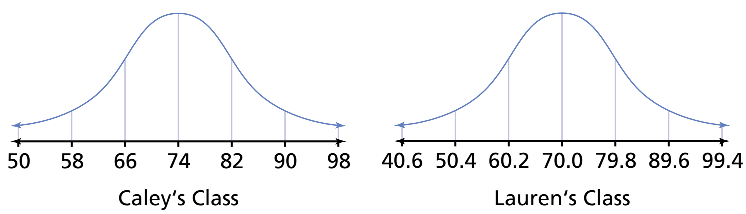
3.5 Applying the Normal Distribution: Z-Scores

In the previous section, you learned about the normal curve and the normal distribution. You know that the area under any normal curve is 1, and that 68% of the data is within one standard deviation of the mean, 95% is within two standard deviations, and almost all (99.7%) of the data is within three standard deviations. How can you use the normal curve to accurately determine the percent of data that lies above or below a given value? How can the normal curve be used to compare two different data from two different data sets?

Two students have been nominated for a $500 Data Management Mathematics award to be presented at graduation. Caley has a mark of 84 and Lauren has a mark of 83. Upon first glance, Caley should be given the award. However, other factors should be taken into consideration to compare the marks of these students. Caley's class has a mean of 74 and a standard deviation of 8, while Lauren's class has a mean of 70 and a standard deviation of 9.8. Assuming that the set of marks in both classes is normally distributed, a fair comparison cannot be made since both distributions are clearly different.

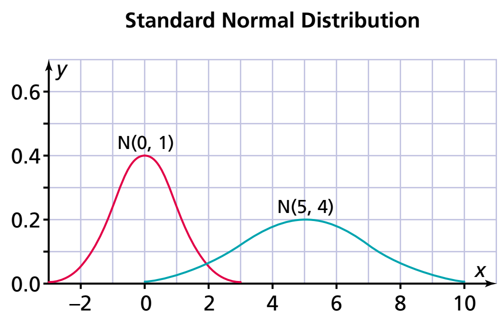


However, a comparison can be made if both students' marks are compared on a common distribution. This can be done using the standard normal distribution.

STANDARD NORMAL DISTRIBUTION

standard normal distribution—a special normal distribution with a mean of 0 and a standard deviation of 1

The normal curve is not the only bell-shaped curve, but it is the most useful one for statistics. A normal curve with a mean of 0 and a standard deviation of 1 is called **standard normal distribution**. As with all normal distributions, it has the property that the area under the whole curve is equal to 1.



The standard normal distribution is written as $X \sim N(0, 1^2)$. N(5, 4) would refer to a population that is normally distributed about a mean of 5 and with a standard deviation of 2. Each element of a normal distribution can be *translated* to the same place on a standard normal distribution by determining the number of standard deviations a given score lies away from the mean.

For a given score, x, from a normal distribution, you know that $x = \bar{x} + z\sigma$, where $\bar{x}$ and σ are the mean and standard deviation of the distribution, respectively. The value z is the number of standard deviations the score lies above or below the mean. Solving for z,

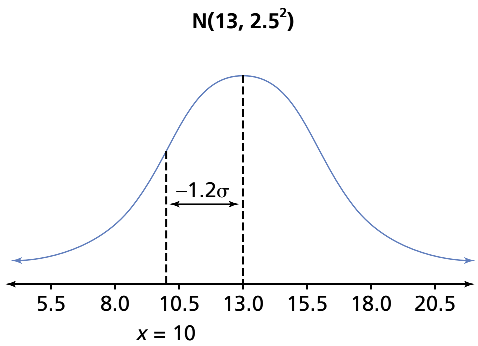
$$\frac{x - \bar{x}}{\sigma} = \frac{z\sigma}{\sigma}$$

