

## JOHN W. TUKEY

### The Philosopher of Data Analysis

He started as a chemist, became a mathematician, and was converted to statistics by what he called “the real problems experience and the real data experience” of war work during the Second World War. *John W. Tukey (1915–2000)*

came to Princeton University in 1937 to study chemistry but took a doctorate in mathematics in 1939. During the war, he worked on the accuracy of range finders and of gunfire from bombers, among other problems. After the war he divided his time between Princeton and nearby Bell Labs, at that time the world’s leading industrial research group.

Tukey devoted much of his attention to the statistical study of messy problems with complex data: the safety of anesthetics used by many doctors in many hospitals on many patients, the Kinsey studies of human sexual behavior, monitoring compliance with a nuclear test ban, and air quality and environmental pollution.

From this “real problems experience and real data experience,” John Tukey developed exploratory data analysis. He invented some of the tools we have met, such as boxplots and stemplots. More important, he developed a philosophy for data analysis that changed the way statisticians think. In this chapter, as in Chapter 1, the approach we take in examining data follows Tukey’s path.

*Tukey was converted to statistics by “the real problems experience and the real data experience” during the Second World War.*

# chapter 2

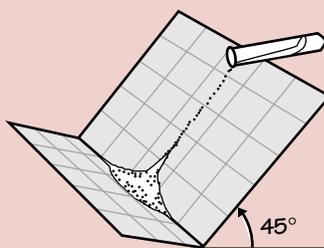
## The Normal Distributions

- 2.1 Density Curves and the Normal Distributions
- 2.2 Standard Normal Calculations
- Chapter Review

**ACTIVITY 2A** A Fine-Grained Distribution

*Materials: Sheet of grid paper; salt; can of spray paint; paint easel; newspapers*

1. Place the grid paper on the easel with a horizontal fold as shown, at about a  $45^\circ$  angle to the horizontal. Provide a “lip” at the bottom to catch the salt. Place newspaper behind the grid and extending out on all sides so you will not get paint on the easel.



2. Pour a stream of salt slowly from a point near the middle of the top edge of the grid. The grains of salt will hop and skip their way down the grid as they collide with one another and bounce left and right. They will accumulate at the bottom, piled against the grid, with the smooth profile of a bell-shaped curve, known as a normal distribution. We will learn about the normal distribution in this chapter.
3. Now carefully spray the grid-salt and all-with paint. Then discard the salt. You should be able to easily measure the height of the curve at different places by simply counting lines on the grid, or you could approximate areas by counting small squares or portions of squares on the grid.

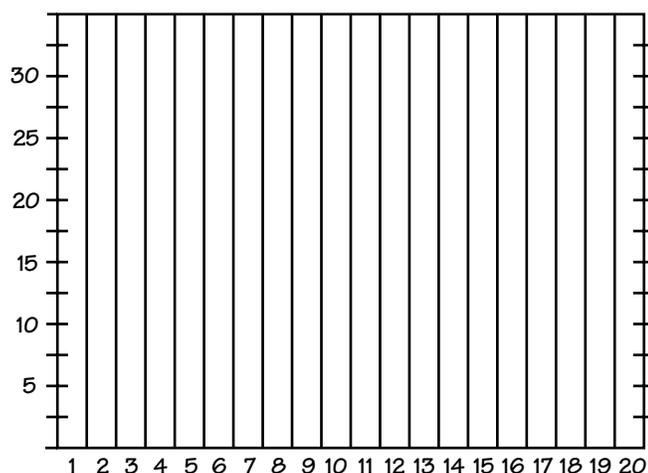
How could you get a tall, narrow curve? How could you get a short, broad curve? What factors might affect the height and breadth of the curve? From the members of the class, collect a set of normal curves that differ from one another.

**ACTIVITY 2B** Roll a Normal Distribution

*Materials: Several marbles, all the same size; two metersticks for a “ramp”; a ruled sheet of paper; a flat table about 4 feet long; carbon paper; Scotch Tape or masking tape*

**ACTIVITY 2B** Roll a Normal Distribution (*continued*)

1. At one end of the table prop up the two metersticks in a “V” shape to provide a ramp for the marbles to roll down. The marble will roll down the chute, continue across the table, and fall off the table to the floor below. Make sure that the ramp is secure and that the tabletop does not have any grooves or obstructions.
2. Roll the marble down the ramp several times to get a good idea of the area of the floor where the marble will fall.
3. Center the ruled sheet of paper (see Figure 2.1) over this area, face up, with the bottom edge toward the table and parallel to the edge of the table. The ruled lines should go in the same direction as the marble’s path. Tape the sheet securely to the floor. Place the sheet of carbon paper, carbon side down, over the ruled sheet.
4. Roll the marble for a class total of 200 times. The spots where it hits the floor will be recorded on the ruled paper as black dots. When the marble hits the floor, it will probably bounce, so try to catch it in midair after the impact so that you don’t get any extra marks. After the first 100 rolls, replace the sheet of paper. This will make it easier for you to count the spots. Make sure that the second sheet is in exactly the same position as the first one.
5. When the marble has been rolled 200 times, make a histogram of the distribution of the points as follows. First, count the number of dots in each column. Then graph this number by drawing horizontal lines in the columns at the appropriate level. Use the scale on the left-hand side of the sheet.

