Standard Scores, the Normal Distribution, and Probability

4

LEARNING OBJECTIVES

- Transform raw scores into standard scores (z scores) and vice versa.
- Describe the normal curve.
- Transform raw scores and standard scores into percentile ranks and vice versa.
- Calculate the probability of an outcome falling in a specified area under the normal curve.

CHAPTER OVERVIEW

This chapter covers standard scores, normal curves, and probability. Most people know a little about bell-shaped curves, what statisticians call normal curves. This chapter explains why normal curves are called normal, what their characteristics are, and how they are used in statistics.

Standard scores, also called z scores, are new territory for most students. Standard scores are useful in statistics because raw scores can be transformed into standard scores (and vice versa). Standard scores allow different kinds of measurements to be compared on a common scale—such as the juiciness of an orange vs. the crispiness of an apple. Standard scores are also useful for measuring distance on the normal curve, which shows how big the pieces of the bell curve are in terms of what percentage of cases fall in different sections of it. This allows scores to be expressed as percentile ranks, which tell what percentage of cases fall at or below a given score. Breaking the normal curve into chunks allows researchers to talk about how common or rare different scores should be, how probable they are. This excursion into probability opens the way to understanding how statistical tests work in upcoming chapters.

4.1 Standard Scores (z Scores)

People often say that you cannot compare apples and oranges. In this chapter, we'll see this can be done thanks to something called a standard score. We'll also see that standard scores allow one to quantify how common or how rare a score is.

Imagine that two fruit growers, Anne who grows apples and Oliver who grows oranges, get into a good-natured argument over whose fruit is better. Anne claims

Corty_3e_Ch04_107-144_Final HR.indd 107 21/01/16 5:23 pm

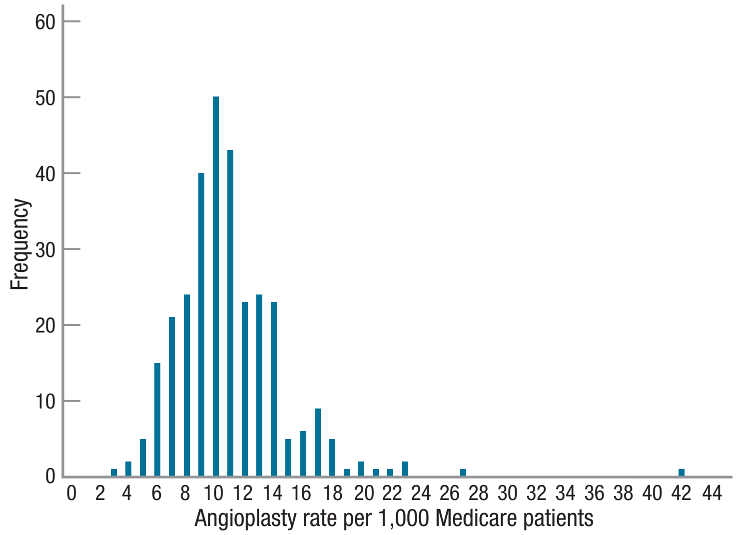

Figure 4.1 Angioplasty Rates for Medicare Patients by City There is one city, far out on the right-hand side that is very different from the others. A standard score, a *z* score, will quantify how unusual this score is. (Data from *The New York Times*, August 18, 2006.)

that the crispiness of apples is what makes them so good. Oliver cites the juiciness of oranges as their best feature. How can one objectively compare the crispiness of an apple to the juiciness of an orange to decide which fruit has more of its desired quality? Standard scores.

Look at the graph in Figure 4.1. It shows the rate with which Medicare patients receive angioplasty in different geographic hospital regions of the United States. It should be apparent that most hospitals cluster around rates of 6 per thousand patients to 14 per thousand patients. And then, there is one city, way off by itself on the right-hand side, treating an unusually large percentage of Medicare patients with angioplasty. How can we describe how unusual that city is? Standard scores.

A **standard score** is a raw score expressed in terms of how many standard deviations it is away from the mean. Standard scores are commonly called ***z* scores**. The formula for calculating a z score is shown in Equations 4.1 and 4.2.

Equation 4.1 Formula for Calculating Standard Scores (z Scores) for a Population

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

where z = the standard score
X = the raw score
μ = the population mean
σ = the population standard deviation

Equation 4.1 is used when the population mean, μ, and population standard deviation, σ, are known. As it is relatively rare that they are known, Equation 4.2 is more commonly used.

Equation 4.2 Formula for Calculating Standard Scores (z Scores) for a Sample

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

where z = the standard score
X = the raw score
M = the population mean
s = the sample standard deviation

Each of these two formulas has two steps:

Step 1 Subtract the mean from the raw score, just as was done in Chapter 3 when calculating deviation scores. Be sure to keep track of the sign, whether the deviation score is positive, negative, or zero.

Step 2 Divide the deviation score by the standard deviation and round to two decimal places. Again, be sure to keep track of the sign, whether the z score is positive, negative, or zero.

Chapter 3 introduced a sample of five Americans where the mean height, in inches, was 67.00 and the standard deviation was 4.18. One of the members of this sample was 62″ tall. For this person, one would calculate the standard score like this:

$$\begin{aligned} z &= \frac{X - M}{s} \\ &= \frac{62 - 67.00}{4.18} \\ &= \frac{-5.0000}{4.18} \\ &= -1.1962 \\ &= -1.20 \end{aligned}$$

This person's z score was -1.20, meaning that his or her score fell 1.20 standard deviations *below* the mean. The scores for all five members of the sample, as z scores, are shown in Table 4.1.

TABLE 4.1 Heights of Five Americans, Expressed as Standard Scores

Height (X)	Deviation Score ($X - M$)	z Score $\left(\frac{X - M}{s}\right)$
62	−5.00	−1.20
65	−2.00	−0.48
66	−1.00	−0.24
69	2.00	0.48
73	6.00	1.44
	$\Sigma = 0.00$	$\Sigma = 0.00$

In this data set, $M = 67.00$ and $s = 4.18$. Standard scores, or z scores, tell the distance from the mean in a standard unit of measurement, the standard deviation. Since z scores are transformed deviation scores, they add up to zero for a data set.

There are two things to note in this table:

1. The z scores quickly show where a person's score falls in relation to the mean. As shown in Figure 4.2, positive z scores indicate a score above the mean, negative z scores a score below the mean, and a z score of 0 is a score right at the mean.
2. If the z scores for a data set are added together, they sum to zero.

z scores standardize *scores. They transform scores to a common unit of measurement, the standard deviation. This allows researchers to compare variables that are measuring different things on different scales.*

The same information found in z scores is contained in the regular deviation scores that are shown in the middle column of Table 4.1. So, what is the advantage of z scores? z scores *standardize* scores. They transform scores to a common unit of measurement, the standard deviation. This allows researchers to compare variables that are measuring different things on different scales.

For example, consider Carlos, who is very smart and very extroverted. Which does Carlos have more of, intelligence or extroversion? To answer this question, Carlos takes two tests. One is an intelligence test and the other is an extroversion scale. Carlos gets a score of 130 on the IQ test and a 75 on the extroversion scale. Do these two scores answer the question of whether Carlos is more unusual in terms of his intelligence or his extroversion?

These scores are out of context, so the question can't be answered. Knowing the means will help put the scores in context. The mean for the intelligence test is 100, and the mean for the extroversion scale is 50. With the means, one can tell that Carlos is above average in terms of intelligence and in terms of extroversion. The next step is to calculate the deviation scores to see how far above average he is. Carlos's deviation score for intelligence is $130 - 100 = 30.00$. For extroversion, his deviation score is $75 - 50 = 25.00$. Based on the deviation scores, it seems as if Carlos is more unusual in terms of intelligence than he is in terms of extroversion. After all, he scored 30 points above the mean on intelligence, but only 25 points above the mean on extroversion. But that would be a premature conclusion. A true comparison can't be made until standard deviation information is considered and z scores are calculated.

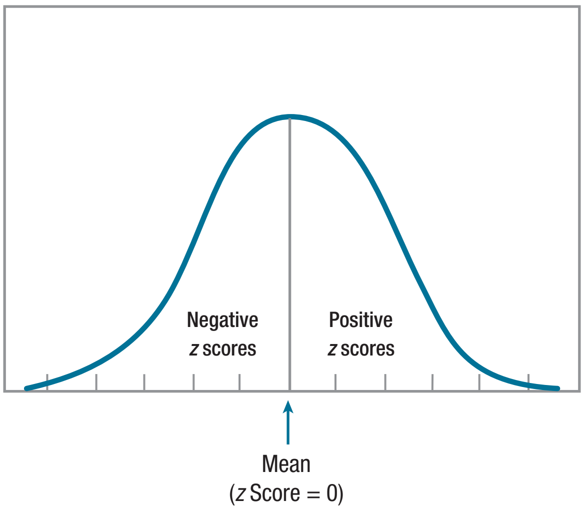

Figure 4.2 Positive, Negative, and Zero z Scores Another term for standard score is *z score*. A positive value for a z score means that the score falls above the mean, a negative value means that the score falls below the mean, and a value of 0 means that the score falls right at the mean.

Here are the standard deviations: 15 for the IQ test and 10 for the extroversion scale. Now, Equation 4.2 can be used to calculate z scores:

$$\begin{aligned} z_{\text{IQ}} &= \frac{130 - 100}{15} \\ &= \frac{30.0000}{15} \\ &= 2.0000 \\ &= 2.00 \\ Z_{\text{Extroversion}} &= \frac{75 - 50}{10} \\ &= \frac{25.0000}{10} \\ &= 2.5000 \\ &= 2.50 \end{aligned}$$

Now the question can be answered using a standard unit, the amount of variability in a set of scores in terms of standard deviation units. Carlos's intelligence test score falls 2 standard deviations above the mean, but his extroversion score falls 2.5 standard deviations above the mean. As shown in **Figure 4.3**, Carlos deviates more from the mean in terms of extroversion than he does in terms of intelligence. His level of extroversion is more unusual than his level of intelligence.

With the z scores, two different variables measured on two different scales can be compared in a statement like this, "Carlos is more above average in his level of extroversion than he is in his level of intelligence." The comparison can be made because the scores have been standardized.

In the same way, the apple and orange example from the start of the chapter would work. If there were a juiciness scale, Oliver's orange could be measured. And, with a crispiness scale, Anne's apple could be measured. Then, each measurement would be converted to a standard score and the one with the higher score would be the winner.

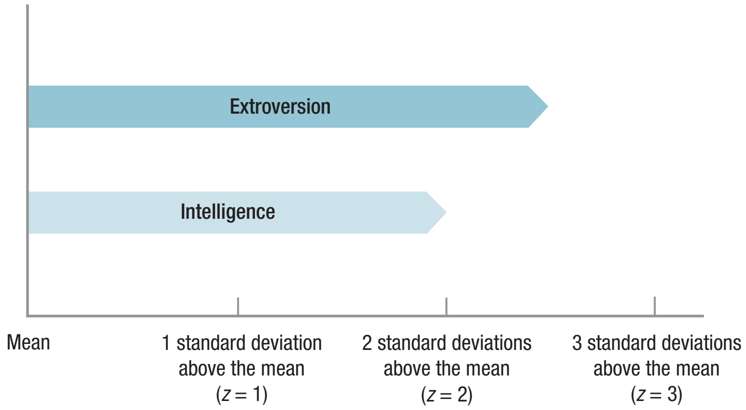

Figure 4.3 Carlos's Scores on an Extroversion Scale and an Intelligence Test
When the extroversion and intelligence scores are transformed into a standard unit of measurement, z scores, it becomes clear that Carlos's score on the extroversion scale is higher than his score on the intelligence test.

Just as raw scores can be turned into standard scores, it is possible to reverse direction and turn z scores into raw scores. Suppose Mei-Li was also extroverted and that her z score was 1.00 on the extroversion scale. What was her raw score?

Here's what to do. A z score of 1.00 is a positive score, which indicates that her score was above the mean. The mean on the extroversion scale is 50.00, so her score will be greater than this. A z score of 1.00 says that Mei-Li's score was 1 standard deviation above the mean. A standard deviation on the extroversion scale is 10.00, so her score was 10 points above the mean. Adding 10.00 (1 standard deviation) to 50.00 (the mean) leads to the conclusion that Mei-Li's score on the test was 60.00. Equation 4.3 formalizes how to turn z scores into raw scores.

Equation 4.3 Formula for Calculating a Raw Score (X) from a Standard Score (z)

$$X = M + (z \times s)$$

where X = the raw score
M = the mean
z = the standard score for which the raw score is being calculated
s = the standard deviation

Following this formula, here is how to calculate Mei-Li's score:

$$\begin{aligned} X &= M + (z \times s) \\ &= 50.00 + (1.00 \times 10.00) \\ &= 50.00 + 10.0000 \\ &= 60.0000 \\ &= 60.00 \end{aligned}$$

Here's one more example of converting a z score into a raw score, this one with a negative z score. Suppose that Tabitha's z score on the IQ test was -0.75. What was her raw score? Using Equation 4.3, here are the calculations:

$$\begin{aligned} X &= 100.00 + (-0.75 \times 15.00) \\ &= 100.00 + (-11.2500) \\ &= 100.00 - 11.2500 \\ &= 88.7500 \\ &= 88.75 \end{aligned}$$

On the IQ test, Tabitha's score was 88.75. Note that the raw score is reported to two decimal places and does not have to be a whole number.

