are no hard-and-fast rules on what to do with outliers, nor is there complete agreement among statisticians on ways to identify them. Obviously, if they occurred as a result of an error, an attempt should be made to correct the error or else the data value should be omitted entirely. When they occur naturally by chance, the statistician must make a decision about whether to include them in the data set.

When a distribution is normal or bell-shaped, data values that are beyond 3 standard deviations of the mean can be considered suspected outliers.

Applying the Concepts 3–3

Determining Dosages

In an attempt to determine necessary dosages of a new drug (HDL) used to control sepsis, assume you administer varying amounts of HDL to 40 mice. You create four groups and label them *low dosage, moderate dosage, large dosage,* and *very large dosage*. The dosages also vary within each group. After the mice are injected with the HDL and the sepsis bacteria, the time until the onset of sepsis is recorded. Your job as a statistician is to effectively communicate the results of the study.

1. Which measures of position could be used to help describe the data results?
2. If 40% of the mice in the top quartile survived after the injection, how many mice would that be?
3. What information can be given from using percentiles?
4. What information can be given from using quartiles?
5. What information can be given from using standard scores?

See page 180 for the answers.

Exercises 3–3

1. What is a *z* score? A *z* score tells how many standard deviations the data value is above or below the mean.

2. Define *percentile rank.* A percentile rank indicates the percentage of data values that fall below the specific rank.

3. What is the difference between a percentage and a percentile? A percentile is a relative measurement of position; a percentage is an absolute measure of the part to the total.

4. Define *quartile.* A quartile is a relative measure of position obtained by dividing the data set into quarters.

5. What is the relationship between quartiles and percentiles? $Q_1 = P_{25}$; $Q_2 = P_{50}$; $Q_3 = P_{75}$

6. What is a decile? A decile is a relative measure of position obtained by dividing the data set into tenths.

7. How are deciles related to percentiles? $D_1 = P_{10}$; $D_2 = P_{20}$; $D_3 = P_{30}$; etc.

8. To which percentile, quartile, and decile does the median correspond? P_{50}; Q_2; D_5

9. Vacation Days If the average number of vacation days for a selection of various countries has a mean of 29.4 days and a standard deviation of 8.6, find the *z* scores for the average number of vacation days in each of these countries.

Canada	26 days	−0.40
Italy	42 days	1.47
United States	13 days	−1.91

Source: www.infoplease.com

10. Age of Senators The average age of Senators in the 108th Congress was 59.5 years. If the standard deviation was 11.5 years, find the *z* scores corresponding to the oldest and youngest Senators: Robert C. Byrd (D, WV), 86, and John Sununu (R, NH), 40. Byrd: $z = 2.30$ Sununu: $z = -1.70$

Source: *CRS Report for Congress.*

11. Driver's License Exam Scores The average score on a state CDL license exam is 76 with a standard deviation of 5. Find the corresponding *z* score for each raw score.

a. 79 0.6
b. 70 −1.2
c. 88 2.4
d. 65 −2.2
e. 77 0.2

12. Teacher's Salary The average teacher's salary in a particular state is $54,166. If the standard deviation is