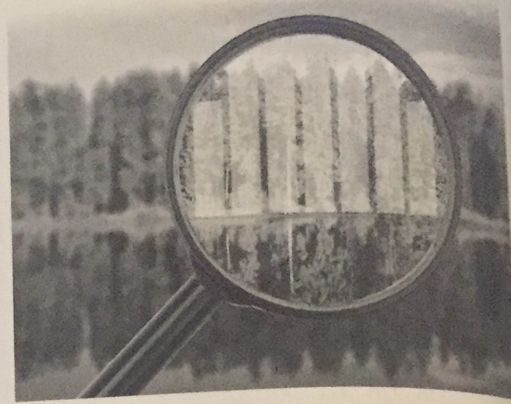
CHAPTER 8



Bell-Shaped Curves and Other Shapes

Thought Questions

1. The heights of adult women in the United States follow, at least approximately, a bell-shaped curve. What do you think this means?
2. What does it mean to say that a man's weight is in the 30th percentile for all adult males?
3. A "standardized score" is simply the number of standard deviations an individual score falls above or below the mean for the whole group. (Values above the mean have positive standardized scores, whereas those below the mean have negative ones.) Adult male heights in the United States have a mean of 70 inches and a standard deviation of 3 inches. Adult female heights in the United States have a mean of 65 inches and a standard deviation of 2 1/2 inches. Thus, a man who is 73 inches tall has a standardized score of 1. What is the standardized score corresponding to your own height, compared to adults of your sex in the United States?
4. Data sets consisting of physical measurements (heights, weights, lengths of bones, and so on) for adults of the same species and sex tend to follow a similar pattern. The pattern is that most individuals are clumped around the average, with numbers decreasing the farther values are from the average in either direction. Describe what shape a histogram of such measurements would have.

8.1 Populations, Frequency Curves, and Proportions

In Chapter 7, we learned how to draw a picture of a set of data and how to think about its shape. In this chapter, we learn how to extend those ideas to pictures and shapes for populations of measurements. For example, in Figure 7.5 we illustrated that, based on a sample of 199 men, heights of adult British males are reasonably bell-shaped. Because the men were a representative sample, the picture for all of the millions of British men is probably similar. But even if we could measure them all, it would be difficult to construct a histogram with so much data. What is the best way to represent the shape of a large population of measurements?

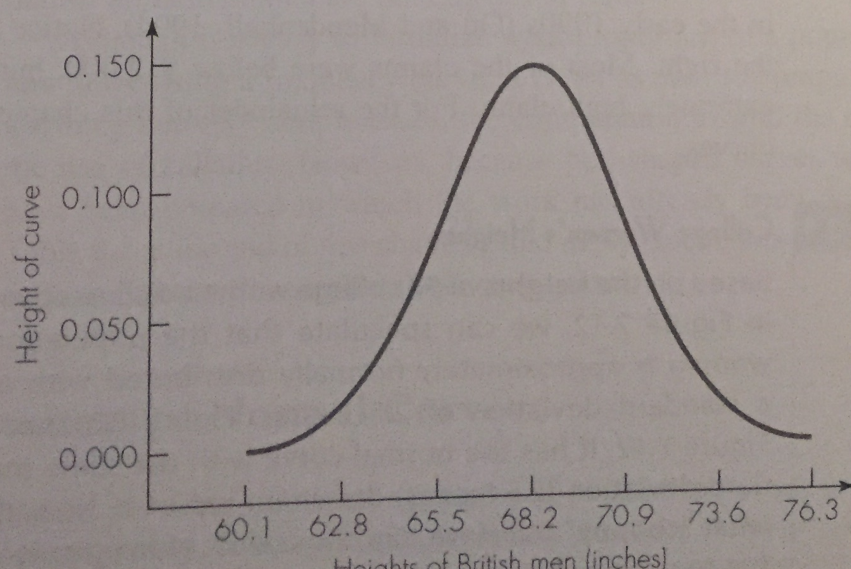
Frequency Curves

The most common type of picture for a population is a smooth **frequency curve.** Rather than drawing lots of tiny rectangles, the picture is drawn as if the tops of the rectangles were connected with a smooth curve. Figure 8.1 illustrates a frequency curve for the population of British male heights, based on the assumption that the heights shown in the histogram in Figure 7.5 are representative of all British men's heights. Notice that the picture is similar to the histogram in Figure 7.5, except that the curve is smooth and the heights have been converted to inches. The mean and standard deviation for the 199 men in the sample are 68.2 inches and 2.7 inches, respectively, and were used to draw Figure 8.1. Later in this chapter you will learn how to draw a picture like this, based on knowing only the mean and standard deviation.