**FIGURE 2.1** Example of ruled sheet for Activity 2B.

## 2.1 DENSITY CURVES AND THE NORMAL DISTRIBUTIONS

We now have a kit of graphical and numerical tools for describing distributions. What is more, we have a clear strategy for exploring data on a single quantitative variable:

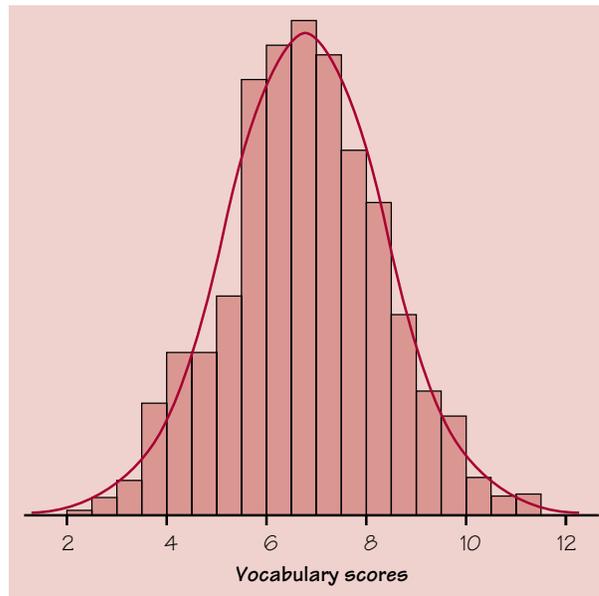
- Always plot your data: make a graph, usually a histogram or a stemplot.
- Look for the overall pattern (shape, center, spread) and for striking deviations such as outliers.
- Calculate a numerical summary to briefly describe the center and spread.

Here is one more step to add to the strategy:

- Sometimes the overall pattern of a large number of observations is so regular that we can describe it by a smooth curve.

### Density curves

Figure 2.2 is a histogram of the scores of all 947 seventh-grade students in Gary, Indiana, on the vocabulary part of the Iowa Test of Basic Skills.<sup>1</sup> Scores of many students on this national test have a quite regular distribution. The histogram is symmetric, and both tails fall off quite smoothly from a single center peak. There are no large gaps or obvious outliers. The smooth curve drawn through the tops of the histogram bars is a good description of the overall pattern of the data.



**FIGURE 2.2** Histogram of the vocabulary scores of all seventh-grade students in Gary, Indiana. The smooth curve shows the overall shape of the distribution.

The curve is a *mathematical model* for the distribution. A mathematical model is an idealized description. It gives a compact picture of the overall pattern of the data but ignores minor irregularities as well as any outliers.

We will see that it is easier to work with the smooth curve in Figure 2.2 than with the histogram. The reason is that the histogram depends on our choice of classes, while with a little care we can use a curve that does not depend on any choices we make. Here's how we do it.

### EXAMPLE 2.1 FROM HISTOGRAM TO DENSITY CURVE

Our eyes respond to the *areas* of the bars in a histogram. The bar areas represent proportions of the observations. Figure 2.3(a) is a copy of Figure 2.2 with the leftmost bars shaded. The area of the shaded bars in Figure 2.3(a) represents the students with vocabulary scores 6.0 or lower. There are 287 such students, who make up the proportion  $287/947 = 0.303$  of all Gary seventh graders.

Now concentrate on the curve drawn through the bars. In Figure 2.3(b), the area under the curve to the left of 6.0 is shaded. Adjust the scale of the graph so that *the total area under the curve is exactly 1*. This area represents the proportion 1, that is, all the observations. Areas under the curve then represent proportions of the observations. The curve is now a *density curve*. The shaded area under the density curve in Figure 2.3(b) represents the proportion of students with score 6.0 or lower. This area is 0.293, only 0.010 away from the histogram result. You can see that areas under the density curve give quite good approximations of areas given by the histogram.

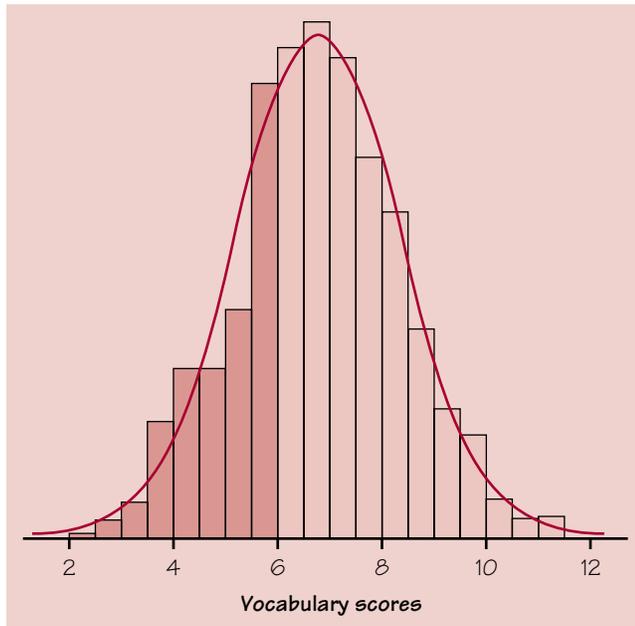
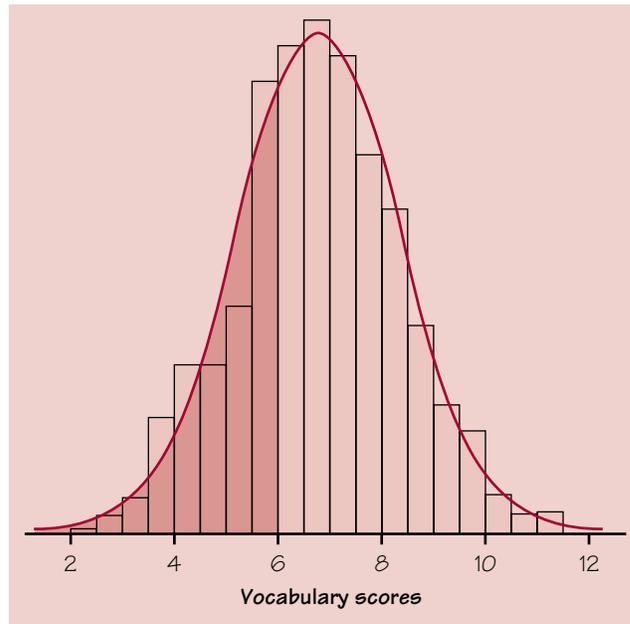


FIGURE 2.3(a) The proportion of scores less than or equal to 6.0 from the histogram is 0.303.