Worked Example 4.1

SAT subtests are handy for practice in converting from raw scores to z scores and vice versa. As of March 2016, there are two parts on the SAT: a subtest that measures math, and a subtest that measures reading and writing. Both subtests are scored on the same scale, where 500 is the average (mean) score and the standard deviation is 100. Suppose a high school senior takes the SAT and gets a 460 on the reading and writing subtest. What is her score as a standard score?

Applying Equation 4.2, here are the calculations:

$$\begin{aligned} z &= \frac{460 - 500}{100} \\ &= \frac{-40.0000}{100} \\ &= -0.4000 \\ &= -0.40 \end{aligned}$$

She scored 0.40 standard deviations below the mean, so her z score would be reported as -0.40.

Suppose she had a friend who took the SAT and did very well on the math section. Her friend's score on the math section, expressed as a z score, was 1.80. What was her friend's score?

Applying Equation 4.3, here are the calculations:

$$\begin{aligned} X &= 500 + (1.80 \times 100) \\ &= 500 + 180.0000 \\ &= 680.0000 \\ &= 680.00 \end{aligned}$$

This student, who had scored 1.80 standard deviations above the mean, obtained a score of 680.00 on the math subtest of the SAT.

Practice Problems 4.1

Review Your Knowledge

4.01 In what units do z scores express raw scores?

4.02 What is the sum of all z scores in a data set?

4.03 If a z score is positive, what does that mean?

Apply Your Knowledge

4.04 The average GPA in a class is 2.75 with a standard deviation of 0.40. One student has a GPA of 3.20. Express this GPA as a z score?

4.05 Another student's GPA, expressed as a z score, is -2.30. What is his GPA?

4.2 The Normal Distribution

Now that calculating and interpreting z scores have been covered, it is time to explore the normal distribution. The **normal distribution** (often called the bell curve) is a specific symmetrical distribution whose highest point occurs in the middle and whose frequencies decrease as the values on the X-axis move away from the midpoint. An example of a normal distribution is shown in Figure 4.4.

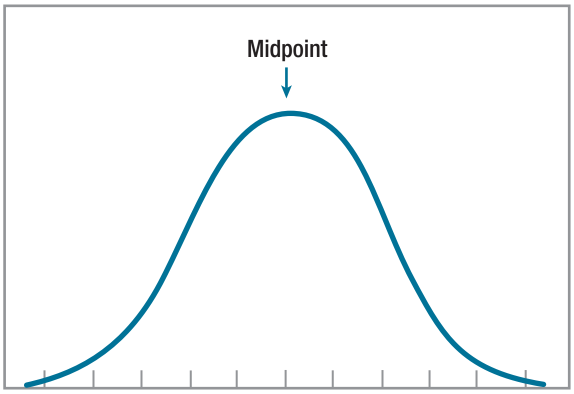

Figure 4.4 The Normal Distribution In a normal distribution, the midpoint is the mean, median, and mode. As scores move away from the midpoint, the frequency of their occurrence decreases symmetrically.

There are several things to note about the shape of a normal distribution:

- The frequencies decrease as the curve moves away from the midpoint, so the midpoint is the mode.
- The distribution is symmetric, so the midpoint is the median—half the cases fall above the midpoint and half fall below it.
- The fact that the distribution is symmetrical also means that the midpoint is the mean. It is the spot that balances the deviation scores.

Figure 4.5 shows a variety of bell-shaped curves that meet these criteria, but not all of them are normal distributions. There's another criterion that must be met for a bell-shaped curve to be considered a normal distribution: a specific percentage of cases has to fall in each region of the curve.

In Figure 4.6, the regions under the normal curve are marked off in *z* score units. In the normal distribution, 34.13% of the cases fall from the mean to 1 standard deviation above the mean, 13.59% in the next standard deviation, and decreasing percentages in subsequent standard deviations. Because the normal distribution is symmetrical, the same percentages fall in the equivalent regions below the mean. Figure 4.7 shows the percentages of cases that fall in each standard deviation unit of the normal distribution as the curve moves away from the mean.

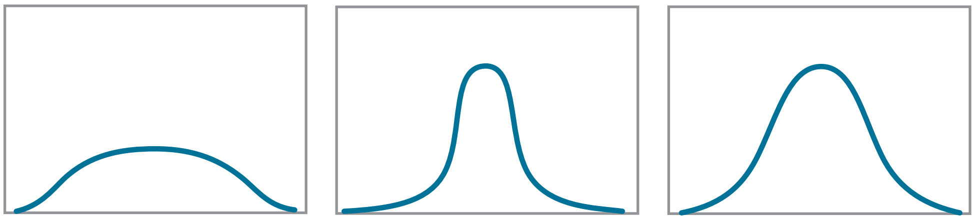

Figure 4.5 Not All Bell-Shaped Curves Are Normal Distributions In bell-shaped curves, the midpoint is the mean, median, and mode, and frequencies decrease symmetrically as values move away from the midpoint. Though all the curves shown here are bell-shaped, not all bell-shaped curves are normal distributions. The normal distribution is a specific bell-shaped curve, defined by the percentage of cases that fall within specified regions. Only the last curve is a normal distribution.

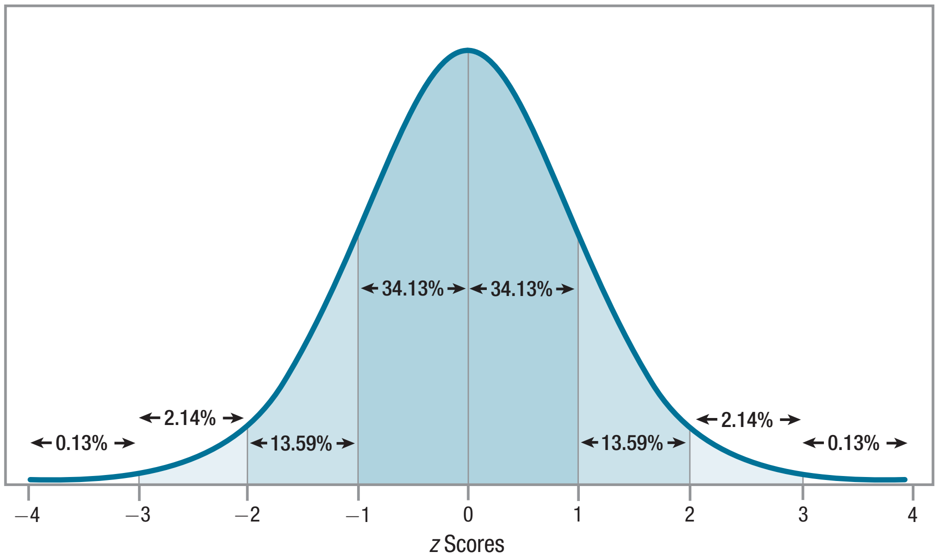

Figure 4.6 Percentage of Cases Falling in Specified Regions of the Normal Distribution The normal distribution is defined by the percentage of cases that fall within specified regions. This figure shows the percentage of cases that fall in each standard deviation as one moves away from the mean. Note the symmetrical nature and how the percentage of cases drops off dramatically as one moves away from the mean. Theoretically, the curve extends out to infinity, with fewer and fewer cases in each standard deviation.

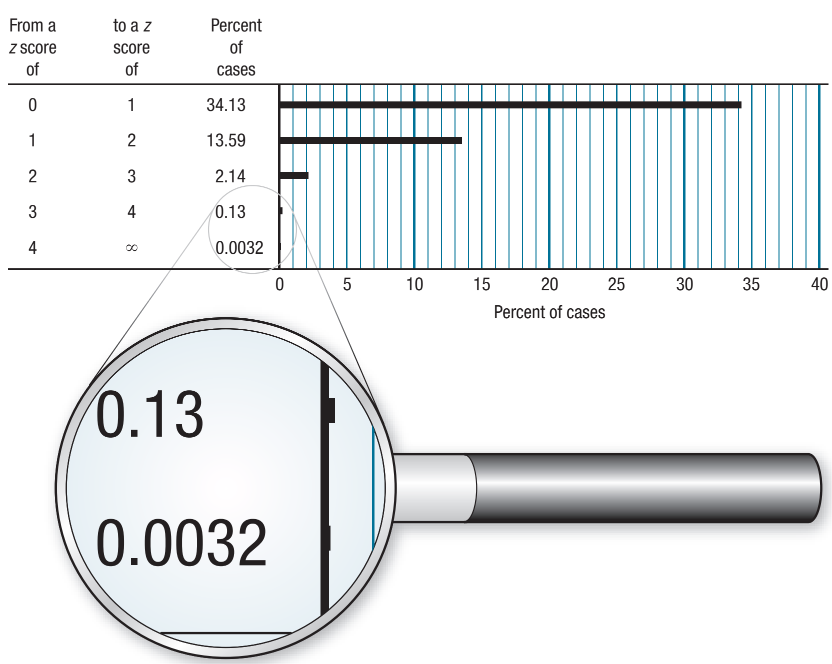

Figure 4.7 Percentage of Cases Falling in Each Standard Deviation of the Normal Distribution as It Moves Away from the Mean Note how quickly the percentage of cases falling in each standard deviation decreases as one moves away from the mean. Only a very small percentage of cases falls more than 3 standard deviations away from the mean.

Corty_3e_Ch04_107-144_Final HR.indd 115 21/01/16 5:23 pm

One important thing to note in Figure 4.7 is how quickly the percentage of cases in a standard deviation drops off as the curve moves away from the mean. The first standard deviation contains about 34% of the cases, the next about 14%, then about 2% in the third, and only about 0.1% in the fourth standard deviation. This distribution has an important implication for where the majority of cases fall:

- About two-thirds of the cases fall within 1 standard deviation of the mean. That's from 1 standard deviation below the mean to 1 standard deviation above it, from $z = -1.00$ to $z = 1.00$. (The exact percentage of cases that falls in this region is 68.26%.)
- About 95% of the cases fall within 2 standard deviations of the mean, from $z = -2.00$ to $z = 2.00$. (The exact percentage is 95.44%.)
- More than 99% of the cases fall within 3 standard deviations, from $z = -3.00$ to $z = 3.00$. In fact, close to 100% of the cases—almost all—fall within 3 standard deviations of the mean. (The exact percentage is 99.73%.)

It is rare, in a normal distribution, for a case to have a score that is more than 3 standard deviations away from the mean. In practical terms, these scores don't happen, so when graphing a normal distribution, setting up the z scores on the X-axis so they range from -3.00 to $z = 3.00$ is usually sufficient.

Here's a simplified version of the normal distribution. Look at **Figure 4.8** and note the three numbers: 34, 14, and 2. These are the approximate percentages of cases that fall in each standard deviation as the curve moves away from the mean:

- ≈34% of cases fall from the mean to 1 standard deviation above it. The same is true for the area from the mean to 1 standard deviation below it. (This symbol, ≈, means approximately.)

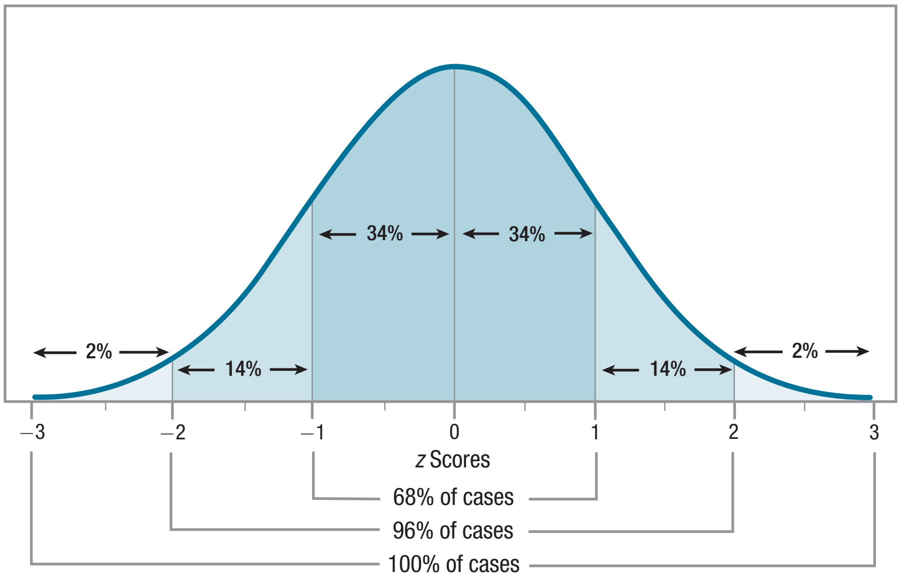

Figure 4.8 Simplified Version of the Normal Distribution This simplified version of the normal distribution takes advantage of the fact that almost all cases fall within 3 standard deviations of the mean and limits the normal distribution to z scores ranging from -3 to $+3$. It uses rounded percentages (34%, 14%, and 2%) for the percentage of cases falling in each standard deviation unit.