Notice that the vertical scale is simply labeled "height of curve." This height is determined by sizing the curve so that the area under the entire curve is 1, for reasons that will become clear in the next few pages. Unlike with a histogram, the height of the curve cannot be interpreted as a proportion or frequency, but is chosen simply to satisfy the rule that the entire area under the curve is 1.

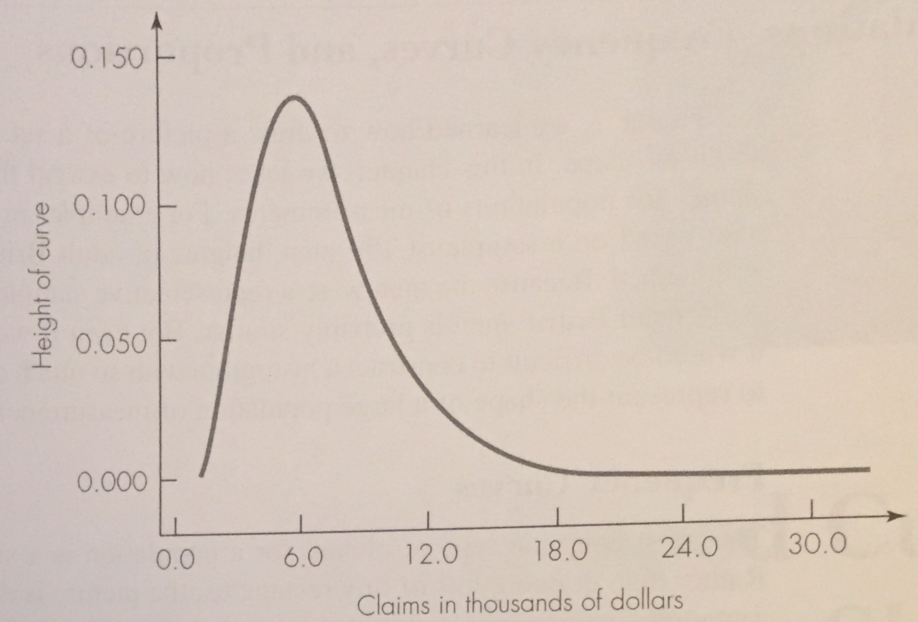
The bell shape illustrated in Figure 8.1 is so common that if a population has this shape, the measurements are said to follow a **normal distribution.** Equivalently,

Figure 8.1
A normal frequency curve



Unless otherwise noted, all content on this page is © Cengage Learning.

Figure 8.2
A nonnormal frequency curve



they are said to follow a **bell-shaped curve,** a **normal curve,** or a **Gaussian curve.** This last name comes from the name of Karl Friedrich Gauss (1777–1855), who was one of the first mathematicians to investigate the shape.

A population of measurements that follow a normal distribution is said to be **normally distributed**. For instance, we would say that the heights of British males are (approximately) normally distributed with a mean of 68.2 inches and a standard deviation of 2.7 inches. Because normal distributions are symmetric, the mean and median are equal. For instance, the median height of British males would also be 68.2 inches, so about half of all British males are taller than this height, and about half are shorter.

Not all frequency curves are bell-shaped. Figure 8.2 shows a likely frequency curve for the population of dollar amounts of car insurance damage claims for a Midwestern city in the United States, based on data from 187 claims in that city in the early 1990s (Ott and Mendenhall, 1994). Notice that the curve is skewed to the right. Most of the claims were below $12,000, but occasionally there was an extremely high claim. For the remainder of this chapter, we focus on bell-shaped curves.