**FIGURE 2.3(b)** The proportion of scores less than or equal to 6.0 from the density curve is 0.293.

### DENSITY CURVE

A **density curve** is a curve that

- is always on or above the horizontal axis, and
- has area exactly 1 underneath it.

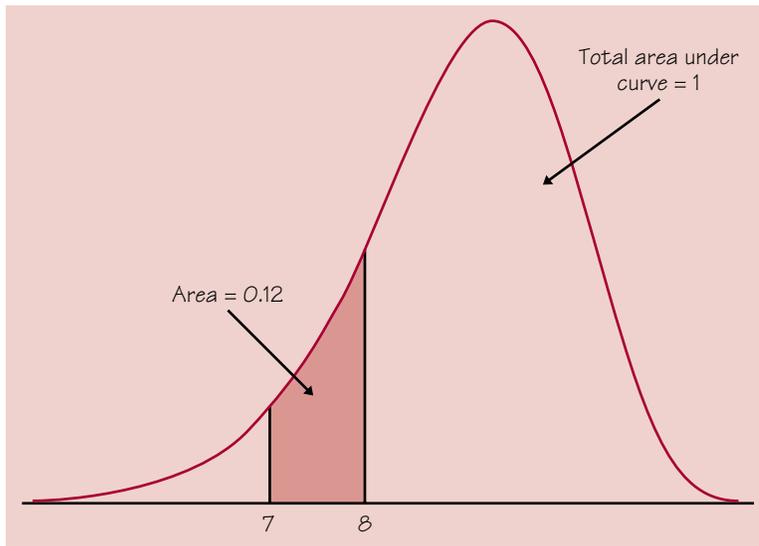
A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values is the proportion of all observations that fall in that range.

### *normal curve*

The density curve in Figures 2.2 and 2.3 is a **normal curve**. Density curves, like distributions, come in many shapes. In later chapters, we will encounter important density curves that are skewed to the left or right, and curves that may look like normal curves but are not.

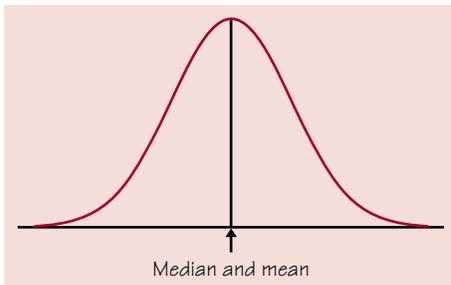
### EXAMPLE 2.2 A SKEWED-LEFT DISTRIBUTION

Figure 2.4 shows the density curve for a distribution that is slightly skewed to the left. The smooth curve makes the overall shape of the distribution clearly visible. The shaded area under the curve covers the range of values from 7 to 8. This area is 0.12. This means that the proportion 0.12 of all observations from this distribution have values between 7 and 8.

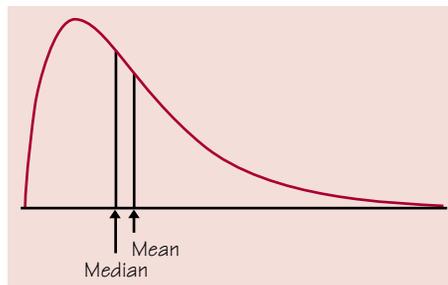


**FIGURE 2.4** The shaded area under this density curve is the proportion of observations taking values between 7 and 8.

Figure 2.5 shows two density curves: a symmetric normal density curve and a right-skewed curve. A density curve of the appropriate shape is often an adequate description of the overall pattern of a distribution. Outliers, which are deviations from the overall pattern, are not described by the curve. Of course, no set of real data is exactly described by a density curve. The curve is an approximation that is easy to use and accurate enough for practical use.



**FIGURE 2.5(a)** The median and mean of a symmetric density curve.



**FIGURE 2.5(b)** The median and mean of a right-skewed density curve.

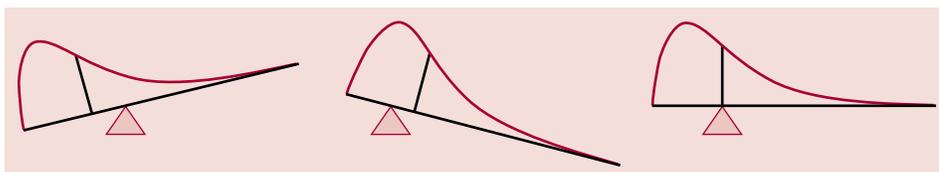
### The median and mean of a density curve

Our measures of center and spread apply to density curves as well as to actual sets of observations. The median and quartiles are easy. Areas under a density curve represent proportions of the total number of observations. The median is the point with half the observations on either side. So **the median of a density curve is the equal-areas point**, the point with half the area under the curve to its left and the remaining half of the area to its right. The quartiles divide the

area under the curve into quarters. One-fourth of the area under the curve is to the left of the first quartile, and three-fourths of the area is to the left of the third quartile. You can roughly locate the median and quartiles of any density curve by eye by dividing the area under the curve into four equal parts.

Because density curves are idealized patterns, a symmetric density curve is exactly symmetric. The median of a symmetric density curve is therefore at its center. Figure 2.5(a) shows the median of a symmetric curve. It isn't so easy to spot the equal-areas point on a skewed curve. There are mathematical ways of finding the median for any density curve. We did that to mark the median on the skewed curve in Figure 2.5(b).