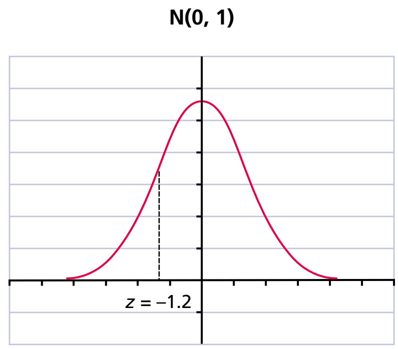
Therefore,

$$z = \frac{x - \bar{x}}{\sigma}$$

z-score—the number of standard deviations a given piece of data is above or below the mean

This new value, z, is called the **z-score** of a piece of data. A positive z-score indicates that the value lies above the mean and a negative z-score indicates that the value lies below the mean.





The graphs above show how the value $x = 10$ is 1.2 standard deviations below the mean. This value will also have the same position on the standard normal distribution shown to the right. Its z-score is calculated using the formula

$$\begin{aligned} z &= \frac{10 - 13}{2.5} \\ &= -1.2 \end{aligned}$$

The process of reducing a normal distribution to the standard normal distribution N(0, 1) is called *standardizing*. In this case, the value $x = 10$ has been standardized to N(0, 1). Remember, the standardized normal distribution has a mean of zero and a standared deviation of one.

Example 1 Calculating Z-Scores

For the distribution $X \sim N(14, 4^2)$, determine the number of standard deviations each piece of data lies above or below the mean.

(a) $X = 11$ **(b)** $X = 21.5$

Solution

(a) $z = \frac{11 - 14}{4}$
$= -0.75$

This piece of data is 0.75 standard deviations below the mean.

(b) $z = \frac{21.5 - 14}{4}$
$= 1.875$

This piece of data is 1.875 standard deviations above the mean.

Example 2 Comparing Data with Z-Scores

Caley scored 84 on her Data Management course, while Lauren, who attends a different Data Management class, scored 83. If Caley's class average is 74 with a standard deviation of 8, and Lauren's class average is 70 with a standard deviation of 9.8, use z-scores to determine who has the better mark.

Solution

To compare these values from two different normal distributions, you need to calculate the z-score for each student.

Caley

$$z = \frac{84 - 74}{8} \doteq 1.25$$

Lauren

$$z = \frac{83 - 70}{9.8} \doteq 1.326$$

Lauren's result is 1.326 standard deviations above the mean, while Caley's is 1.25 standard deviations above the mean. Lauren's result is slightly better.

Z-scores can be used to compare data values from different normal distributions, but they are more frequently used to estimate how many other pieces of data in a population are above or below a given value.

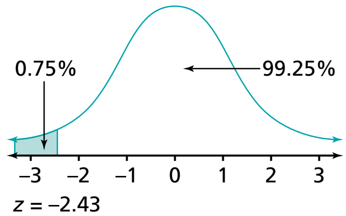
Z-SCORE TABLE

Technolink
The z-score table is included on the textbook CD, as well as in Appendix B.1 on page 398.

The z-score table shown below is used to find the proportion of data that has an equal or lesser z-score than a given value.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06
–2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015
–2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.002
–2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.002
–2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.003
–2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.005
–2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.006
–2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	[illegible]
–2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.01	
–2.1	0.0179				0.0162		

For example, if Patrick's error total in keyboarding class is 2.43 standard deviations below the mean, then his total has a z-score of −2.43. He can compare this value to the chart above, find −2.4 in the column on the left, and then move across to the 0.03 column. The z-score table states that only 0.75% of a normal distribution has a lower z-score.



Therefore, only 0.75% of the class has fewer errors. Similarly, 99.25% (1 − 0.0075) has more errors than Patrick.