- ≈14% of cases fall from 1 standard deviation above the mean to 2 standard deviations above it. Ditto for the area from 1 standard deviation below to 2 standard deviations below.
- ≈2% of cases fall from 2 standard deviations above the mean to 3 standard deviations above it. Again, ditto for the same area below the mean.

Here's a real example of how scores only range from 3 standard deviations below the mean to 3 standard deviations above it. Think of the SAT where subtests have a mean of 500 and a standard deviation of 100. A z score of -3.00 is a raw score of 200 on a subtest. This can be calculated using Equation 4.3:

$$\begin{aligned} X &= 500 + (-3.00 \times 100) \\ &= 500 + (-300.0000) \\ &= 500 - 300.0000 \\ &= 200.0000 \\ &= 200.00 \end{aligned}$$

A z score of 3.00 is a raw score of 800 on a subtest. This can be calculated using Equation 4.3:

$$\begin{aligned} X &= 500 + (3.00 \times 100) \\ &= 500 + 300.0000 \\ &= 800.0000 \\ &= 800.00 \end{aligned}$$

This means that almost all scores on SAT subtests should range from 200 to 800. And that is how the SAT is scored. No one receives a score lower than a 200 or higher than an 800 on an SAT subtest.

The normal distribution is called "normal" because it is a naturally occurring distribution that occurs as a result of random processes. For example, consider taking 10 coins, tossing them, and counting how many turn up heads. It could be anywhere from 0 to 10. The most likely outcome is 5 heads and the least likely outcome is either 10 heads or no heads. If one tossed the 10 coins thousands of times and graphed the frequency distribution, it would approximate a normal distribution (see Figure 4.9).

Psychologists don't care that much about coins, but they do care about variables like intelligence, conscientiousness, openness to experience, neuroticism,

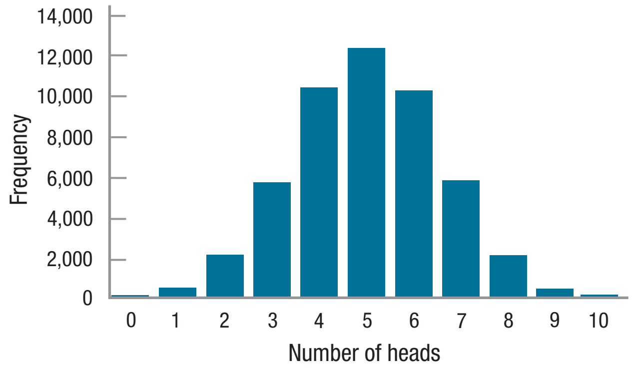

Figure 4.9 Approximation of Normal Distribution for Number of Heads for 10 Coins Tossed Together 50,000 Times If one tossed 10 coins 50,000 times and counted the number of heads, the distribution might look like this. Though this is a bar graph for discrete data, it has a normal shape like that seen for a continuous variable in Figure 4.5.

It is often assumed that psychological variables are normally distributed.

agreeableness, and extroversion. It is often assumed that these psychological variables (and most other psychological variables) are normally distributed. Here's an example that shows how helpful this assumption is:

1. If 34.13% of the cases fall from the mean to a standard deviation above the mean, and
2. if intelligence is normally distributed, and
3. if intelligence has a mean of 100 and a standard deviation of 15,
4. then 34.13% of people have IQs that fall in the range from 100 to 115, from the mean to 1 standard deviation above it.

If a person were picked at random, there's a 34.13% chance that he or she would have an IQ that falls in the range from 100 to 115. That's true as long as intelligence is normally distributed.

Finding percentages using the normal distribution wouldn't be useful if the percentages were only known for each whole standard deviation. Luckily, the normal curve has been sliced into small segments and the area in each segment has been calculated. Appendix Table 1, called a *z* score table, contains this information. A small piece of it is shown in Table 4.2.

There are a number of things to note about this *z* score table:

- It is organized by *z* scores to two decimal places. Each row is for a *z* score that differs by 0.01 from the row above or below.
- There are no negative *z* scores listed. Because the normal distribution is symmetrical, the same row can be used whether looking up the area under the curve for $z = 0.50$ or -0.50.
- The columns report the percentage of area under the curve that falls in different sections. This is the same as the percentage of cases that have scores in that region.

TABLE 4.2 From Appendix Table 1: Area Under the Normal Distribution

	A	B	C
z Score	Below $+z$ or Above $-z$	From Mean to z	Above $+z$ or Below $-z$
0.50	69.15%	19.15%	30.85%
0.51	69.50%	19.50%	30.50%
0.52	69.85%	19.85%	30.15%
0.53	70.19%	20.19%	29.81%
0.54	70.54%	20.54%	29.46%
0.55	70.88%	20.88%	29.12%
0.56	71.23%	21.23%	28.77%
0.57	71.57%	21.57%	28.43%
0.58	71.90%	21.90%	28.10%
0.59	72.24%	22.24%	27.76%

This is a section of Appendix Table 1. Note that there are three columns of information for each *z* score.

- The percentages reported are always greater than 0%. In any region of the normal distribution, there is always some percentage of cases that fall in it. Sometimes the percentage is very small, but it is always greater than zero.
- The percentages reported can't be greater than 100%. In any region, there can't be more than 100% of the cases.
- There are three columns, A, B, and C, each containing information about the percentage of cases that fall in a different section of the normal distribution. For each row in the table, there are three different views for each *z* score.

Column A

For a *positive z* score, column A tells the percentage of cases that have scores *at* or *below* that value. A positive *z* score falls *above* the midpoint, so the area *below* a positive *z* score also includes the 50% of cases that fall below the midpoint. This means the area below a positive *z* score will always be greater than 50% as it includes both the 50% of cases that fall below the mean, and an additional percentage that falls above the mean. Look at Figure 4.10, where *A* marks the area from the bottom of the curve to a *z* score of 0.50. This area is shaded ╱╱╱ and it should be apparent that it covers more than 50% of the area under the curve. According to column A in Appendix Table 1, the exact percentage of cases that fall in the region below a *z* score of 0.50 is 69.15%.

Because the normal distribution is symmetric, column A also tells the percentage of cases that fall *at* or *above* a *negative z* score. Figure 4.11 marks off area *A* for $z = -0.50$. Now compare Figure 4.10 to Figure 4.11, which are mirror images of each other: the same areas are marked off, just on different sides of the normal distribution. So, 69.15% of cases fall at or above a *z* score of −0.50.

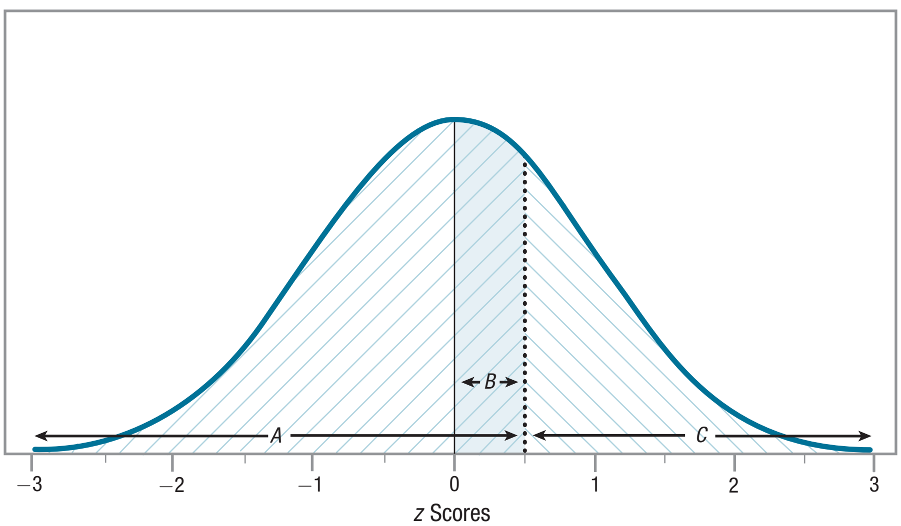

Figure 4.10 Finding Areas Under the Curve for a *z* Score of 0.50 Area *A*, marked ╱╱╱, is the portion of the normal curve that falls below $z = 0.50$. Area *B*, the shaded area, is the section from the mean to $z = 0.50$. Area *C*, marked ╲╲╲, is the section that falls above $z = 0.50$. These areas, *A*, *B*, and *C*, correspond to the columns A, B, and C in Appendix Table 1.

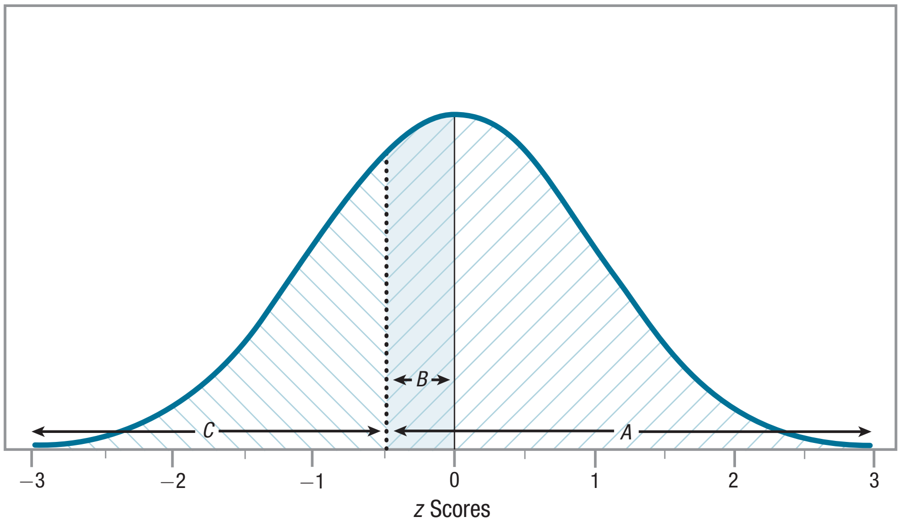

Figure 4.11 Finding Areas Under the Curve for a *z* Score of −0.50 Compare this figure to Figure 4.10. Note that the area *A* is the same in Figure 4.10 as here in Figure 4.11, but the side of the normal distribution it is on has changed as the sign of the *z* score changes. This shows the symmetry between positive and negative *z* scores. The same is true for area *B* in the two figures and for area *C*.

Column B

In Appendix Table 1, column B tells the percentage of cases that fall from the mean to a *z* score, whether positive or negative. This value can never be greater than 50% because it can't include more than half of the normal distribution. Figures 4.10 and 4.11 show the shaded area *B* for a *z* score of ±0.50. Looking in Table 4.2, one can see that this area captures 19.15% of the normal distribution.

Column C

Column C tells the percentage of cases that fall *at* or *above* a *positive z* score. Because of the symmetrical nature of the normal distribution, column C also tells the percentage of cases that fall *at* or *below* a *negative z* score. Look at Figures 4.10 and 4.11, which show area *C*, marked \\\, for a *z* score of ±0.50. It should be apparent that area *C* values will always be less than 50%. The *z* score table in Appendix Table 1 shows that the exact value is 30.85% for the percentage of cases that fall above a *z* score of 0.50 or below a *z* score of −0.50.

There are three questions that the *z* score table in Appendix Table 1 can be used directly to answer about a normal distribution:

1. What percentage of cases fall above a *z* score?
2. What percentage of cases fall from the mean to a *z* score?
3. What percentage of cases fall below a *z* score?

Figure 4.12 is a flowchart that walks through using the *z* score table in Appendix Table 1 to answer those questions.