What about the mean? The mean of a set of observations is their arithmetic average. If we think of the observations as weights strung out along a thin rod, the mean is the point at which the rod would balance. This fact is also true of density curves. **The mean is the point at which the curve would balance if made of solid material.** Figure 2.6 illustrates this fact about the mean. A symmetric curve balances at its center because the two sides are identical. **The mean and median of a symmetric density curve are equal**, as in Figure 2.5(a). We know that the mean of a skewed distribution is pulled toward the long tail. Figure 2.5(b) shows how the mean of a skewed density curve is pulled toward the long tail more than is the median. It's hard to locate the balance point by eye on a skewed curve. There are mathematical ways of calculating the mean for any density curve, so we are able to mark the mean as well as the median in Figure 2.5(b).



**FIGURE 2.6** The mean is the balance point of a density curve.

#### MEDIAN AND MEAN OF A DENSITY CURVE

The **median** of a density curve is the equal-areas point, the point that divides the area under the curve in half.

The **mean** of a density curve is the balance point, at which the curve would balance if made of solid material.

The median and mean are the same for a symmetric density curve. They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.

We can roughly locate the mean, median, and quartiles of any density curve by eye. This is not true of the standard deviation. When necessary, we can once again call on more advanced mathematics to learn the value of the standard deviation. The study of mathematical methods for doing calculations with density curves is part of theoretical statistics. Though we are concentrating on statistical practice, we often make use of the results of mathematical study.

Because a density curve is an idealized description of the distribution of data, we need to distinguish between the mean and standard deviation of the density curve and the mean  $\bar{x}$  and standard deviation  $s$  computed from the actual observations. The usual notation for the mean of an idealized distribution is  $\mu$  (the Greek letter mu). We write the standard deviation of a density curve as  $\sigma$  (the Greek letter sigma).

*mean  $\mu$*   
*standard deviation  $\sigma$*

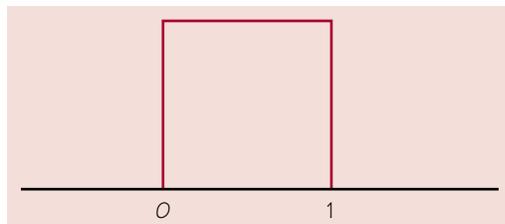
## EXERCISES

### 2.1 DENSITY CURVES

- (a) Sketch a density curve that is symmetric but has a shape different from that of the curve in Figure 2.5(a).
- (b) Sketch a density curve that is strongly skewed to the left.

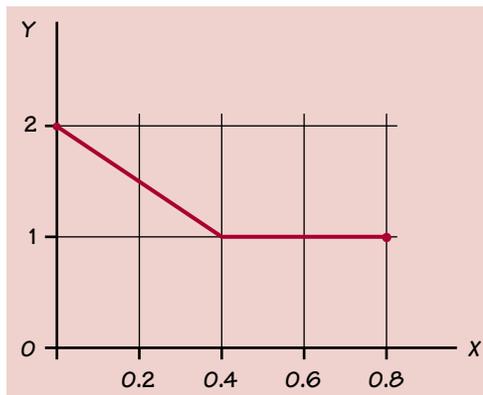
**2.2 A UNIFORM DISTRIBUTION** Figure 2.7 displays the density curve of a *uniform distribution*. The curve takes the constant value 1 over the interval from 0 to 1 and is zero outside the range of values. This means that data described by this distribution take values that are uniformly spread between 0 and 1. Use areas under this density curve to answer the following questions.

- (a) Why is the total area under this curve equal to 1?
- (b) What percent of the observations lie above 0.8?
- (c) What percent of the observations lie below 0.6?
- (d) What percent of the observations lie between 0.25 and 0.75?
- (e) What is the mean  $\mu$  of this distribution?



**FIGURE 2.7** The density curve of a uniform distribution.

**2.3 A WEIRD DENSITY CURVE** A line segment can be considered a density “curve,” as shown in Exercise 2.2. A “broken line” graph can also be considered a density curve. Figure 2.8 shows such a density curve.



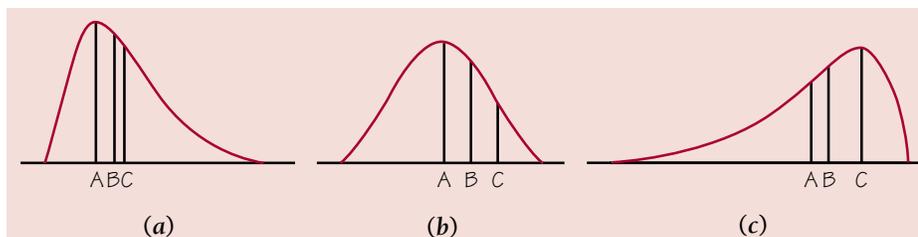
**FIGURE 2.8** An unusual “broken line” density curve.

(a) Verify that the graph in Figure 2.8 is a valid density curve.

For each of the following, use areas under this density curve to find the proportion of observations within the given interval:

- (b)  $0.6 \leq X \leq 0.8$
- (c)  $0 \leq X \leq 0.4$
- (d)  $0 \leq X \leq 0.2$
- (e) The median of this density curve is a point between  $X = 0.2$  and  $X = 0.4$ . Explain why.

**2.4 FINDING MEANS AND MEDIANS** Figure 2.9 displays three density curves, each with three points indicated. At which of these points on each curve do the mean and the median fall?



**FIGURE 2.9** Three density curves.

**2.5 ROLL A DISTRIBUTION** In this exercise you will pretend to roll a regular, six-sided die 120 times. Each time you roll the die, you will record the number on the up-face. The numbers 1, 2, 3, 4, 5, and 6 are called the *outcomes* of this chance experiment.