EXAMPLE 8.1 **College Women's Heights**

Based on the heights of 94 college women discussed in Example 7.7 and displayed in Figure 7.12, we can speculate that the population of heights of all college women is approximately normally distributed with a mean of 64.5 inches and a standard deviation of 2.5 inches. Figure 8.3 shows the same histogram as Figure 7.12. It has the normal curve with the same mean (64.5 inches) and standard deviation (2.5 inches) drawn on top of it. Note that the histogram is somewhat "choppy" but if we had thousands of heights it would look much more like the smooth curve. ■

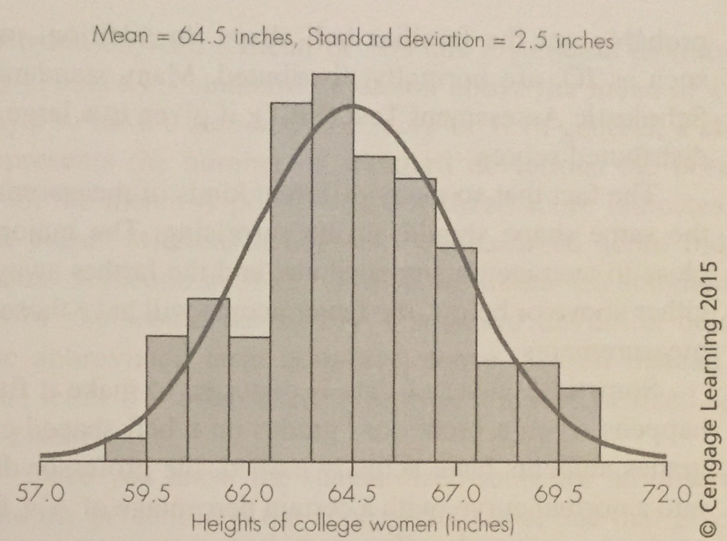


Figure 8.3
Heights of 94 college women (histogram) and normal curve for heights of all college women (smooth curve)

Proportions

Frequency curves are quite useful for determining what **proportion** or percentage of the population of measurements falls into a certain range. If we wanted to find out what proportion of the data fell into any particular range with a stemplot, we would count the number of leaves that were in that range and divide by the total. If we wanted to find the proportion in a certain range using a histogram, we would simply add up the heights of the rectangles for that range, assuming we had used proportions instead of counts for the heights. If not, we would add up the counts for that range and divide by the total number in the sample.

What if we have a frequency curve instead of a stemplot or histogram? Frequency curves are, by definition, drawn to make it easy to represent the proportion of the population falling into a certain range. Recall that they are drawn so the entire area underneath the curve is 1, or 100%. Therefore, to figure out what percentage or proportion of the population falls into a certain range, all you have to do is figure out how much of the area is situated over that range. For example, in Figure 8.1, half of the area is in the range above the mean height of 68.2 inches. In other words, about half of all British men are 68.2 inches or taller.

Although it is easy to visualize what proportion of a population falls into a certain range using a frequency curve, it is not as easy to compute that proportion. For anything but very simple cases, the computation to find the required area involves the use of calculus. However, because bell-shaped curves are so common, tables have been prepared in which the work has already been done (see, for example, Table 8.1 at the end of this chapter), and many calculators and computer applications such as Excel will compute these proportions.

8.2 The Pervasiveness of Normal Curves

Nature provides numerous examples of populations of measurements that, at least approximately, follow a normal curve. If you were to create a picture of the shape of almost any physical measurement within a homogeneous population, you would probably get the familiar bell shape. In addition, many psychological attributes, such as IQ, are normally distributed. Many standard academic tests, such as the Scholastic Assessment Test (SAT), if given to a large group, will result in normally distributed scores.

The fact that so many different kinds of measurements all follow approximately the same shape should not be surprising. The majority of people are somewhere close to average on any attribute, and the farther away you move from the average, either above or below, the fewer people will have those more extreme values for their measurements.

Sometimes a set of data is distorted to make it fit a normal curve. That's what happens when a professor "grades on a bell-shaped curve." Rather than assign the grades students have actually earned, the professor distorts them to make them fit into a normal curve, with a certain percentage of A's, B's, and so on. In other words, grades are assigned *as if* most students were average, with a few strong ones at the top and a few weak ones at the bottom. Unfortunately, this procedure has a tendency to artificially spread out clumps of students who are at the top or bottom of the scale, so that students whose original grades were very close together may receive different letter grades.

8.3 Percentiles and Standardized Scores

Percentiles

Have you ever wondered what percentage of the population of your sex is taller than you are, or what percentage of the population has a lower IQ than you do? Your **percentile** in a population represents the position of your measurement in comparison with everyone else's. It gives the percentage of the population that falls *below* you. If you are in the 50th percentile, it means that exactly half of the population falls below you. If you are in the 98th percentile, 98% of the population falls below you and only 2% is above you.