Here's an example of the first question: "What is the percentage of cases in a normal distribution that have scores above a *z* score of −2.30?" Before turning to the

Corty_3e_Ch04_107-144_Final HR.indd 120 21/01/16 5:23 pm

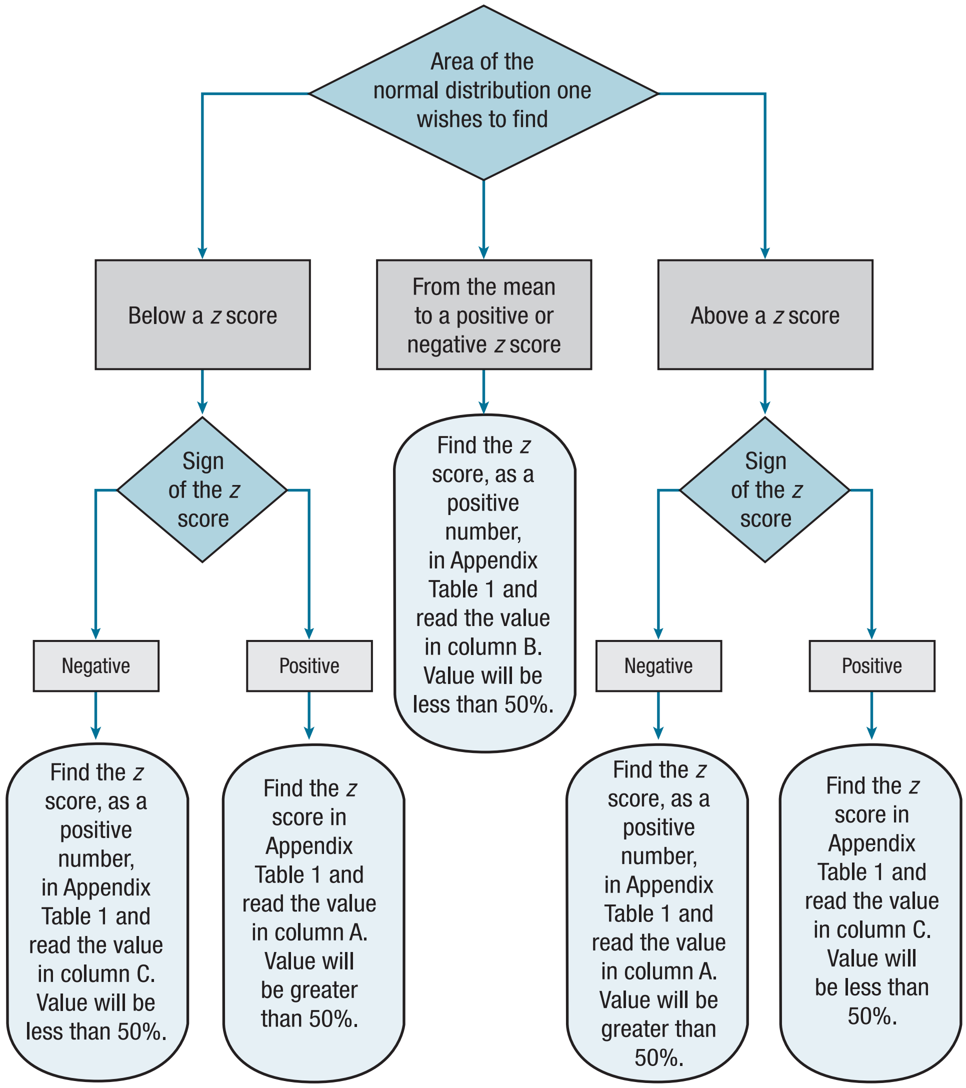

Figure 4.12 How to Choose: Flowchart for Finding Areas Under the Normal Curve It is always a good idea to make a sketch of the normal distribution and shade in the area one wishes to find in order to make a rough estimate. Then, use this flowchart to determine which column (A, B, or C) of Appendix Table 1 will lead to the exact answer.

flowchart in Figure 4.12, it is helpful to make a quick sketch of the area one is trying to find. Figure 4.13 is a normal distribution, where the *X*-axis is marked from $z = -3$ to $z = 3$. A vertical line is drawn at $z = -2.30$, and the area to the right of the line all the way to the far right side of the distribution is shaded in. That's the area above a *z* score of -2.30. The percentage of cases that fall in this shaded in area is greater than 50% and is probably fairly close to 100%.

Now, turn to the flowchart in Figure 4.12. It directs one to use column A of the *z* score table to find the exact answer. Turn to Appendix Table 1 and go down the column of *z* scores to 2.30. Then look in column A, which tells the area above a negative *z* score, to find the value of 98.93%. And that's the answer: 98.93% of cases in a normal distribution fall above a *z* score of -2.30.

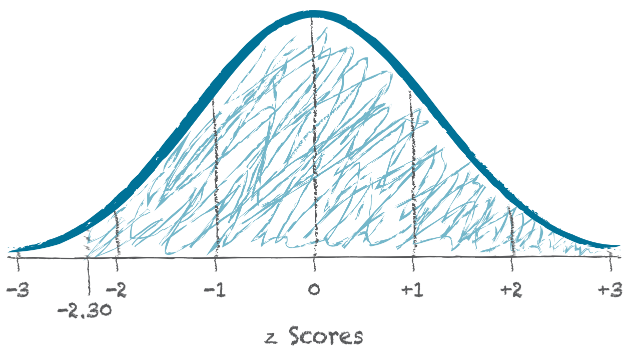

Figure 4.13 Sketch for Finding Area Above a *z* Score of −2.30 This figure marks the area in a normal distribution that falls above a *z* score of −2.30. The area includes the section to the right of the midpoint, so the total has to be greater than 50%. Inspection of the figure suggests that the area will be close to 100%.

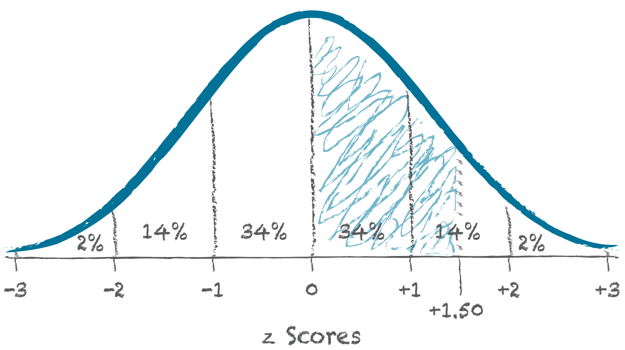

Figure 4.14 Sketch for Finding Area from the Mean to a z Score of 1.50 This figure marks the area in a normal distribution falling from the mean to a *z* score of 1.50. The numbers (34%, 14%, and 2%) that represent the approximate percentages in each standard deviation help estimate how much area the shaded area represents.

The second type of problem involves finding the area from the mean to a *z* score. For example, what's the area from the mean to a *z* score of 1.50?

Start with a sketch. In **Figure 4.14**, there's a normal curve. The *X*-axis is labeled with *z* scores from −3 to +3, and each standard deviation unit is marked with the approximate percentages (34, 14, or 2) that fall in it. There's a line at $z = 1.50$, and the area from $z = 0.00$ to $z = 1.50$ is shaded in. Inspection of this figure shows that the area being looked for will be greater than 34% but less than 48%.

To find the exact answer, use the flowchart in Figure 4.12: the answer can be found in column B of the *z* score table. The flowchart also indicates that the answer will be less than 50% (which is consistent with what was found in Figure 4.14). Finally, turn to the intersection of the row for $z = 1.50$ and column B in Appendix Table 1 for the answer: 43.32%. In a normal distribution, 43.32% of the cases fall from the mean to 1.50 standard deviations above it.

The third type of question asks one to find the area below a *z* score. For example, what is the area below a *z* score of −0.75?

As always, start with a sketch. **Figure 4.15** shows a normal distribution, filled in with the approximate areas in each standard deviation. There's a vertical line at −0.75, and the area to the left of this line (the area below this *z* score) is shaded in. That's the area to be found. This area will be larger than 16% because it includes the bottom 2%, the next 14%, and then some more.

To find the exact answer, turn to the flowchart in Figure 4.12. The flowchart says to use column C to find the *z* score (a positive number) in Appendix Table 1. Looking at the intersection of the row where $z = 0.75$ and column C, we see that the answer is 22.66%. In a normal distribution, 22.66% of the scores are lower than a *z* score of −0.75.

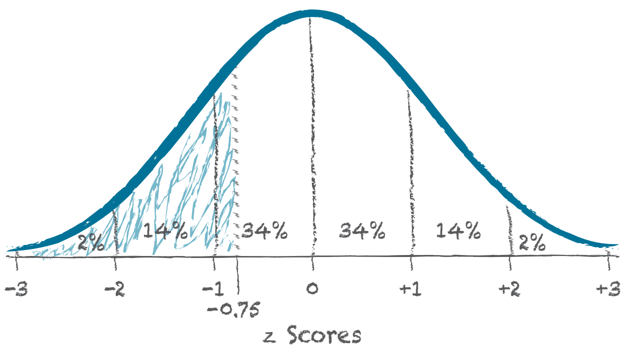

Figure 4.15 Sketch for Finding Area Less Than a z Score of −0.75 This figure marks the area in a normal distribution falling below a z score of −0.75. The numbers (34%, 14%, and 2%) that represent the approximate percentages in each standard deviation help estimate how much area the shaded area represents. The shaded area will be a little bit more than 16%.

After some practice finding the area under the normal distribution curve, we will move on to two new ways to use z scores and the normal curve. One, percentile ranks, is a new way to describe a person's performance. The other, probability, underpins how statistical tests work.

Worked Example 4.2

Here's some practice finding the area under the normal distribution using a real-life example. The expected length of a pregnancy from the day of ovulation to childbirth is 266 days, with a standard deviation of 16 days. Assuming that the length of pregnancy is normally distributed, what percentage of women give birth at least two weeks later than average? Two weeks is equivalent to 14 days, so the question is asking for the percentage of women who deliver their babies on day 280 or later.

The first step is to transform 280 into a z score:

$$\begin{aligned} z &= \frac{280 - 266}{16} \\ &= \frac{14.0000}{16} \\ &= 0.8750 \\ &= 0.88 \end{aligned}$$

Next, make a sketch to isolate the area we're looking for. In Figure 4.16, the area above a z score of 0.88 is shaded in. This area will be less than 50%. The area includes the distance from 2 standard deviations above the mean to 3 standard deviations above it (2%), the distance from 1 standard deviation above the mean to 2 standard deviations above it (14%), and then a little bit more. The area will be a bit more than 16%.

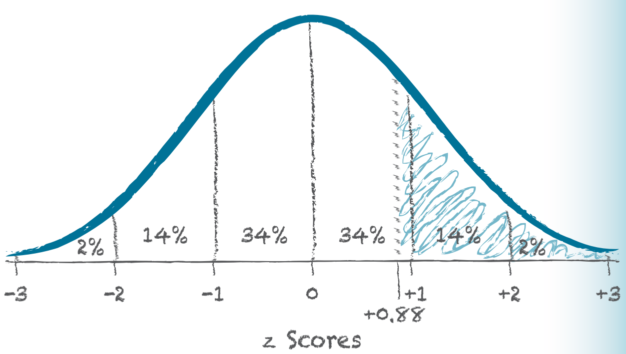

Figure 4.16 Sketch for Finding Probability of Having a Pregnancy That Lasts Two or More Weeks Longer Than Average The dotted line at $z = 0.88$ marks the area where deliveries two weeks later than average start. The numbers (34%, 14%, and 2%) that represent the approximate percentages in each standard deviation help estimate how much area the shaded area represents. The area will be a little bit more than 16%.

The area being looked for is the area above a z score. Consulting Figure 4.12, it is clear that the answer may be found in column C of Appendix Table 1. There, a z score of 0.88 has the corresponding value of 18.94%. And that's the conclusion—almost 19% of pregnant women deliver two weeks or more later than the average length of pregnancy.

Practice Problems 4.2

Review Your Knowledge

4.06 Describe the shape of a normal distribution.

4.07 What measure of central tendency is the midpoint of a normal distribution?

4.08 What do numbers 34, 14, and 2 represent in a normal curve?

Apply Your Knowledge

4.09 What percentage of area in a normal distribution falls at or above a z score of -1.33?

4.10 What percentage of area in a normal distribution falls at or above a z score of 2.67?

4.11 What percentage of area in a normal distribution falls from the mean to a z score of -0.85?

4.3 Percentile Ranks

So far, the z score table in Appendix Table 1 has been used as it was designed. First, the z score is found in the table and then the area in the normal distribution is found above, below, or to that z score. Now, the table will be used to calculate percentile ranks.

A **percentile rank** tells the percentage of cases whose scores are at or below a given level in a frequency distribution. Percentile ranks provide a new way of thinking about a person's score. Imagine a student, Leah, who obtained a score of 600 on an SAT subtest. (Remember, SAT subtests have a mean of 500 and a standard deviation of 100.) There have been two ways thus far of expressing Leah's score: (1) as a raw score ($X = 600$), or (2) as a standard score ($z = 1.00$).

Here's a third way, as a percentile rank. Using the z score of 1.00 and looking in column A of Appendix Table 1, 84.13% of the area under the normal curve falls at or below Leah's score. Leah's score as a percentile rank, abbreviated *PR*, is 84.13. Leah scored higher on this SAT subtest than about 84% of students. When her score is expressed in percentile rank form, it's clear that Leah did well on this test.

Now let's say that Joshua took the same test and his percentile rank was 2. Joshua didn't do very well—he only performed better than 2% of other students. Comparing Leah to Joshua shows a major advantage of percentile ranks—they put scores into an easily interpretable context.