*outcomes*

In 120 rolls, how many of each number would you expect to roll? The TI-83 and TI-89 are useful devices for conducting chance experiments, especially ones like this that involve performing many repetitions. Because you are only pretending to roll the die repeatedly, we call this chance experiment a *simulation*. There will be a more formal treatment of simulations in Chapter 5.

*simulation*

- Begin by clearing  $L_1$  or list1 on your calculator.
- Use your calculator's random integer generator to generate 120 random whole numbers between 1 and 6 (inclusive), and then store these numbers in  $L_1$  or list1.



### TI-83

- Press **MATH**, choose PRB, then  $5:\text{RandInt}$ .
- Complete the command  $\text{RandInt}(1, 6, 120)$  **STO**  $\rightarrow$   $L_1$ .

### TI-89

- Press **CATALOG** **F3** and choose  $\text{randInt} \dots$
- Complete the command  $\text{tistat.randint}(1, 6, 120)$  **STO**  $\rightarrow$  list1.

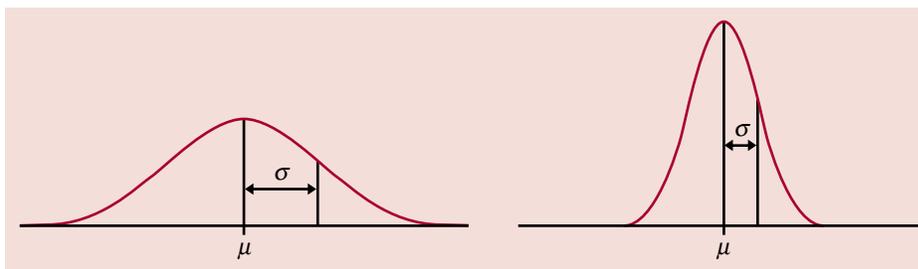
- Set the viewing window parameters:  $X[1, 7]_1$  by  $Y[-5, 25]_5$ .
- Specify a histogram using the data in  $L_1$ /list1.
- Then graph. Are you surprised? This is called a **frequency histogram** because it plots the **frequency** of each outcome (number of times each outcome occurred).
- Repeat the simulation several times. You can recall and reuse the previous command by pressing **2nd** **ENTER**. It's a good habit to clear  $L_1$ /list1 before you roll the die again.

In theory, of course, each number should come up 20 times. But in practice, there is chance variation, so the bars in the histogram will probably have different heights. Theoretically, what should the distribution look like?

## Normal distributions

One particularly important class of density curves has already appeared in Figures 2.2, 2.3, and 2.5(a) and the “fine-grained distribution” of Activity 2A. These density curves are symmetric, single-peaked, and bell-shaped. They are called *normal curves*, and they describe **normal distributions**. All normal distributions have the same overall shape. The exact density curve for a particular normal distribution is described by giving its mean  $\mu$  and its standard deviation  $\sigma$ . The mean is located at the center of the symmetric curve, and is the same as the median. Changing  $\mu$  without changing  $\sigma$  moves the normal curve along the horizontal axis without changing its spread. The standard deviation  $\sigma$  controls the spread of a normal curve. Figure 2.10 shows two normal curves

*normal distributions*

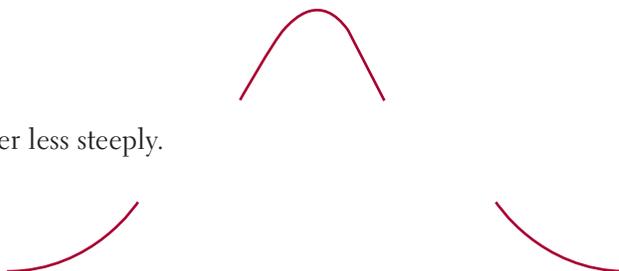


**FIGURE 2.10** Two normal curves, showing the mean  $\mu$  and standard deviation  $\sigma$ .

with different values of  $\sigma$ . The curve with the larger standard deviation is more spread out.

The standard deviation  $\sigma$  is the natural measure of spread for normal distributions. Not only do  $\mu$  and  $\sigma$  completely determine the shape of a normal curve, but we can locate  $\sigma$  by eye on the curve. Here's how. As we move out in either direction from the center  $\mu$ , the curve changes from falling ever more steeply

to falling ever less steeply.



### *inflection points*

The points at which this change of curvature takes place are called **inflection points** and are located at distance  $\sigma$  on either side of the mean  $\mu$ . Figure 2.10 shows  $\sigma$  for two different normal curves. You can feel the change as you run a pencil along a normal curve, and so find the standard deviation. Remember that  $\mu$  and  $\sigma$  alone do not specify the shape of most distributions, and that the shape of density curves in general does not reveal  $\sigma$ . These are special properties of normal distributions.

Why are the normal distributions important in statistics? Here are three reasons. First, normal distributions are good descriptions for some distributions of *real data*. Distributions that are often close to normal include scores on tests taken by many people (such as SAT exams and many psychological tests), repeated careful measurements of the same quantity, and characteristics of biological populations (such as lengths of cockroaches and yields of corn). Second, normal distributions are good approximations to the results of many kinds of *chance outcomes*, such as tossing a coin many times. Third, and most important, we will see that many *statistical inference* procedures based on normal distributions work well for other roughly symmetric distributions. However, even though many sets of data follow a normal distribution, many do not. Most income distributions, for example, are skewed to the right and so are not normal. Nonnormal data, like nonnormal people, not only are common but are sometimes more interesting than their normal counterparts.