Example 3 Using the Z-Score Table

Perch in a lake have a mean length of 20 cm and a standard deviation of 5 cm. Find the percent of the population that is less than or equal to the following lengths (the **percentile**).

percentile—the *k*th percentile is the least data value that is greater than *k*% of the population

(a) 22 cm **(b)** 16 cm **(c)** 28 cm **(d)** 4 cm

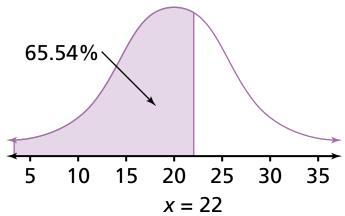
Solution 1 *Without technology*

	0.00	0.01	0.0
0.0	0.5000	0.5040	0.50
0.1	0.5398	0.5438	0.54
0.2	0.5793	0.5832	0.58
0.3	0.6179	0.6217	0.62
0.4	0.6554	0.6591	0.6
0.5	0.6915	0.6	

(a) The z-score for this length is

$$z = \frac{22 - 20}{5} = 0.40.$$

Checking the z-score table, you see that 0.6554 or 65.54% of the data is equal to or less than this z-score value.



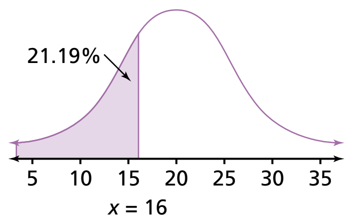
Of the fish in the lake, 65.54% are 22 cm long or less. This fish is in the 66th percentile.

–1.0	0.1587	0.1562
–0.9	0.1841	0.1814
–0.8	0.2119	0.2090
–0.7	0.2420	0.2389
–0.6	0.2743	0.2709
–0.5	0.3085	0.3050
–0.4	0.3446	0.3409
–0.3	0.3821	

(b) The z-score for this length is

$$z = \frac{16 - 20}{5} = -0.80.$$

Checking the z-score table, you see that 0.2119 or 21.19% of the data is equal to or less than this z-score value.



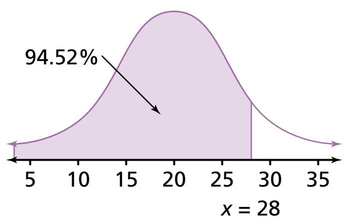
Of the fish in the lake, 21.19% are 16 cm long or less. This fish is in the 21st percentile.

1.3	0.9032	0.9049	0.
1.4	0.9192	0.9207	0.92
1.5	0.9332	0.9345	0.935
1.6	0.9452	0.9463	0.947
1.7	0.9554	0.9564	0.95
1.8	0.9641	0.9649	0.96
1.9	0.9713	0.9719	0.97

(c) The z-score for this length is

$$z = \frac{28 - 20}{5} = 1.60.$$

Checking the z-score table, you see that 0.9452 or 94.52% of the data is equal to or less than this z-score value.



Of the fish in the lake, 94.52% are 28 cm long or less. This fish is in the 95th percentile.

(d) The z-score for this length is $z = \frac{4 - 20}{5} = -3.20$.

Checking the z-score table, you notice that the values only go down to −2.99. Any value whose z-score is less than −3 is considered an outlier. Its percentile is considered to be 0%. Likewise, any variable whose z-score is greater than 2.99 is also considered an outlier. Its percentile is considered to be 100%.

Solution 2 *Using a TI-83 Plus calculator*

Technolink For more information on using **normalcdf(**, see Appendix C.8 on page 406.

Once you have calculated the z-scores manually, you can use the **normalcdf(** function to calculate the proportion of data between that z-score and the left most extreme (use −1E99, which means -1×10^{99}).

Press [2nd] [VARS] [2] and then enter the parameters as shown.