Figure 4.17 shows a normal distribution that is marked off (1) with the approximate percentages in each standard deviation (34%, 14%, and 2%), (2) with z scores (from -3 to $+3$) on the X-axis, and (3) with approximate percentile ranks, also on the X-axis. There are several things to note in this figure:

- In this simplified version of the normal distribution, there are no cases that fall more than 3 standard deviations below the mean and none that fall more than 3 standard deviations above it.
- This means that the percentile rank associated with a z score of -3 is 0, and the percentile rank associated with a z score of $+3$ is 100.

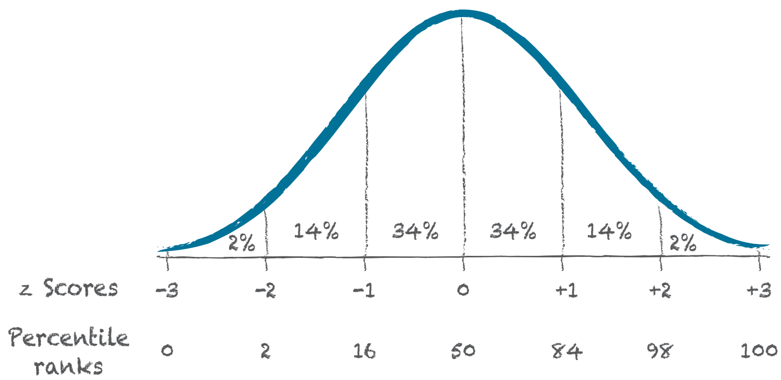

Figure 4.17 Relationship Among the Area Under the Normal Curve, *z* Scores, and Percentile Ranks Percentile ranks, which can be computed from *z* scores, express the same information but in a different format.

- The midpoint is the median. This means that a score at the midpoint has a percentile rank of 50.
- ≈2% of cases fall from the bottom of the normal curve to $z = -2$, so the percentile rank for $z = -2$ is 2.
- Moving up the normal distribution, it's clear that percentile ranks grow as one moves to the right along the *X*-axis.

A Common Question

Q What is the level of measurement for percentile ranks?

A Percentile ranks don't have equality of units, so they are ordinal. Look at Figure 4.16 and note that the distance (in *z* score units) from a *PR* of 2 to a *PR* of 16 is the same as the distance from a *PR* of 16 to a *PR* of 50. In one case, a single *z* score covers 14 *PR* units and in the other it covers 34 *PR* units.

A percentile rank can be calculated from a raw score or a *z* score. Alternatively, a raw score or a *z* score can be calculated from a percentile rank by using Appendix Table 1. The flowchart in Figure 4.18 shows how to do this.

Imagine another student, Kelli, who took the same SAT subtest and whose score had a percentile rank of 40. What is her *z* score? A sketch always helps. Figure 4.19 shows a normal distribution. The *X*-axis is marked with *z* scores, percentile ranks, and SAT subtest scores, which shows how these three scores are equivalent. For example, a *z* score of 1 is equivalent to a percentile rank score of 84 and an SAT subtest score of 600. There is also an X on the *X*-axis about where a percentile rank of 40 would be. The X shows that this percentile rank will have an associated *z* score that falls somewhere between −1 and 0 and an associated SAT score that falls somewhere between 400 and 500.

Now, it is time to calculate Kelli's *z* score more exactly. According to the flowchart in Figure 4.18, to find the *z* score associated with a percentile rank less than 50, go to column C in Appendix Table 1, which shows that the *z* score will be negative. Move down column C. The row closest to a value of 40.00 shows a *z* score of 0.25. This score needs to be treated as a negative number, so the answer is −0.25. Kelli's grade on the test, as a *z* score, is −0.25.

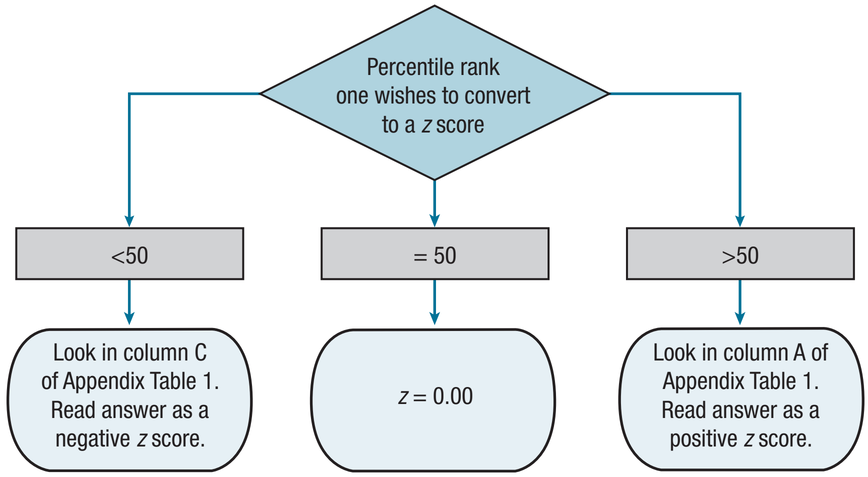

Figure 4.18 How to Choose: Flowchart for Calculating *z* Scores from Percentile Ranks Before calculating a *z* score for a percentile rank, it is a good idea to sketch out a normal distribution with both *z* scores and percentile ranks on the *X*-axis in order to estimate the answer. Once the answer is in *z* score format, it can be transformed to a raw score using Equation 4.3.

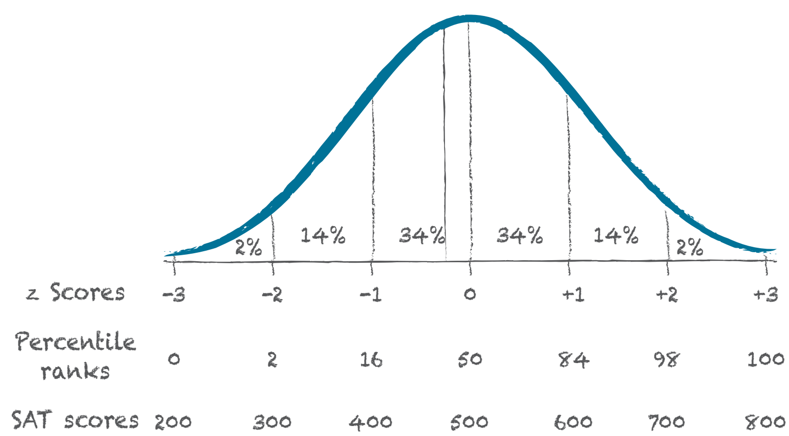

Figure 4.19 Sketch for Figuring Out the SAT Score for a Student with a Percentile Rank of 40 By labeling the *X*-axis with percentile rank scores as well as with SAT scores, it is possible to make a rough estimate as to what SAT score corresponds to a given percentile rank score.

Kelli's grade is now known in two formats: as a percentile rank (40) and as a *z* score (−0.25). What is her score as a raw score on the SAT subtest? To answer that question, use Equation 4.3:

$$\begin{aligned} X &= 500 + (-0.25 \times 100) \\ &= 500 + (-25.0000) \\ &= 500 - 25.0000 \\ &= 475.0000 \\ &= 475.00 \end{aligned}$$

Kelli obtained a 475.00 on this SAT subtest.

Worked Example 4.3

Imagine that Clayton takes an IQ test, with a mean of 100 and a standard deviation of 15, and does better than 95% of the population. If he did better than 95%, then his score, as a percentile rank, is 95. What was his IQ score if his score was a 95 as a percentile rank?

First, make a sketch. In Figure 4.20, the *X*-axis is marked with *z* scores, percentile ranks, and IQ scores and a dotted line is drawn about where *PR* 95 should be. The line is between *z* scores of 1 and 2, closer to 2 than to 1. In IQ units, it is between 115 and 130, closer to 130.

To calculate the raw score, use the flowchart in Figure 4.18. Clayton's score has a percentile rank above 50, so turn to column A of Appendix Table 1. In that column, look for the value closest to 95%. In this case, there are two values, 1.64 and 1.65, that are equally close to 95.00. Logic can be used to decide which of these to select:

- A *z* score of 1.64 is higher than 94.95% of scores. This doesn't quite capture the 95% value that is Clayton's score.
- A *z* score of 1.65 is higher than 95.05% of scores. This captures the 95% value that is Clayton's score.

To select the score that is as good or better than 95% of the cases in a normal distribution, go with 1.65. Clayton's score, as a *z* score, is 1.65.

To convert that *z* score to an IQ score, use Equation 4.3:

$$
\begin{aligned}
X_{IQ} &= 100 + (1.65 \times 15) \\
&= 100 + 24.7500 \\
&= 124.7500 \\
&= 124.75
\end{aligned}
$$

Clayton's score on the IQ test is 124.75, a score that was better than that of 95% of the population.

Remember the angioplasty graph from the beginning of the chapter? It showed that most cities had "normal" rates of angioplasty and one city was unusually high. Let's see how unusual that city is.

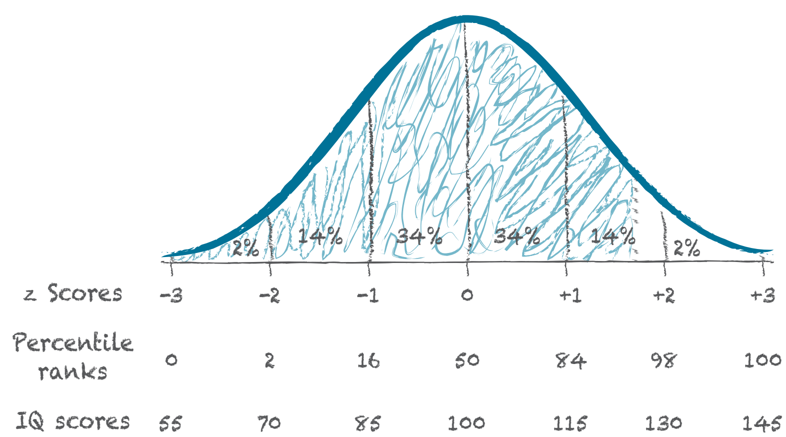

Figure 4.20 Sketch for Finding the IQ Score Associated with a Percentile Rank of 95 One can use the percentile ranks marked on the *x*-axis to estimate where to draw the line for the sketch for the IQ score associated with a percentile rank of 95.

To do so, we need to know the mean angioplasty rate for the sample of cities (11) and the standard deviation (4). This allows us to calculate a z score for the outlier city that has an angioplasty rate of 42:

$$z = \frac{X - M}{s} = \frac{42 - 11}{4} = 7.75$$

This city has an angioplasty rate that is 7.75 standard deviations above the national average. This is a very extreme score. The z score table in the back of the book only goes up to a z score of 5.00; and "only" 0.00003% of cases fall above this. A score of 7.75 is farther out in the tail than this, so it is an incredibly rare event. To a statistician's mind, such extreme events don't just happen by chance. There's something fishy going on in that city, something that has inflated the rate at which angioplasties are performed.

Practice Problems 4.3

Review Your Knowledge

4.12 What is a percentile rank?

4.13 What is the percentile rank for a score at the midpoint of a set of normally distributed scores?

Apply Your Knowledge

4.14 If a person's score, as a percentile rank, is 80, what is her z score?

4.15 What is the percentile rank associated with a z score of -0.45?

4.4 Probability

The focus of this chapter now shifts to probability. To understand why probability is important, consider a study in which a sample of people with depression is divided into two groups, one of which is given psychotherapy and the other isn't. After treatment, the two groups are assessed on a depression scale and the psychotherapy participants are found to be less depressed than the no-psychotherapy participants. Which conclusion, A or B, would a statistician draw?

A. Psychotherapy is effective in the treatment of depression.

B. Psychotherapy is probably effective in the treatment of depression.

A statistician would pick statement B, that psychotherapy is *probably* effective in the treatment of depression. Why? Because she would know that psychotherapy was effective for the *sample* tested, but she couldn't be sure it would be effective for all the depressed people not in the sample, for the larger *population* of depressed people.

Statisticians' conclusions involve *probability*, just as weather forecasters say it will "probably" rain tomorrow. The conditions may be right to predict that it will rain, but rain is not guaranteed. **Probability** concerns how likely it is that an event or outcome will occur. Probability is defined as the number of ways a specific outcome (or set of outcomes) can occur, divided by the total number of possible outcomes. The formula, with probability abbreviated as p and the specific outcome abbreviated as A, is given in Equation 4.4.