### The 68–95–99.7 rule

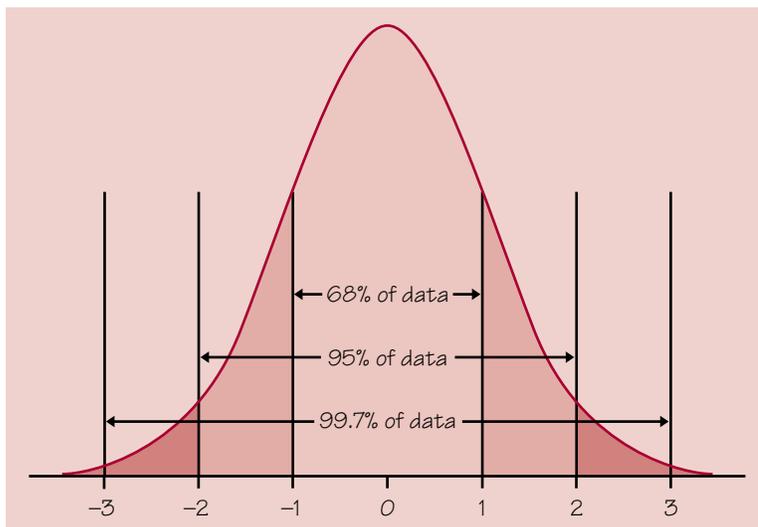
Although there are many normal curves, they all have common properties. In particular, all normal distributions obey the following rule.

**THE 68–95–99.7 RULE**

In the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ :

- 68% of the observations fall within  $\sigma$  of the mean  $\mu$ .
- 95% of the observations fall within  $2\sigma$  of  $\mu$ .
- 99.7% of the observations fall within  $3\sigma$  of  $\mu$ .

Figure 2.11 illustrates the 68–95–99.7 rule. Some authors refer to it as the “empirical rule.” By remembering these three numbers, you can think about normal distributions without constantly making detailed calculations, and when rough approximations will suffice.

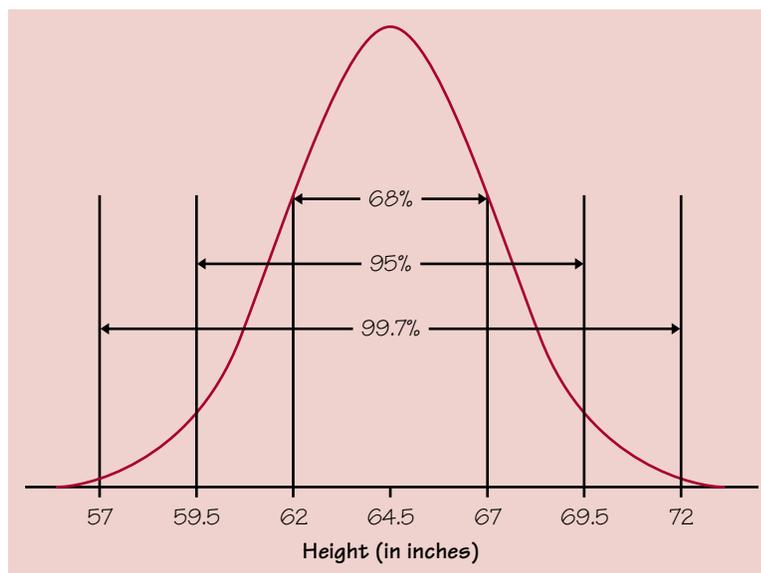


**FIGURE 2.11** The 68–95–99.7 rule for normal distributions.

**EXAMPLE 2.3 YOUNG WOMEN'S HEIGHTS**

The distribution of heights of young women aged 18 to 24 is approximately normal with mean  $\mu = 64.5$  inches and standard deviation  $\sigma = 2.5$  inches. Figure 2.12 shows the application of the 68–95–99.7 rule in this example.

Two standard deviations is 5 inches for this distribution. The 95 part of the 68–95–99.7 rule says that the middle 95% of young women are between  $64.5 - 5$  and  $64.5 + 5$  inches tall, that is, between 59.5 and 69.5 inches. This fact is exactly true for an exactly normal distribution. It is approximately true for the heights of young women because the distribution of heights is approximately normal.



**FIGURE 2.12** The 68–95–99.7 rule applied to the distribution of the heights of young women. Here,  $\mu = 64.5$  and  $\sigma = 2.5$ .

The other 5% of young women have heights outside the range from 59.5 to 69.5 inches. Because the normal distributions are symmetric, half of these women are on the tall side. So the tallest 2.5% of young women are taller than 69.5 inches.

The 99.7 part of the 68-95-99.7 rule says that almost all young women (99.7% of them) have heights between  $\mu - 3\sigma$  and  $\mu + 3\sigma$ . This range of heights is 57 to 72 inches.

Because we will mention normal distributions often, a short notation is helpful. We abbreviate the normal distribution with mean  $\mu$  and standard deviation  $\sigma$  as  $N(\mu, \sigma)$ . For example, the distribution of young women's heights is  $N(64.5, 2.5)$ .

National test scores are frequently reported in terms of percentiles, rather than raw scores. If your score on the math portion of such a test was reported as the 90th percentile, then 90% of the students who took the math test scored *lower than or equal to* your score. Percentiles are used when we are most interested in seeing where an individual observation stands relative to the other individuals in the distribution. Typically, in practice, the number of observations is quite large so that it makes sense to talk about the distribution as a density curve. The median score would be the 50th percentile because half the scores are to the left of (i.e., lower than) the median. The first quartile is the 25th percentile and the third quartile is the 75th percentile.

**EXERCISES**

**2.6 MEN'S HEIGHTS** The distribution of heights of adult American men is approximately normal with mean 69 inches and standard deviation 2.5 inches. Draw a normal curve on which this mean and standard deviation are correctly located. (Hint: Draw the curve first, locate the points where the curvature changes, then mark the horizontal axis.)