(a) normalcdf(−1E99, 0.4) = 0.6554

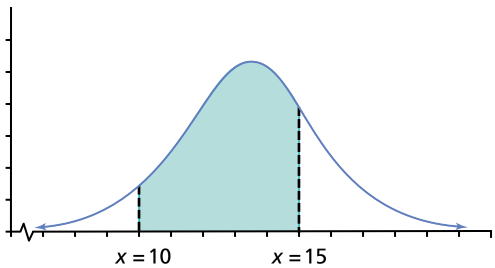
(b) normalcdf(−1E99, −0.8) = 0.2119

(c) normalcdf(−1E99, 1.6) = 0.9452

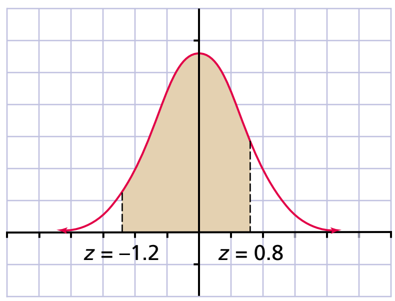
(d) normalcdf(−1E99, −3.2) = 0.0007

Using your problem-solving skills, you can also determine the percent of the population that lies between two values.

N(13, 2.5)

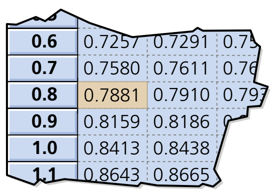


N(0, 1)

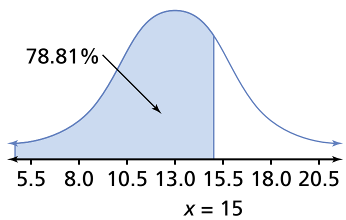


What percent of the data lies between $x = 10$ and $x = 15$? If you first find the percent of data that is less than 15 ($z = 0.8$), you can then subtract the percent of data that is less than 10 ($z = -1.2$).

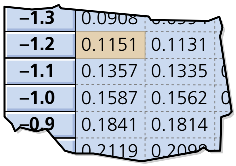
The z-score table shows that 78.81% of the data is to the left of the x-value 15.

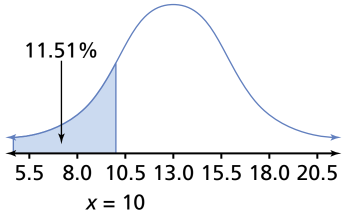


0.6	0.7257	0.7291	0.75
0.7	0.7580	0.7611	0.76
0.8	0.7881	0.7910	0.797
0.9	0.8159	0.8186	0.
1.0	0.8413	0.8438	
1.1	0.8643	0.8665	



The z-score table shows that 11.51% of the data is to the left of the x-value 10.





Therefore, the difference between the two percents that lie below $x = 10$ and $x = 15$ is $78.81\% - 11.51\% = 67.3\%$. Therefore, 67.3% of the data lies between $x = 10$ and $x = 15$.

Example 4 Using the Z-Score Table

Using the normal distribution $X \sim N(7, 2.2^2)$, find the percent of data that is within the given intervals.

(a) $3 < X < 6$ **(b)** $7 < X < 15$

Solution 1 *Without technology*

(a) For $x = 3$, $z = \frac{3 - 7}{2.2} \doteq -1.81$.

For $x = 6$, $z = \frac{6 - 7}{2.2} \doteq -0.45$.

By subtracting the corresponding z-score values, you obtain $0.3264 - 0.0351 = 0.2913$. Therefore, 29.13% of the data fills this interval.

(b) For $x = 7$, $z = \frac{7 - 7}{2.2} = 0.0$.

For $x = 15$, $z = \frac{15 - 7}{2.2} \doteq 3.6$.

The z-score for $z = 3.6$ is off the chart. Its percentile can be considered equal to 100%. By definition, the z-score for $z = 0$ must be 50%. Therefore, 50% of the data lies in this interval.

Solution 2 *Using a TI-83 Plus calculator*

Press [2nd] [VARS] [2] and key in the z-score values to determine the percent of data between the two given z-score values.