Equation 4.4 Formula for Calculating Probability

$$p(A) = \frac{\text{Number of ways outcome } A \text{ can occur}}{\text{Total number of possible outcomes}}$$

where p = probability
A = specific outcome

Coins are often used to teach probability. A coin has two possible outcomes, heads or tails, and they are *mutually exclusive*. This means the coin can only land heads or tails—not both—on a given toss. (*Fairness,* that each outcome is equally likely, is going to be assumed for all examples.) Applying Equation 4.4, if the coin has two possible outcomes, heads or tails, and there is only one way it can land heads, the probability of it landing heads is $p(\text{heads}) = \frac{1}{2}$. This could be reported a number of ways:

- Probability can be reported as a fraction. In this case, one would say, "There is a 1 out of 2 chance that a coin will turn up heads."
- Probability can be reported as a proportion by doing the math in the fraction, by dividing the numerator by the denominator: $\frac{1}{2} = .50$. One could report, "The probability of a coin turning up heads is .50." Or, one could say, "$p(\text{heads}) = .50$."
- Finally, by multiplying the proportion by 100, the probability can be turned into a percentage: $.50 \times 100 = 50$. One could say, "There is a 50% chance that a coin will land heads."

To get a little more complex than a single coin, let's toss two coins so that two outcomes occur at once. We'll assume that the two outcomes are *independent*. **Independence** means that the occurrence of one outcome has no impact on the other outcome. This means that how the first coin turns out, heads or tails, has no impact on how the second coin turns out.

As shown in **Figure 4.21**, there are four possible outcomes for the two coins: (1) the first is heads and the second is heads (HH), (2) the first is heads and the second is tails (HT), (3) the first is tails and the second is heads (TH), and (4) the first is tails and the second is tails (TT).

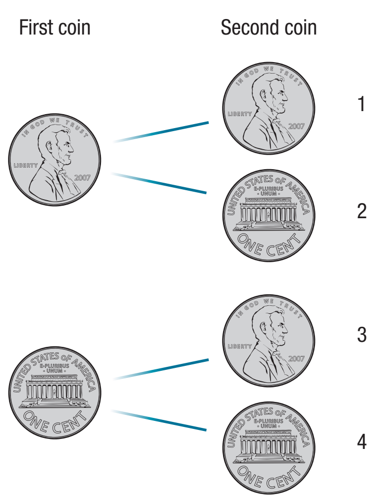

Figure 4.21 Four Possible Outcomes for Two Coins Being Tossed The first coin can be either heads or tails and so can the second coin, leading to four unique outcomes. Each of the four outcomes has one chance in four of occurring, so the probability of each outcome is $\frac{1}{4}$.

TABLE 4.3 Probabilities, Expressed as Proportions, for the Three Outcomes for Two Coins Being Tossed

Outcome	Probability
Two heads (HH)	.25
One heads and one tails (HT or TH)	.50
Two tails (TT)	.25
	$\Sigma = 1.00$

There are three possible outcomes if two coins are tossed. The sum of the probabilities for all three outcomes is 1.00.

What is the probability of obtaining two heads? There's one way such a result can occur out of four possible outcomes, so $p(\text{two heads}) = \frac{1}{4} = .25$. What is the probability of getting a heads and a tails? There are two ways this can occur (HT or TH) out of four possible outcomes, so $p(\text{one heads and one tails}) = \frac{2}{4} = .50$. Finally, there's one other outcome possible for the two coins: $p(\text{two tails}) = \frac{1}{4} = .25$. Table 4.3 summarizes the probabilities of the three outcomes for the two coins.

The probability for any given outcome can't be less than zero or greater than 1.

If a probability for an outcome is 1.00, or expressed as a percentage 100%, that means this outcome is a sure thing. In fact, any given outcome can't have a probability higher than 1.00, and the sum of the probabilities of all the possible outcomes is 1.00. If two coins are tossed, it is guaranteed that one of these outcomes—two heads, *or* a heads and a tails, *or* two tails—will turn out. In addition, the probability of an outcome occurring can't be less than .00, or 0%. A probability of zero means that an outcome can't happen. Putting together the two facts in this paragraph, the probability for any given outcome can't be less than zero or greater than 1. Stated more formally,

$$.00 \leq p(A) \leq 1.00$$

A Common Question

Q When probabilities are reported as a proportion, they are written without a leading zero, for example, as .25 not as 0.25. Why?

A APA format says to put a zero before the decimal point for numbers that are less than 1 only when the numbers can be greater than 1. Probabilities can't exceed 1, so they don't get a leading zero.

The most common way that statisticians calculate and use probability involves the normal distribution. They often ask, if a variable is normally distributed and a case is selected at random from a population, how likely is it—how probable is it—that this case would have a score that falls in a certain range?

The material covered in Section 4.2 of this chapter can be construed in just this way. For example, if 50% of cases fall at or above the mean, then the probability of a case selected at random having a score at or above the mean is .50. If approximately 84% of the cases fall in the region at or below a z score of 1, then the probability of picking a case at random and it having a z score less than or equal to 1 is .84. (To move from a percentage to a probability, move the decimal two places to the left; to move from a probability to a percentage, move the decimal place two spots to the right.)

Statisticians use probabilities to determine if a result is common or rare. By *common*, they mean the result occurs frequently, that it falls somewhere near the middle of a normal distribution. Typically, statisticians consider something that has a 95% chance of occurring as common. A *rare* result is one that is unlikely to occur, one that falls in the ends, the tails, of the distribution. As will be seen in Chapter 6, how rare something must be to be considered rare is a decision that should be made on an experiment-by-experiment basis. But, usually, statisticians consider something rare if it has a 5% or less chance of occurring. Phrased in terms of probability, if $p < .05$, it is usually considered a rare outcome to a statistician.

Let's find the z scores that mark off the middle 95% of scores in a normal distribution. This will lead us to one of the most important numbers in statistics, 1.96.

Follow the procedure in Table 4.4 to find the z scores that mark off the middle 95% of cases. The first step is to take the percentage, here 95%, and split it in two:

$$\frac{95}{2} = 47.5000 = 47.50$$

TABLE 4.4 How to Choose: Finding z Scores Associated with Middle and Extreme Percentages of the Normal Distribution

	Finding z Scores Associated with a Middle Percentage of the Normal Distribution
Step 1	Take the middle percentage and cut it in half.
Step 2	Use column B of Appendix Table 1 to find the percentage closest to the ½ value calculated in Step 1.
Step 3	Report the z score for the percentage in ± format.
	Finding z Scores Associated with an Extreme Percentage of the Normal Distribution
Step 1	Take the extreme percentage and cut it in half.
Step 2	Use column C of Appendix Table 1 to find the percentage closest to the ½ value calculated in Step 1.
Step 3	Report the z score for the percentage in ± format.

These guidelines help one calculate the percentage of cases, in a normal distribution, that fall in the middle or in the two tails.

Why split it in two? Because that tells us how much of the area is above the mean and how much is below it. We now know that the 95% consists of 47.5% (half) above the mean and 47.5% (the other half) below the mean.

The second step is to look in column B of Appendix Table 1 to find the value closest to this percentage. Column B contains the exact value of 47.50, and the z score associated with it is 1.96. The final step in Table 4.4 is to report the results as a range from a negative z score to a positive z score: in a normal distribution, the middle 95% of cases fall from a z score of -1.96 to a z score of 1.96 (see Figure 4.22). Or, thinking in terms of probability, if one took a case at random from the population, the probability is .95 that it came from the interval ranging from $z = -1.96$ to $z = 1.96$.

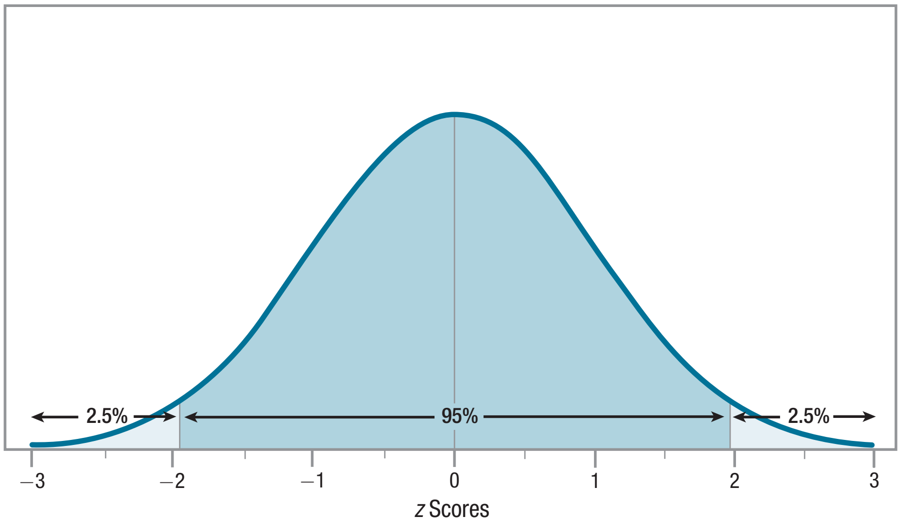

Figure 4.22 Middle and Extreme Sections of the Normal Curve Associated with z scores of ± 1.96 The same z scores, ± 1.96, that mark off the middle 95% of scores also mark off the extreme 5% of scores.

Look at Figure 4.22 that shows the middle 95%. If 95% of the cases fall from a z score of -1.96 to a z score of 1.96, isn't that the same as saying 5% of the cases don't fall in that region? The complement to a **common zone**, the middle section, is the **rare zone**, the section in the ends, the two tails, of the distribution. This section in the extremes of the distribution is evenly divided between the two tails. The extreme 5%, for example, consists of the 2.5% at the very bottom of the normal distribution (the far left side in Figure 4.22) *and* the 2.5% at the very top of it (the far right side in Figure 4.22). Of course, this can also be expressed as a probability: $p = .05$ that a case selected at random from the population falls below –1.96 or above 1.96 in a normal distribution. (Table 4.4, which shows how to calculate a **middle percentage**, also contains guidelines on how to calculate an **extreme percentage**.)

The z scores, ± 1.96, that mark off the middle 95% of scores in a normal distribution are also the z scores that mark off the extreme 5% of scores in a normal distribution.

The z scores, ± 1.96, that mark off the middle 95% of scores in a normal distribution are also the z scores that mark off the extreme 5% of scores in a normal distribution. Keep your eyes open, these values will appear a lot more in later chapters. You'll see 1.96 used in formulas; you'll see 95% used in confidence intervals; and you'll see two probabilities, $p < .05$ and $p > .05$, to indicate, respectively, whether a result is rare or common.

Worked Example 4.4

Mount Pleasant High School instituted a new math curriculum. The principal was not worried about how the average students would fare, but was concerned about the poorer and the better students. He wanted to assess the impact on the extreme 50% of students as measured by IQ. He needed to find the IQ scores that mark off the extreme 50% of cases, the 25% at the top and the 25% at the bottom.

Figure 4.23, where the X-axis is marked off with z scores and IQ scores, is the sketch he made. The approximate areas falling in each standard deviation are noted, and dotted lines are drawn about where the cut-off points will be. Based on the sketch, one cut-off point is expected to occur between z scores of

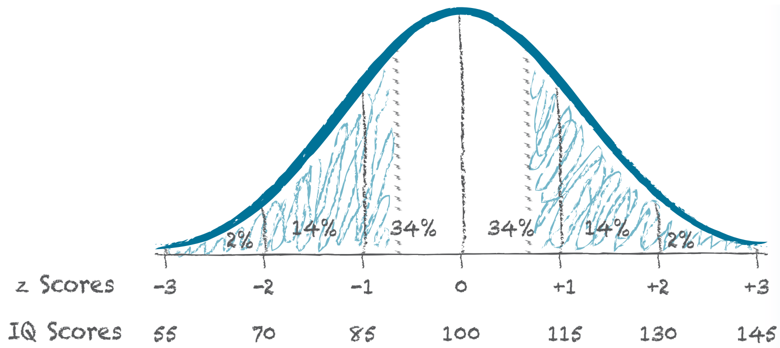

Figure 4.23 Sketch for Finding the Extreme 50% of IQ Scores The extreme 50% of scores consist of the bottom 25% of scores, the shaded area on the left of the curve, and the top 25%, the shaded area on its right.

-1 and 0. The other cut-off point will occur, symmetrically, on the positive side of the curve.

How did he know where to sketch the cut-off points? The objective is to have half of the 50% of extreme scores, or 25%, in each tail. He knew that below, or to the left of, a z score of -1, about 16% of the scores fall. That isn't 25%. So, he needed to move partway into the next standard deviation, enough to add 9% to the existing 16%. This puts the z score somewhere between $z = -1$ and $z = 0$.

Now, with a rough idea of what the answer was, he followed the guidelines in Table 4.4 to find the exact answer. First, he cut the percentage in half. Then, he used column C of Appendix Table 1 to find the z score associated with the half percentage value. Looking in column C of Appendix Table 1, he found that the value closest to 25% was 25.14 and the z score associated with it was 0.67. In z score format, he had his answer: below a z score of -0.67 *and* above a z score of 0.67, the extreme 50% of cases fall.

But, to answer the original question, he needed to turn the z scores into IQ scores using Equation 4.3:

$$\begin{aligned} X &= 100 + (-0.67 \times 15) \\ &= 100 + (-10.0500) \\ &= 100 - 10.0500 \\ &= 89.9500 \\ &= 89.95 \end{aligned}$$

$$\begin{aligned} X &= 100 + (0.67 \times 15) \\ &= 100 + 10.0500 \\ &= 110.0500 \\ &= 110.05 \end{aligned}$$

He now had his answer: The IQ scores that mark off the extreme 50% of scores are 89.95 and 110.05. He would need to follow the progress of students with IQs lower than 90 or higher than 110 to see how the new math curriculum affected the extreme 50% of students.