Your percentile is easy to find if the population of values has an approximate bell shape and if you have just three pieces of information. All you need to know are your own value and the mean and standard deviation for the population. Although there are obviously an unlimited number of potential bell-shaped curves, depending on the magnitude of the particular measurements, each one is completely determined once you know its mean and standard deviation. In addition, each one can be "standardized" in such a way that the same table can be used to find percentiles for any of them.

Standardized Scores

Suppose you knew your IQ was 115, as measured by the 5th edition of the Stanford-Binet IQ test. (Earlier editions had a standard deviation of 16, but the more recent 5th edition has a standard deviation of 15.) Scores from that test have a normal distribution with a mean of 100 and a standard deviation of 15. Therefore, your IQ is exactly 1 standard deviation above the mean of 100. In this case, we would say you have a *standardized score* of 1. In general, a **standardized score** simply represents the number of standard deviations the observed value or score falls from the mean. A positive standardized score indicates an observed value above the mean, whereas a negative standardized score indicates a value below the mean. Someone with an IQ of 85 would have a standardized score of −1 because he or she would be exactly 1 standard deviation below the mean. Sometimes the abbreviated term **standard score** is used instead of "standardized score." The letter z is often used to represent a standardized score, so another synonym is **z-score**.

Once you know the standardized score for an observed value, all you need to find the percentile is the appropriate table, one that gives percentiles for a normal distribution with a mean of 0 and a standard deviation of 1. (Calculators, statistical software, and free websites that provide these percentiles are also available.) A normal curve with a mean of 0 and a standard deviation of 1 is called a **standard normal curve.** It is the curve that results when any normal curve is converted to standardized scores. In other words, the standardized scores resulting from any normal curve will have a mean of 0 and a standard deviation of 1 and will retain the bell shape.

Table 8.1, presented at the end of this chapter, gives percentiles for standardized scores. For example, with an IQ of 115 and a standardized score of +1, you would be at the 84th percentile. In other words, your IQ would be higher than that of 84% of the population. If we are told the percentile for a score but not the value itself, we can also work backward from the table to find the value. Let's review the steps necessary to find a percentile from an observed value, and vice versa.

To find the percentile from an observed value:

1. Find the standardized score: (observed value − mean)/s.d., where s.d. = standard deviation. Don't forget to keep the plus or minus sign.
2. Look up the percentile in Table 8.1 (page 175).

To find an observed value from a percentile:

1. Look up the percentile in Table 8.1, and find the corresponding standardized score.
2. Compute the observed value: mean + (standardized score)(s.d), where s.d. = standard deviation.

EXAMPLE 8.2 **Is High Cholesterol Too Common?**

According to a publication from the World Health Organization (Lawes et al, 2004), cholesterol levels for women aged 30 to 44 in North American are approximately normally distributed with a mean of about 185 mg/dl and standard deviation of about 35 mg/dl. High cholesterol is defined as anything over 200 mg/dl. What proportion of women in this age group has high cholesterol? In other words, if a woman has an observed level of 200 mg/dl, what is her percentile?

Standardized score = *(observed value – mean)/(s.d.)*
Standardized score = *(200 – 185)/35*
Standardized score = *15/35* = *0.43.*

From Table 8.1, we see that a standardized score of 0.43 is between the 66th percentile score of 0.41 and the 67th percentile score of 0.44. Therefore, about 66.7%, or about two-thirds of women in this age group, do not have high cholesterol. That means that about one-third do have high cholesterol. Thus, if you fall into this category, you are not alone! In fact, according to the United States Center for Disease Control, the average total cholesterol level is 200 mg/dl. That means that about half of all adults are categorized as having high cholesterol. (*Source*: http://www.cdc.gov/cholesterol/facts.htm, accessed June 10, 2013) ■

EXAMPLE 8.3 **Tragically Low IQ**

In the Edinburgh newspaper the *Scotsman* on March 8, 1994, a headline read, "Jury urges mercy for mother who killed baby" (p. 2). The baby had died from improper care. One of the issues in the case was that "the mother . . . had an IQ lower than 98 percent of the population, the jury had heard." From this information, let's compute the mother's IQ. If it was lower than 98% of the population, it was higher than only 2%, so she was in the 2nd percentile. From Table 8.1, we see that her standardized score was −2.05, or 2.05 standard deviations below the mean of 100. We can now compute her IQ:

observed value = *mean* + *(standardized score)(s.d.)*
observed value = *100* + *(−2.05)(15)*
observed value = *100* + *(−30.8)* = *100* − *30.8*
observed value = *69.2*

Thus, her IQ was about 69. The jury was convinced that her IQ was, tragically, too low to expect her to be a competent mother. ■