**2.7 MORE ON MEN'S HEIGHTS** The distribution of heights of adult American men is approximately normal with mean 69 inches and standard deviation 2.5 inches. Use the 68–95–99.7 rule to answer the following questions.

- (a) What percent of men are taller than 74 inches?
- (b) Between what heights do the middle 95% of men fall?
- (c) What percent of men are shorter than 66.5 inches?
- (d) A height of 71.5 inches corresponds to what percentile of adult male American heights?

**2.8 IQ SCORES** Scores on the Wechsler Adult Intelligence Scale (WAIS, a standard “IQ test”) for the 20 to 34 age group are approximately normally distributed with  $\mu = 110$  and  $\sigma = 25$ . Use the 68–95–99.7 rule to answer these questions.

- (a) About what percent of people in this age group have scores above 110?
- (b) About what percent have scores above 160?
- (c) In what range do the middle 95% of all IQ scores lie?

**2.9 WOMEN'S HEIGHTS** The distribution of heights of young women aged 18 to 24 is discussed in Example 2.3. Find the percentiles for the following heights.

- (a) 64.5 inches
- (b) 59.5 inches
- (c) 67 inches
- (d) 72 inches

**2.10 FINE-GRAINED DISTRIBUTION** You can do this exercise if you spray-painted a normal distribution in Activity 2A. On your “fine-grained distribution,” first count the number of whole squares and parts of squares under the curve. Approximate as best you can. This represents the total area under the curve.

- (a) Mark vertical lines at  $\mu - 1\sigma$  and  $\mu + 1\sigma$ . Count the number of squares or parts of squares between these two vertical lines. Now divide the number of squares within one standard deviation of  $\mu$  by the total number of squares under the curve and express your answer as a percent. How does this compare with 68%? Why would you expect your answer to differ somewhat from 68%?
- (b) Count squares to determine the percent of area within  $2\sigma$  of  $\mu$ . How does your answer compare with 95%?
- (c) Count squares to determine the percent of area within  $3\sigma$  of  $\mu$ . How does your answer compare with 99.7%?

## SUMMARY

We can sometimes describe the overall pattern of a distribution by a **density curve**. A density curve always remains on or above the horizontal axis and has total area 1 underneath it. An area under a density curve gives the proportion of observations that fall in a range of values.

A density curve is an idealized description of the overall pattern of a distribution that smooths out the irregularities in the actual data. Write the mean of a density curve as  $\mu$  and the standard deviation of a density curve as  $\sigma$  to distinguish them from the mean  $\bar{x}$  and the standard deviation  $s$  of the actual data.

The **mean**, the **median**, and the **quartiles** of a density curve can be located by eye. The mean  $\mu$  is the balance point of the curve. The median divides the area under the curve in half. The quartiles with the median divide the area under the curve into quarters. The **standard deviation**  $\sigma$  cannot be located by eye on most density curves.

The mean and median are equal for symmetric density curves. The mean of a skewed curve is located farther toward the long tail than is the median.

The **normal distributions** are described by a special family of bell-shaped symmetric density curves, called **normal curves**. The mean  $\mu$  and standard deviation  $\sigma$  completely specify a normal distribution  $N(\mu, \sigma)$ . The mean is the center of the curve, and  $\sigma$  is the distance from  $\mu$  to the inflection points on either side.

In particular, all normal distributions satisfy the **68–95–99.7 rule**, which describes what percent of observations lie within one, two, and three standard deviations of the mean.

An observation's percentile is the percent of the distribution that is at or to the left of the observation.

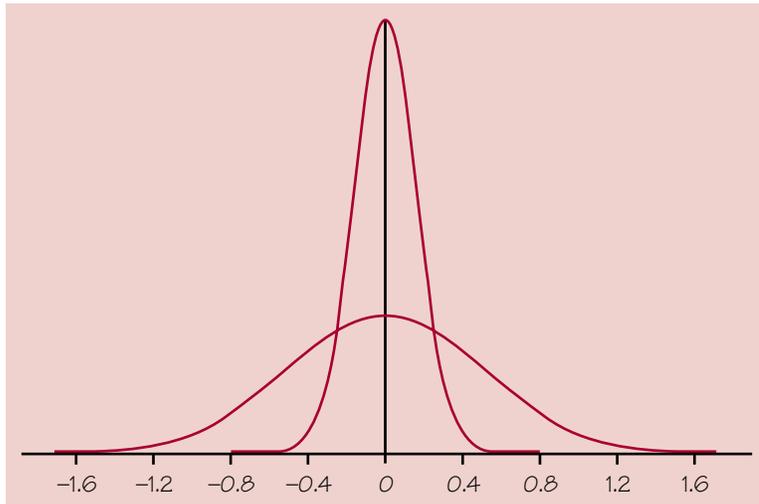
## SECTION 2.1 EXERCISES

**2.11 ESTIMATING STANDARD DEVIATIONS** Figure 2.13 shows two normal curves, both with mean 0. Approximately what is the standard deviation of each of these curves?

**2.12 HELMET SIZES** The army reports that the distribution of head circumference among male soldiers is approximately normal with mean 22.8 inches and standard deviation 1.1 inches. Use the 68–95–99.7 rule to answer the following questions.

- (a) What percent of soldiers have head circumference greater than 23.9 inches?
- (b) A head circumference of 23.9 inches would be what percentile?
- (c) What percent of soldiers have head circumference between 21.7 inches and 23.9 inches?

**2.13 GESTATION PERIOD** The length of human pregnancies from conception to birth varies according to a distribution that is approximately normal with mean 266 days and standard deviation 16 days. Use the 68–95–99.7 rule to answer the following questions.