Practice Problems 4.4

Review Your Knowledge

4.16 What is the definition of probability?

4.17 What's the smallest possible probability for an outcome? The largest?

4.18 If the z scores that mark off the middle 60% of scores are $\pm.84$, what are the z scores that mark off the extreme 40% of scores?

Apply Your Knowledge

For Practice Problems 4.19–4.20, assume that in any given year, there are 52 weeks in a year, 365 days in a year, and equal numbers of every day of the week.

4.19 If a researcher picked one date at random out of a year, what is the probability of it being a Monday?

4.20 If a researcher picked one date at random out of a year, what is the probability of it not being a Monday?

4.21 What are the z scores that mark the middle 34% of scores?

4.22 If someone is picked at random, what is the probability that her IQ will be 123 or higher? (Use a mean of 100 and a standard deviation of 15.)

Application Demonstration

Let's use some historical data about IQ to see material from this chapter in action. In the past, intelligence was used as the sole criterion of whether a person was classified as having an intellectual disability or being gifted. These days, multiple criteria are used. Let's see what the implications of the former criteria are, based on an IQ test with a mean of 100 and a standard deviation of 15.

Let's start with intellectual disability. At one point, people were considered intellectually disabled if their IQs were less than 70. Now, adaptive behavior—the ability to function in everyday life—is also taken into account. But, using the old definition, if a person were picked at random, what is the probability that he or she would be classified as intellectually disabled?

An IQ score below 70 means an IQ score less than or equal to 69, so our question becomes, "What is the probability of having an IQ score $\leq$ 69?" It always helps to visualize what one is doing, so sketch out a normal distribution. Figure 4.24 shows a normal distribution with the X-axis marked in both IQ scores and z scores. For the IQ scores, the midpoint is 100 and then the standard deviation is added or subtracted in both directions 3 times. Thus, the IQ scores range from 55 to 145. There is a dotted vertical line marking an IQ score of 69 and the area below the line, to the left of it, is shaded in. This is the area to be found. It should be apparent that this is not a large percentage of cases. In fact, it should be close to 2%.

Now, it is time to calculate more exactly. Using Equation 4.2, convert the IQ score to a z score:

$$\begin{aligned} z_{IQ} &= \frac{X - M}{s} \\ &= \frac{69 - 100}{15} \\ &= \frac{-31.0000}{15} \\ &= -2.0667 \\ &= -2.07 \end{aligned}$$

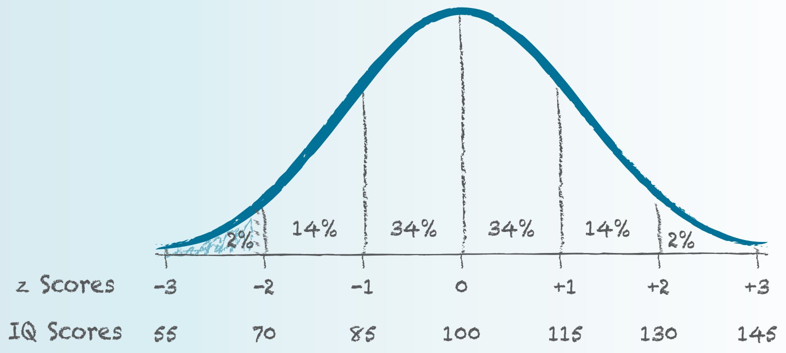

Figure 4.24 Sketch for Finding Percent of Cases with IQ Scores at or Below 69
Shading in the area at or below an IQ score of 69 allows one to estimate that less than 2% of the population falls in that area.

An IQ score of 69 is equivalent to a z score of -2.07. (Remember, the fact that the z score is negative means that it falls below the mean.) Now, take that z score and find the percentage of cases at or below it. The flowchart in Figure 4.12 directs one to column C of Appendix Table 1. Turn to that table and find the row with a z score of 2.07. (Remember, the normal distribution is symmetrical, so one can consider 2.07 the same as -2.07.) The answer as a percentage, 1.92%, is found in column C. We can turn 1.92% into a probability of .0192 by moving the decimal place two places to the left. And that's the answer—if a person were picked at random out of the world, the probability is .0192 that he or she would be classified as intellectually disabled if that classification were based entirely on IQ.

Here's one more way to use this probability. If the population of the United States is 320 million, how many have IQs of 69 or lower? Answer this by taking the probability, .0192, and multiplying it by the population:

$$.0192 \times 320{,}000{,}000 = 6{,}144{,}000$$

That's a lot of people! Using the old standard, more than 6 million people in the United States should have IQs of 69 or lower and would be classified as intellectually disabled.

Let's look at the other side of the intelligence continuum—genius. How common are geniuses? One cut-off for genius, on a standard IQ test (mean = 100 and standard deviation = 15), is a score of 145 or higher. This is sketched in Figure 4.25. It should be apparent that this is the upper limit of the simplified version of the normal distribution. There are very few people with IQ scores of 145 or higher.

To find out what percentage of people could be classified as geniuses, use Equation 4.2 to turn the IQ score into a z score:

$$\begin{aligned} Z_{IQ} &= \frac{145 - 100}{15} \\ &= \frac{45.0000}{15} \\ &= 3.0000 \\ &= 3.00 \end{aligned}$$

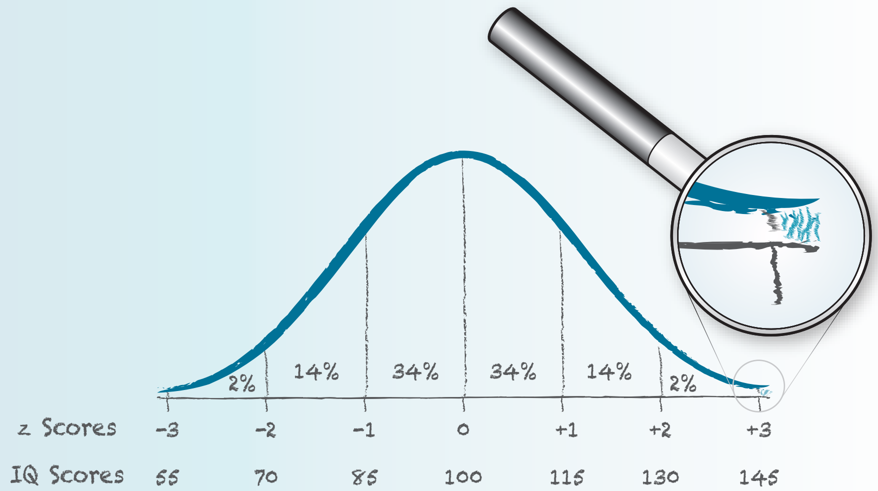

Figure 4.25 Sketch for Finding Percentage of Population with IQs of 145 or Higher Trying to shade in the area to the right of an IQ score of 145 makes one aware of how few scores fall in that area.

The flowchart in Figure 4.11 indicates that the exact percentage may be found in column C of Appendix Table 1. Reading across the row in the *z* score table for a *z* score of 3.00, one finds that the value in column C is 0.135%. And that's the answer: assuming intelligence is normally distributed and considering anyone with an IQ of 145 or higher a genius, then 0.135% of the population will be geniuses.

What does that mean for the 320 million Americans? Turn 0.135% into a proportion (.00135) and multiply that by the population:

$$.00135 \times 320{,}000{,}000 = 432{,}000$$

Fewer than half a million Americans, only 432,000 to be exact, should have IQs high enough to be classified as geniuses.

One last thing, let's suppose someone takes an IQ test and gets a 145. What's his or her score as a percentile rank? The flowchart in Figure 4.18 says to consult column A of Appendix Table 1. Reading column A for a *z* score of 3.00, we see that the person's score is 99.865, as a percentile rank. Not a bad percentile rank to have.

DIY

Dig out the standardized test scores, either the SAT or the ACT, that you used on your college application. How did you do? Can you turn your score into a *z* score and a percentile rank? For the combined SAT score, which is the sum of the math, writing, and critical reading tests, use a mean of 1,500 and a standard deviation of 300 (If you have taken the new SAT, which has only two subtests, use a mean of 1000 and a standard deviation of 200). For the ACT composite score, use a mean of 20.5 and a standard deviation of 5.5.

SUMMARY

Transform raw scores into standard scores (z scores) and vice versa.

- A standard score, called a z score, transforms a raw score so that it is expressed in terms of how many standard deviations it falls away from the mean. A positive z score means the score falls above the mean, a negative z score means it falls below the mean, and a z score of zero means the score falls right at the mean. z scores standardize scores, allowing different variables to be expressed in a common unit of measurement.

Describe the normal curve and calculate the likelihood of an outcome falling in specified areas of it.

- The normal distribution is important because it is believed that many psychological variables are normally distributed. The normal curve is a specific bell-shaped curve, defined by the percentage of cases that fall in specified areas. About 34% of cases fall from the mean to 1 standard deviation above the mean, ≈14% fall from 1 to 2 standard deviations above the mean, and ≈2% fall from 2 standard deviations above to 3 above the mean. Because the normal distribution is symmetric, the same percentages fall below the mean. It is rare that a case has a score more than 3 standard deviations from the mean.
- Appendix Table 1 lists the percentage of cases that fall in specified segments of the normal distribution. A flowchart, Figure 4.12, can be used as a guide to find the area under a normal curve that is above a z score, below a z score, or from the mean to a z score.

Transform raw scores and standard scores into percentile ranks and vice versa.

- Percentile ranks tell the percentage of cases in a distribution that have scores at or below a given level. They provide easy-to-interpret information about how well a person performed on a test.

Calculate the probability of an outcome falling in a specified area under the normal curve.

- Conclusions that statisticians draw are probabilistic.
- Probability is defined as the number of ways a specific outcome or event can occur, divided by the total number of possible outcomes.
- The normal curve can be divided into a middle section and extreme sections. A middle percentage of scores is symmetric around the midpoint, while an extreme percentage is evenly divided between the two tails of the distribution. The same z score, for example, defines the middle 60% and the extreme 40%. The middle 60% consists of the first 30% above the midpoint *and* the first 30% below the midpoint, while the extreme 40% consists of the 20% of scores above the middle 60% *and* the 20% below it.
- z scores can be used to find the probability that a score, selected at random, falls in a certain section of the normal curve. For example, $p = .50$ that a score selected at random has a score at or below the mean.
- The rare zone of a distribution is the part in the tails where scores rarely fall. Commonly, statisticians consider something rare if it happens less than 5% of the time. If rare is written as $p < .05$, then a common outcome is written as $p > .05$, meaning it happens more than 5% of the time. The common zone of a distribution is the central part where scores commonly fall.

KEY TERMS

common zone – the section of a distribution where scores usually fall; commonly set to be the middle 95%.

extreme percentage – percentage of the normal distribution that is found in the two tails and is evenly divided between them.

independence – in probability, when the occurrence of one outcome does not have any impact on the occurrence of a second outcome.

middle percentage – percentage of the normal distribution found around the midpoint, evenly divided into two parts, one just above the mean and one just below it.

normal distribution – also called the normal curve; a specific bell-shaped curve defined by the percentage of cases that fall in specific areas under the curve.

percentile rank – percentage of cases with scores at or below a given level in a frequency distribution.

probability – how likely an outcome is; the number of ways a specific outcome can occur, divided by the total number of possible outcomes.

rare zone – section of a distribution where scores do not usually fall; in most instances, set to be the extreme 5%.

standard score – raw score expressed in terms of how many standard deviations it falls away from the mean; also known as a *z* score.

***z* score** – raw score expressed in terms of how many standard deviations it falls away from the mean; also known as a standard score.

CHAPTER EXERCISES

Answers to the odd-numbered exercises appear at the back of the book.

Review Your Knowledge

4.01 A standard score is a ____ score expressed in terms of how many standard deviations it is away from the mean.

4.02 ____ is another term for standard score.

4.03 To calculate a *z* score, one needs to know the raw score, the ____, and the ____.

4.04 A positive *z* score means the score is ____ the mean; a negative *z* score means the score is ____ the mean.

4.05 If a raw score is right at the mean, it has a *z* score of ____.

4.06 The sum of *z* scores for a data set is ____.

4.07 Given a *z* score, it is possible to turn it into a raw score as long as one knows the sample ____ and ____.

4.08 The normal curve is symmetrical, has the highest point in the ____, and has frequencies that ____ as one moves away from the midpoint.

4.09 Because the normal curve is symmetric, the midpoint is also the ____ and the ____.

4.10 In a normal curve, 34.13% of the cases fall from the mean to ____ above the mean.

4.11 As one moves away from the mean in a normal distribution, the percentage of cases that fall in each standard deviation ____.