EXAMPLE 8.4 **Calibrating Your GRE Score**

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. For people who took the exam between August 2011 and April 2012, the mean for the quantitative reasoning portion of the exam was 151.3 and the standard deviation was 8.7 (Educational Testing Service, 2012). If you had received a score of 163 on that GRE exam, what percentile would you be in, assuming the scores were bell-shaped? We can compute your percentile by first computing your standardized score:

standardized score = (observed value − mean)/(s.d.)
standardized score = (163 − 151.3)/8.7
standardized score = 11.7/8.7 = 1.34

From Table 8.1, we see that a standardized score of 1.34 is at the 91st percentile. In other words, your score was higher than about 91% of the population. Figure 8.4 illustrates the GRE score of 163 for the population of GRE scores and the corresponding standardized score of 1.34 for the standard normal curve. Notice the similarity of the two pictures.

The Educational Testing Service publishes tables showing the exact percentile for various scores on the GRE. For those who took the exam between August 2011 and April 2012, 88% scored below 163 on the quantitative reasoning part. The actual value of 88% is thus very close to the value of 91% that we calculated based only on knowing the mean and standard deviation of the scores. ■

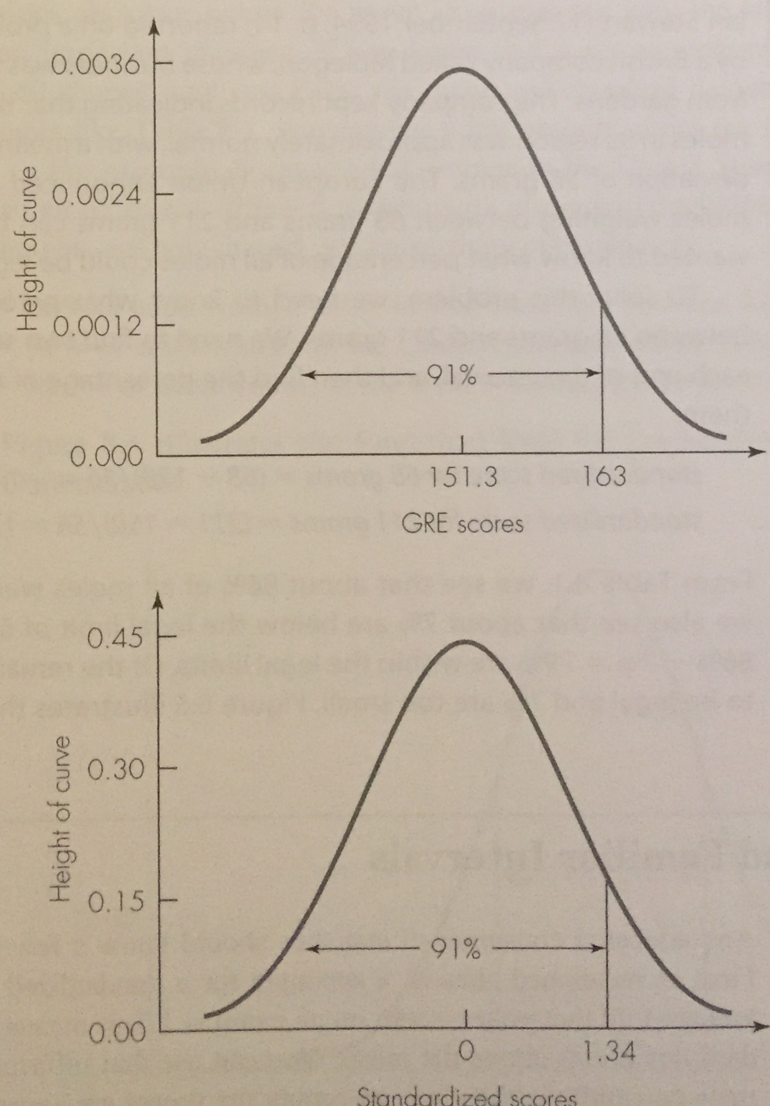
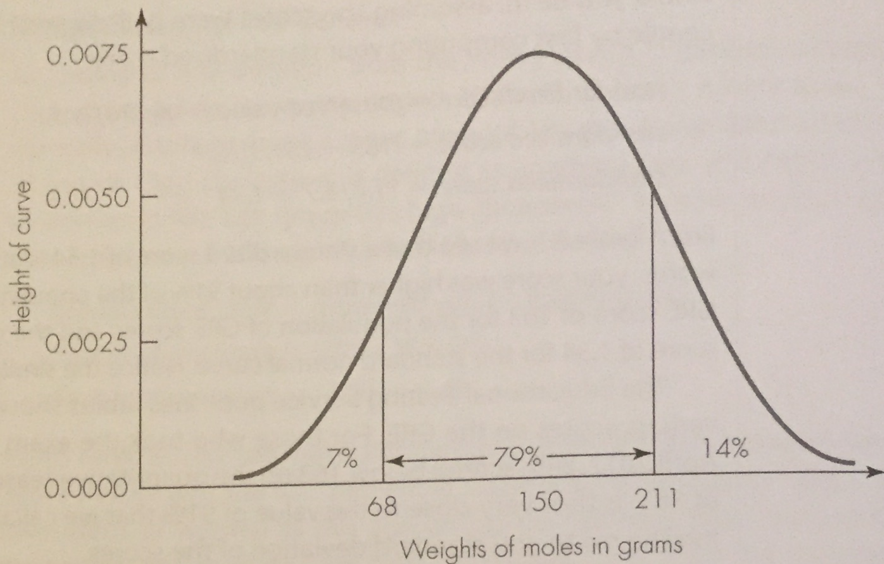