(a) normalcdf$(-1.81, -0.45) = 0.2912$

(b) normalcdf$(0, 3.6) = 0.4998$

Technolink
Solution 2
You can also find the answer without calculating the z-scores by evaluating the following:
(a) normalcdf(3, 6, 7, 2.2)
(b) normalcdf(7, 15, 7, 2.2)

KEY IDEAS

standard normal distribution—symmetrical, approaching zero at the extremes; has a mean of 0 and a standard deviation of 1 N(0, 1); total area under the curve is equal to 1; used to compare data from different data sets

z-score—the number of standard deviations a given piece of data is above or below the mean; calculated using the formula $z = \frac{x - \bar{x}}{\sigma}$

z-score table—lists proportion of data values, in a normal distribution, with an equal or smaller z-score; normalcdf(*lowerbound*, *upperbound*) performs the same function on a TI-83 Plus calculator; found in Appendix B.1 on page 398

percentile—*k*th percentile is the least data value that is greater than *k*% of the population

3.5 Exercises

A

1. **Knowledge and Understanding** Calculate a z-score for each *x*-value, correct to one decimal place, given the mean and standard deviation provided.

(a) $\bar{x} = 6, \sigma = 2$

(i) $x = 5.3$ **(ii)** $x = 7.2$ **(iii)** $x = 9.9$ **(iv)** $x = 0.8$

(b) $\bar{x} = 75, \sigma = 4$

(i) $x = 65.5$ **(ii)** $x = 77.9$ **(iii)** $x = 86.0$ **(iv)** $x = 70.7$

(c) $\bar{x} = 24, \sigma = 8$

(i) $x = 20.1$ **(ii)** $x = 5.5$ **(iii)** $x = 37.9$ **(iv)** $x = 8.0$

(d) $\bar{x} = 6.6, \sigma = 2.5$

(i) $x = 8.0$ **(ii)** $x = -0.4$ **(iii)** $x = 10.6$ **(iv)** $x = 6.7$

2. Which of the following statements are properties of a standard normal distribution? Explain.

(a) The area under the curve is infinite.

(b) The mean is 0 and the standard deviation is 1.

(c) The mean, median, and mode are approximately equal.

(d) Standard notation for a standard normal distribution is N(0, 1).

3. Tanya's baby brother has a length of 55 cm. This puts him in the 95th percentile. In a group of 800 babies, how many of them would have a length less than Tanya's brother if the lengths are normally distributed?

4. Using the z-score table, find the percentile that corresponds to each of the following z-scores.

(a) $z = 0.44$ **(b)** $z = 2.33$ **(c)** $z = -0.83$ **(d)** $z = -1.85$

5. Using the z-score table, find the z-score that corresponds to each of the following percentiles.

(a) 45^{th} **(b)** 73^{rd} **(c)** 7^{th} **(d)** 98^{th}

6. Given a normally distributed data set whose mean is 50 and whose standard deviation is 10, what value of x would each of the following z-scores have?

(a) $z = 1.00$ **(b)** $z = -1.00$ **(c)** $z = 2.50$
(d) $z = -0.50$ **(e)** $z = -1.81$ **(f)** $z = 0.20$
(g) $z = 1.62$ **(h)** $z = -2.24$

B

7. Adrian's average bowling score is 174, and is normally distributed with a standard deviation of 35.

(a) What z-score corresponds with the following scores?
(i) 180 points **(ii)** 264 points

(b) In what percent of games does Adrian score less than 200 points? More than 200 points?

(c) The top 10% of bowlers in Adrian's league get to play in an all-star game. If the league average is 170, with a standard deviation of 11 points, and is normally distributed what average score does Adrian need to have to obtain a spot in the all-star game?

Think about IQ Scores

How many geniuses are there in Toronto? What problems would you run into answering this question?

8. IQ scores of people around the world are normally distributed, with a mean of 100 and a standard deviation of 15. A genius is someone with an IQ greater than or equal to 140. What percent of the population is considered genius?

9. The number of red blood cells (in millions per cubic microlitre) is normally distributed, with a mean of 4.8 and a standard deviation of 0.3.