4.12 Within 2 standard deviations of the mean in a normal distribution, about ____% of the cases fall.

4.13 In a normal distribution, it is ____ to have a *z* score that is greater than 3.

4.14 Three numbers that are good to remember for estimating areas in the first 3 standard deviations of a normal distribution are ____, ____, and ____.

4.15 The normal distribution is a naturally occurring distribution that is the result of ____ processes.

4.16 Most psychological variables are considered to be ____.

4.17 Appendix Table 1 slices the normal distribution into segments that are ____ *z* score units wide.

4.18 The area that falls above a positive *z* score is the same as the area that falls ____ a negative *z* score.

4.19 The area that falls from the mean to a positive *z* score is ____ as the area from the mean to the same value as a negative *z* score.

4.20 The area that falls from the mean to a negative *z* score can't exceed ____%.

4.21 The area that falls below a negative *z* score will always be ____ 50%.

4.22 One can use the *z* score table in reverse, to find a ____ when given a percentile rank.

4.23 A percentile rank tells the percentage of cases whose scores are ____ a given level.

4.24 In a normal distribution, the percentile rank at the ____ is 50.

4.25 Statistical conclusions involve ____.

4.26 ____ concerns how likely an outcome is to occur.

4.27 Probability is defined as the number of ways a specific outcome can occur divided by ____.

4.28 If two outcomes are ____, then only one outcome can occur at a time.

4.29 Probability can be reported as a ____, a ____, or a ____.

4.30 If two outcomes are ____, then what the first outcome is has no impact on what the second outcome is.

4.31 The highest possible probability for an outcome is ____ and the lowest possible probability is ____.

4.32 If 45% of the scores fall in a certain section of the normal distribution, then $p =$ _____that a case selected at random from the population would have a score that falls in that section.

4.33 Statisticians usually think of something that has a probability of _____as rare and something with a probability of .95 as________.

4.34 The middle 16% of scores in a normal distribution consist of ____% just above the mean and ____% just below the mean.

4.35 The extreme 6% of scores in a normal distribution consist of ____% in each tail of the distribution.

4.36 ____ is a very important number in statistics.

Apply Your Knowledge

Calculating standard scores

4.37 If $M = 7$, $s = 2$, and $X = 9.5$, what is z?

4.38 If $X = 17.34$, $s = 5.45$, and $M = 24.88$, what is z?

Calculating raw scores

4.39 If $z = 1.45$, $s = 3.33$, and $M = 12.75$, what is X?

4.40 If $s = 25$, $M = 150$, and $z = -0.75$, what is X?

Calculating z scores

4.41 If John's score on an IQ test is 73, what is his score as a *z* score? (For this and subsequent IQ questions, use a mean of 100 and a standard deviation of 15.)

4.42 If Ebony's score on an IQ test is 113, what is her score as a *z* score?

Calculating raw scores

4.43 Chantelle's score on an IQ test, expressed as a *z* score, is 0. What was her score?

4.44 Sven's score on an IQ test, expressed as a *z* score, is −.38. What was his score?

Comparing scores

4.45 Hillary's score on the math subtest of the SAT was 620. On a spelling test ($M = 60$, $s = 15$), she got a 72. Is she better at math or spelling? (On subtests of the SAT, the mean is 500 and the standard deviation is 100.)

4.46 Jong-Il was feeling depressed and anxious. His therapist gave him a depression test ($M = 10$, $s = 3$, higher scores mean more depression) and an anxiety inventory ($M = 80$, $s = 25$, higher scores mean more anxiety). His score was a 15 on the depression test and a 110 on the anxiety inventory. Is Jong-Il more depressed than anxious, or more anxious than depressed?

Finding the area above, below, or to a z score

4.47 What percentage of scores in a normal distribution fall at or above a z score of 1.34?

4.48 What percentage of cases in a normal distribution fall at or below a z score of 2.34?

4.49 What percentage of cases in a normal distribution fall at or above a z score of -0.85?

4.50 What percentage of scores in a normal distribution fall at or below a z score of -2.57?

4.51 What percentage of scores in a normal distribution fall from the mean to a z score of -1.96?

4.52 What percentage of cases in a normal distribution fall from the mean to a z score of 2.58?

4.53 The mean diastolic blood pressure in the United States is 77, with a standard deviation of 11. If diastolic blood pressure is normally distributed and a diastolic blood pressure of 90 or higher is considered high blood pressure, what percentage of Americans have high diastolic blood pressure?

4.54 The average American male is 5′ 9″. The standard deviation for height is 3″. If a basketball coach only wants to have American men on his team who are at least 6′ 6″, what percentage of the U.S. male population is eligible for recruitment?

Solving for percentile rank

4.55 If $z = -2.12$, what is the score as a percentile rank?

4.56 What is the percentile rank for a z score of 1.57?

4.57 If Gabrielle's score on an IQ test is 113, what is her score as a percentile rank?

4.58 Werner's score on an SAT subtest was 480. What is his score as a percentile rank?

Moving from percentile rank to raw score

4.59 Justin's score as a percentile rank on an SAT subtest was 88.5. What is his score on the SAT subtest?

4.60 Assume systolic blood pressure is normally distributed, with a mean of 124 and a standard deviation of 16. Sarita's systolic blood pressure, as a percentile rank, is 20. What is her systolic blood pressure?

Calculating probabilities

4.61 If there are 28 students in a classroom and 18 of them are boys, what is the probability—if a student is picked at random—of selecting a boy?

4.62 Paolo, a kindergarten student was given a pack of colored construction paper with 10 pages each of red, orange, yellow, green, blue, indigo, violet, black, and white. His favorite color is green. What is the probability, if he selects a page at random, of picking a green one?

Finding standard scores for middle and extreme sections of the normal curve

4.63 What are the z scores associated with the middle 10% of scores?

4.64 What are the z scores associated with the middle 54% of scores?

4.65 What are the IQ scores associated with the middle 84% of scores?

4.66 What are the IQ scores associated with the middle 61% of scores?

4.67 What are the z scores associated with the extreme 4% of scores?

4.68 What are the z scores associated with the extreme 18% of scores?

4.69 What are the SAT subtest scores associated with the extreme 15% of scores?

4.70 What are the SAT subtest scores associated with the extreme 8% of scores?

Finding probabilities

4.71 What is the probability, for a score picked at random from a normal distribution, that it

Corty_3e_Ch04_107-144_Final HR.indd 140 21/01/16 5:23 pm

falls at or above 3 standard deviations above the mean.

4.72 What is the probability, for a score picked at random from a normal distribution, that it falls at or below 2.7 standard deviations below the mean.

4.73 What is the probability, for a score picked at random from a normal distribution, that it falls within .38 standard deviations of the mean.

4.74 What is the probability, for a score picked at random from a normal distribution, that it falls within 2.1 standard deviations of the mean?

4.75 What is the probability, for a score picked at random from a normal distribution, that it does not fall within half a standard deviation of the mean?

4.76 What is the probability, for a score picked at random from a normal distribution, that it falls at least 4 standard deviations away from the mean?

Defining common and rare

4.77 A researcher decides to call something rare if it should happen no more than 10% of the time. What *z* scores should she use as cut-off scores?

4.78 Another researcher decides to call something rare if it should happen no more than 1% of the time. What *z* scores should he use as cut-off scores?

4.79 Dr. Noyes wants his rare zone to cover an area with a probability of .05. He would like this whole area to be on the positive side of the normal curve. What *z* score will be the cut-off for this area?

4.80 Dr. Hicks will call something rare if it happens less than or equal to 3% of the time and is greater than the mean. Expressed as a *z* score, what is the cut-off score?

Expand Your Knowledge

4.81 What is the standard deviation of a set of *z* scores?

4.82 Answer this without consulting Appendix Table 1. In a normal distribution, does a higher percentage of scores fall between *z* scores of 1.1 and 1.2, or between *z* scores of 2.1 and 2.2?

a. 1.1 and 1.2
b. 2.1 and 2.2
c. The percentages are the same.
d. This can't be answered without consulting Appendix Table 1.

4.83 A researcher obtained a set of scores and has made a frequency polygon for the distribution. The distribution is symmetric, has the highest percentage of scores occurring at the midpoint, and the frequency of scores decreases as it moves away from the midpoint. The distribution is which of the following?

a. Normally distributed
b. Not normally distributed
c. May be normally distributed
d. Not enough information to tell

4.84 Marilyn vos Savant claims that her IQ was once measured at 228 and that she is the smartest person in the world. (vos Savant is a real person and writes a weekly column for *Parade* magazine.) Let's assume that she took a standard IQ test. How credible is her claim that her IQ is 228?

4.85 A researcher has a sample of five cases. She has used the sample mean and sample standard deviation to calculate *z* scores. Four of the *z* scores are 0.50, 1.00, 1.50, and 2.00. What is the missing *z* score?

4.86 A researcher wants to study people who are very jealous. She believes that jealousy is normally distributed. She uses a jealousy scale that has a mean of 75 and a standard deviation of 25. She administers the jealousy scale as a screening device to a large group of people. For her study, she will only select people with scores of 125 or higher. How many people will she need to screen in order to end up with 50 people in her sample?

4.87 Assume that there is an animal intelligence test. A psychologist administers it to a representative sample of cats and a representative sample of dogs. Both species turn out to be, on average, equally smart. (That is, $M_{\text{Dogs}} = M_{\text{Cats}}$.) But, more variability occurred in the IQs of dogs than cats. (That is, $s_{\text{Dogs}} > s_{\text{Cats}}$.) If we consider animals with IQs above 110 to be "geniuses," will there be more genius cats or more genius dogs?

4.88 If a person is selected at random, what is the probability that he is in the top 7% of the world in terms of height?

4.89 A scale to measure perseverance has been developed. Perseverance is normally distributed and scores on the measure range from 15 to 45. It is very, very rare that anyone scores lower than 15 or higher than 45. What are the scale's mean and standard deviation?

4.90 What percentage of cases in the normal distribution have *z* scores in the range from 1.25 through 1.75?

SPSS

SPSS doesn't do a lot with *z* scores or percentile ranks. It does, however, offer a way of converting raw scores to *z* scores and saving them in the data file so that they can be used in other analyses. To do so, go to the menu under "Analyze," then "Descriptive Statistics," and finally "Descriptives" as seen in Figure 4.26.

Clicking on Descriptives opens up the menu seen in Figure 4.27. Notice that the variable "height" has already been moved to the "Variable(s)" box and that the box for "Save standardized values as variables" has been checked. Clicking "OK" starts the analysis.

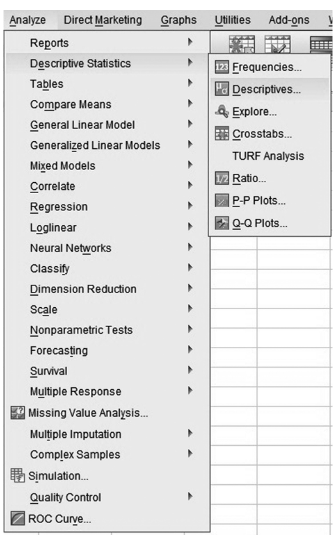

Figure 4.26 Commands for Saving *z* Scores in SPSS The "Descriptives" command can be found under "Analyze" and then "Descriptive Statistics." (Source: SPSS)

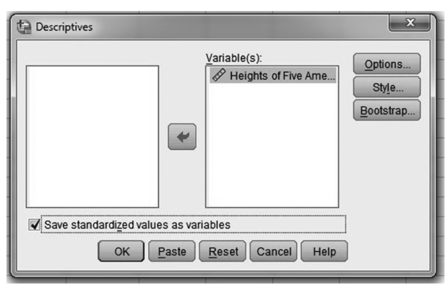

Figure 4.27 Menu for the "Descriptives" Command Clicking on "Descriptives" in Figure 4.26 opens up this menu. Note that the variable for which the *z* scores, or "standardized values" as SPSS calls them here, are to be calculated has been moved to the "Variable(s)" box. And, importantly, the box has been checked to "Save standardized values as variables." (Source: SPSS)

Corty_3e_Ch04_107-144_Final HR.indd 142 21/01/16 5:23 pm

Height	ZHeight
62	-1.19523
65	-.47809
66	-.23905
69	.47809
73	1.43427

Figure 4.28 SPSS Output for *z* scores for Heights of Five Americans Compare the number of variables in this data file to the number seen in the upper left of Figure 4.27. The z scores for height have been given the name Zheight and added to the data set. (Source: SPSS)

Look at Figure 4.28, which shows the data editor for the demographer's sample of five Americans. Notice that there are two variables listed: height and Zheight. Zheight is the name assigned to the *z* score for height by SPSS. The *z* scores that SPSS calculates carry more decimal places throughout the calculations, but they match up well to the ones found for these data earlier in this chapter.

Corty_3e_Ch04_107-144_Final HR.indd 144 21/01/16 5:23 pm