Figure 8.4
The percentile for a GRE quantitative reasoning score of 163 and corresponding standardized score

Unless otherwise noted, all content on this page is © Cengage Learning.

Figure 8.5
Moles inside and outside the legal limits of 68 to 211 grams



EXAMPLE 8.5 Ian Stewart (17 September 1994, p. 14) reported on a problem posed to a statistician by a British company called Molegon, whose business was to remove unwanted moles from gardens. The company kept records indicating that the population of weights of moles in its region was approximately normal, with a mean of 150 grams and standard deviation of 56 grams. The European Union announced that, starting in 1995, only moles weighing between 68 grams and 211 grams can be legally caught. Molegon wanted to know what percentage of all moles could be legally caught.

To solve this problem, we need to know what percentage of all moles weigh between 68 grams and 211 grams. We need to find two standardized scores, one for each end of the interval, and then find the percentage of the curve that lies between them:

$$\textit{standardized score for 68 grams} = (68 - 150)/56 = -1.46$$
$$\textit{standardized score for 211 grams} = (211 - 150)/56 = 1.09$$

From Table 8.1, we see that about 86% of all moles weigh 211 grams or less. But we also see that about 7% are below the legal limit of 68 grams. Therefore, about 86% − 7% = 79% are within the legal limits. Of the remaining 21%, 14% are too big to be legal and 7% are too small. Figure 8.5 illustrates this situation. ■

8.4 *z*-Scores and Familiar Intervals

Any educated consumer of statistics should know a few facts about normal curves. First, as mentioned already, a synonym for a standardized score is a **z-score.** Thus, if you are told that your *z*-score on an exam is 1.5, it means that your score is 1.5 standard deviations above the mean. You can use that information to find your approximate percentile in the class, assuming the scores are approximately bell-shaped.

Unless otherwise noted, all content on this page is © Cengage Learning

Second, some easy-to-remember intervals can give you a picture of where values on any normal curve will fall. This information is known as the **Empirical Rule.**

> **Empirical Rule**
>
> For any normal curve, approximately
>
> 68% of the values fall within 1 standard deviation of the mean in either direction;
>
> 95% of the values fall within 2 standard deviations of the mean in either direction;
>
> 99.7% of the values fall within 3 standard deviations of the mean in either direction.