(a) What percent of people have a red blood cell count of less than 4?
(b) What percent of people have a count between 4.7 and 5.0?
(c) To be in the top 5%, what count would someone need to have?

10. **(a)** A student's score is 675 on a standardized test known to be normally distributed with a mean of 500 and a standard deviation of 110. What percentile is she in?

(b) Another student taking the same test wants to score in the 90^{th} percentile. What score must he get?

11. Each 450-g box of cereal is routinely returned if its mass has a z-score of -2.7 or less. Research has shown that the standard deviation of masses is 8 g and is normally distributed.

(a) What is the minimum-sized cereal box that is not returned?
(b) What percent of cereal boxes is returned?

12. The weights of 75 model planes at a local convention are normally distributed. The average weight is 4.4 kg, with a standard deviation of 0.41 kg.

(a) How many planes have a mass less than 4 kg?
(b) How many planes would be disqualified if it were against the rules to have a plane with a mass of more than 5.5 kg?
(c) How many planes have a mass between 3.5 kg and 5 kg?

13. **Communication** The mean temperature in Collingwood is normally distributed and is 21.7 °C in July, with a standard deviation of 3.1 °C. If your friend spent the weekend in Collingwood and said that the temperature was more than 30 °C all three days, would you believe him? Explain.

C **14.** Mr. Median is a very precise teacher. Each class he teaches *has* to be normally distributed with a class average of 71 and a standard deviation of 11. What would the quartiles of Mr. Median's class be?

15. A snake farm advertises that 25% of their snakes are longer than 2.5 m and 10% of them are longer than 3 m. The lengths are normally distributed. What is the mean length of snakes at this farm? What is the standard deviation?

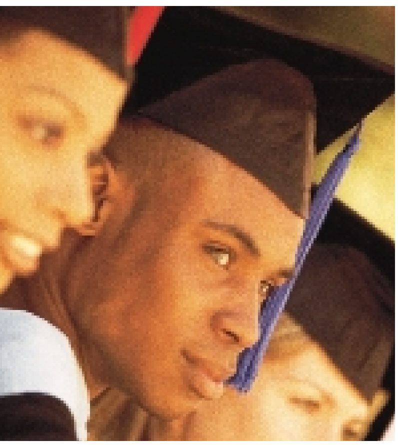
ADDITIONAL ACHIEVEMENT CHART QUESTIONS

16. **Knowledge and Understanding** For the distribution $X \sim N(14, 3^2)$, calculate the corresponding z-scores for the following *x*-values.
(a) $x = 11.5$ **(b)** $x = 17$ **(c)** $x = 20.4$ **(d)** $x = 13.2$

17. **Application** For the distribution N(3, 0.55), determine the percent of the data that is within the given interval.
(a) $X > 2.44$ **(b)** $1.8 < X < 2.3$ **(c)** $X < 1.91$

18. **Thinking, Inquiry, Problem Solving** On the final exam, Jalice's class had an average of 61.1% with a standard deviation of 11.4%.
(a) If there are 27 students in the class, how many of them scored less than 50%?
(b) If the teacher were to adjust everyone's grade so that the class average is 65% with a standard deviation of 11.4%, how many would score less than 50%?

19. **Communication** The teacher of a Data Management class, in which the class average is 68% with a standard deviation of 8.5%, has offered to help up to 10% of the students after school with their projects. Using the normal distribution, how should the teacher decide who will get help? Explain.



Chapter Problem

Comparing Marks

Refer to the data on page 140.

CP13. Calculate a z-score for Justin's average mark using the mean and standard deviation you used in Section 3.3.
CP14. What percentile is Justin in?
CP15. Using the z-score table, what percent of the class should have a lower average mark than Justin? Is this true? Why or why not?
CP16. What would Justin have to raise his average mark to if he wanted to be in the 90th percentile?