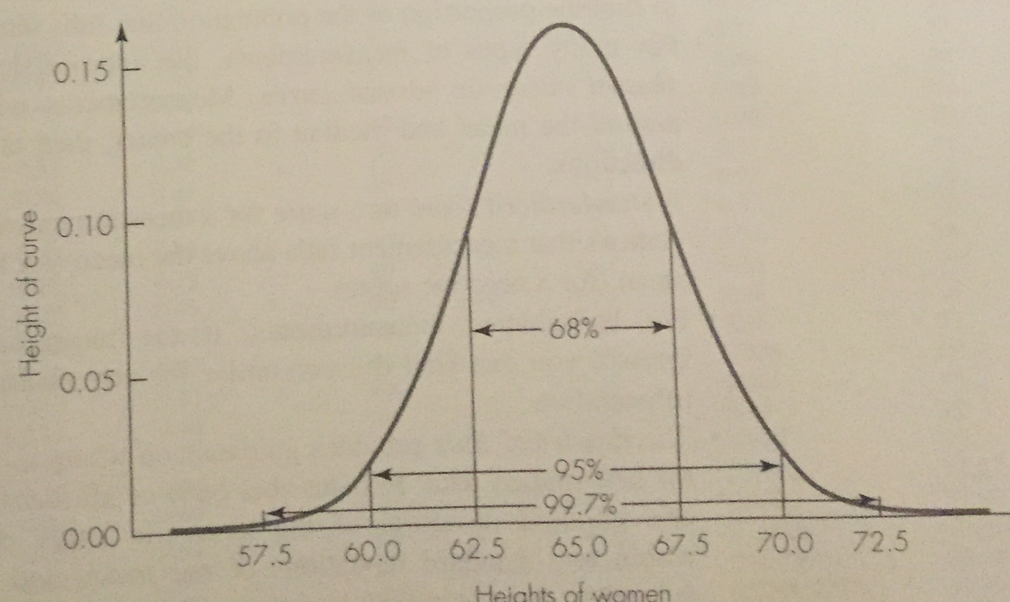
A measurement would be an extreme outlier if it fell more than 3 standard deviations above or below the mean. You can see why the standard deviation is such an important measure. If you know that a set of measurements is approximately bell-shaped, and you know the mean and standard deviation, then even without a table like Table 8.1, you can say a fair amount about the magnitude of the values.

For example, because adult women in the United States have a mean height of about 65 inches (5 feet 5 inches) with a standard deviation of about 2.5 inches, and heights are bell-shaped, we know that approximately

- 68% of adult women in the United States are between 62.5 inches and 67.5 inches
- 95% of adult women in the United States are between 60 inches and 70 inches
- 99.7% of adult women in the United States are between 57.5 inches and 72.5 inches

Figure 8.6 illustrates the Empirical Rule for the heights of adult women in the United States.

Figure 8.6
The Empirical Rule for heights of adult women



Unless otherwise noted, all content on this page is © Cengage Learning.

The mean height for adult males in the United States is about 70 inches and the standard deviation is about 3 inches. You can easily compute the ranges into which 68%, 95%, and almost all men's heights should fall.

Using Computers to Find Normal Curve Proportions

There are computer programs and websites that will find the proportion of a normal curve that falls below a specified value, above a value, and between two values. For example, here are two useful Excel functions:

> **NORMSDIST(value)** provides the proportion of the standard normal curve below the value. Example: NORMSDIST(1) = .8413, which rounds to .84, shown in Table 8.1 for $z = 1$.
>
> **NORMDIST(value,mean,s.d.,1)** provides the proportion of a normal curve with the specified mean and standard deviation (s.d.) that lies below the value given. (If the last number in parentheses is 0 instead of 1, it gives you the height of the curve at that value, which isn't much use to you. The "1" tells it that you want the proportion below the value.) Example: NORMDIST(67.5,65,2.5,1) = .8413, representing the proportion of adult women with heights of 67.5 inches or less.

Thinking About Key Concepts

- A *frequency curve* shows the possible values for a measurement and can be used to find the proportion of the population that falls into various ranges.
- For many types of measurements, the appropriate frequency curve is a *bell-shaped curve* or *normal curve*. Measurements with this shape are clustered around the mean and median in the center, then tail off symmetrically in both directions.
- A *standardized score* or *z-score* for a measurement is the number of standard deviations that measurement falls above the mean (for a positive score) or below the mean (for a negative score).
- For bell-shaped measurements, if the mean and standard deviation are known, you can find the percentile for any measurement with no additional information.
- The *Empirical Rule* provides guidance on where to expect measurements to fall for bell-shaped data. It states that 68% of all measurements should fall within one standard deviation of the mean (in either direction), 95% should fall within two standard deviations of the mean, and almost all (99.7%) should fall within three standard deviations of the mean.

TABLE 8.1 Proportions and Percentiles for Standard Normal Scores

Standard Score, z	Proportion Below z	Percentile	Standard Score, z	Proportion Below z	Percentile
−6.00	0.000000001	0.0000001	0.03	0.51	51
−5.20	0.0000001	0.00001	0.05	0.52	52
−4.26	0.00001	0.001	0.08	0.53	53
−3.00	0.0013	0.13	0.10	0.54	54
−2.576	0.005	0.50	0.13	0.55	55
−2.33	0.01	1	0.15	0.56	56
−2.05	0.02	2	0.18	0.57	57
−1.96	0.025	2.5	0.20	0.58	58
−1.88	0.03	3	0.23	0.59	59
−1.75	0.04	4	0.25	0.60	60
−1.64	0.05	5	0.28	0.61	61
−1.55	0.06	6	0.31	0.62	62
−1.48	0.07	7	0.33	0.63	63
−1.41	0.08	8	0.36	0.64	64
−1.34	0.09	9	0.39	0.65	65
−1.28	0.10	10	0.41	0.66	66
−1.23	0.11	11	0.44	0.67	67
−1.17	0.12	12	0.47	0.68	68
−1.13	0.13	13	0.50	0.69	69
−1.08	0.14	14	0.52	0.70	70
−1.04	0.15	15	0.55	0.71	71
−1.00	0.16	16	0.58	0.72	72
−0.95	0.17	17	0.61	0.73	73
−0.92	0.18	18	0.64	0.74	74
−0.88	0.19	19	0.67	0.75	75
−0.84	0.20	20	0.71	0.76	76
−0.81	0.21	21	0.74	0.77	77
−0.77	0.22	22	0.77	0.78	78
−0.74	0.23	23	0.81	0.79	79
−0.71	0.24	24	0.84	0.80	80
−0.67	0.25	25	0.88	0.81	81
−0.64	0.26	26	0.92	0.82	82
−0.61	0.27	27	0.95	0.83	83
−0.58	0.28	28	1.00	0.84	84
−0.55	0.29	29	1.04	0.85	85
−0.52	0.30	30	1.08	0.86	86
−0.50	0.31	31	1.13	0.87	87
−0.47	0.32	32	1.17	0.88	88
−0.44	0.33	33	1.23	0.89	89
−0.41	0.34	34	1.28	0.90	90
−0.39	0.35	35	1.34	0.91	91
−0.36	0.36	36	1.41	0.92	92
−0.33	0.37	37	1.48	0.93	93
−0.31	0.38	38	1.55	0.94	94
−0.28	0.39	39	1.64	0.95	95
−0.25	0.40	40	1.75	0.96	96
−0.23	0.41	41	1.88	0.97	97
−0.20	0.42	42	1.96	0.975	97.5
−0.18	0.43	43	2.05	0.98	98
−0.15	0.44	44	2.33	0.99	99
−0.13	0.45	45	2.576	0.995	99.5
−0.10	0.46	46	3.00	0.9987	99.87
−0.08	0.47	47	3.75	0.9999	99.99
−0.05	0.48	48	4.26	0.99999	99.999
−0.03	0.49	49	5.20	0.9999999	99.99999
0.00	0.50	50	6.00	0.999999999	99.9999999

Focus on Formulas

Notation for a Population

The lowercase Greek letter "mu" (μ) represents the **population mean.**

The lowercase Greek letter "sigma" (σ) represents the **population standard deviation.**

Therefore, the **population variance** is represented by σ^2.

A **normal distribution** with a mean of μ and variance of σ^2 is denoted by $N(\mu, \sigma^2)$.

For example, the **standard normal distribution** is denoted by $N(0, 1)$.

Standardized Score z for an Observed Value x

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

Observed Value x for a Standardized Score z

$$x = \mu + z\sigma$$

Empirical Rule

If a population of values is $N(\mu, \sigma^2)$, then approximately:

68% of values fall within the interval $\mu \pm \sigma$

95% of values fall within the interval $\mu \pm 2\sigma$

99.7% of values fall within the interval $\mu \pm 3\sigma$

Exercises

Exercises with numbers divisible by 3 (3, 6, 9, etc.) are included in the Solutions at the back of the book. They are marked with an asterisk (*).

1. Using Table 8.1, a computer, or a calculator, determine the percentage of the population falling *below* each of the following standard scores:
   a. −1.00
   b. 1.96
   c. 0.84
2. In each of the following cases, explain how you know that the population of measurements could not be normally distributed.
   a. The population is all families in the world that have exactly four children, and the measurement is the number of boys in the family.
   b. The population is all college students and the measurement is number of hours per week the student exercises. The mean is 4.5 hours, and the standard deviation is also 4.5 hours.

*3. Using Table 8.1, a computer, or a calculator, determine the percentage of the population falling *above* each of the following standard scores:
   *a. 1.28
   *b. −0.25
   *c. 2.33