**FIGURE 2.13** Two normal curves with the same mean but different standard deviations, for Exercise 2.11.

- (a) Between what values do the lengths of the middle 95% of all pregnancies fall?
- (b) How short are the shortest 2.5% of all pregnancies?
- (c) How long are the longest 2.5% of all pregnancies?

**2.14 IQ SCORES FOR ADULTS** Wechsler Adult Intelligence Scale (WAIS) scores for young adults are  $N(110, 25)$ .

- (a) If someone's score were reported as the 16th percentile, about what score would that individual have?
- (b) Answer the same question for the 84th percentile and the 97.5th percentile.

**2.15 WEIGHTS OF DISTANCE RUNNERS** A study of elite distance runners found a mean body weight of 63.1 kilograms (kg), with a standard deviation of 4.8 kg.

- (a) Assuming that the distribution of weights is normal, sketch the density curve of the weight distribution with the horizontal axis marked in kilograms.
- (b) Use the 68–95–99.7 rule to find intervals centered at the mean that will include 68%, 95%, and 99.7% of the weights of the runners.

**2.16 CALCULATOR GENERATED DENSITY CURVE** Like Minitab and similar computer utilities, the TI-83/TI-89 has a “random number generator” that produces decimal numbers between 0 and 1.

- On the TI-83, press  $\boxed{\text{MATH}}$ , then choose PRB and 1:Rand.
- On the TI-89, press  $\boxed{2\text{nd}} \boxed{5}$  (MATH), then choose 7:Probability and 4:Rand(. Be sure to close the parentheses.

Press  $\boxed{\text{ENTER}}$  several times to see the results. The command `2rand` (`2rand()` on the TI-89) produces a random number between 0 and 2. The density curve of the outcomes has constant height between 0 and 2, and height 0 elsewhere.

- What is the height of the density curve between 0 and 2? Draw a graph of the density curve.
- Use your graph from (a) and the fact that areas under the curve are relative frequencies of outcomes to find the proportion of outcomes that are less than 1.
- What is the median of the distribution? What are the quartiles?
- Find the proportion of outcomes that lie between 0.5 and 1.3.

**2.17 FLIP50** The program FLIP50 simulates flipping a fair coin 50 times and counts the number of times the coin comes up heads. It prints the number of heads on the screen. Then it repeats the experiment for a total of 100 times, each time displaying the number of heads in 50 flips. When it finishes, it draws a histogram of the 100 results. (You have to set up the plot first on the TI-89.)

- What outcomes are likely? What outcome(s) are the most likely? If you made a histogram of the results of the 100 replications, what shape distribution would you expect?
- The program is listed below. Enter the program carefully, or link it from a classmate or your teacher. Run the program and observe the variations in the results of the 100 replications.
- When the histogram appears, TRACE to see the classes and frequencies. Record the results in a frequency table.
- Describe the distribution: symmetric versus nonsymmetric; center; spread; number of peaks; gaps; suspected outliers. What shape density curve would best fit your distribution?

## TI-83

```
prgm:FLIP50
100→DIM(L1)
For(I,1,100)
0→H
For(J,1,50)
randInt(0,1)→N
If N=1:H+1→H
End
Disp H
H→L1(I)
End
PlotsOff
10→Xmin
40→Xmax
2→Xscl
-6→Ymin
```

## TI-89

```
flip50()
Prgm
tistat.clrlist(list1)
For i,1,100
0→h
For j,1,50
tistat.randint(0,1)→n
If n=1
h+1→h
EndFor
Disp h
h→list1[i]
EndFor
PlotsOff
10→xmin
40→xmax
```

```

25→Ymax          2→xscl
5→Yscl           -6→ymin
Plot1(Histogram,L1) 25→ymax
DispGraph        5→yscl
                  EndPrgm

```

Set up Plot 1 to be a histogram of list1 with a bucket width of 2. Then press  $\blacklozenge$   $\boxed{\text{F3}}$  (GRAPH).

**2.18 NORMAL DISTRIBUTION ON THE CALCULATOR** The normal density curves are defined by a particular equation:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We can obtain individual members of this family of curves by specifying particular values for the mean  $\mu$  and the standard deviation  $\sigma$ . If we specify the values  $\mu = 0$  and  $\sigma = 1$ , then we have the equation for the *standard* normal distribution. This exercise will explore two functions.

- Enter as  $Y_1$  the following equation for the standard normal distribution:

$$Y_1 = (1/\sqrt{(2\pi)}) ( e^{(-.5x^2)} )$$

- For  $Y_2$ , position your cursor after  $Y_2=$ . On the TI-83, press  $\boxed{2\text{nd}} \boxed{\text{VARS}}$  (DISTR) and choose  $1:\text{normalpdf}$  (. On the TI-89, press  $\boxed{\text{CATALOG}} \boxed{\text{F3}}$  (Flash Apps) and choose  $\text{normpdf}$  (. Finish defining  $Y_2$  as  $\text{normPdf}(x)$  ( $\text{tistat.normPdf}(x)$  on the TI-89).
- Turn off all plots and any functions other than  $Y_1$  and  $Y_2$ . Change the graph style for  $Y_2$  to a thick line by highlighting the slash  $\backslash$  to the left of  $Y_2$  and pressing  $\boxed{\text{ENTER}}$  once. (On the TI-89, press  $\boxed{2\text{nd}} \boxed{\text{F1}}$  ( $\boxed{\text{F6}}$ ) and choose 4:Thick.)
- Specify a viewing window  $X[-3,3]_1$  and  $Y[-0.1,0.5]_{0.1}$ .
- Press  $\boxed{\text{GRAPH}} (\blacklozenge \boxed{\text{F3}})$  on the TI-89.)

Write a sentence that describes the connection between these two functions.  
*Note:* normalpdf stands for “normal probability density function.” We’ll learn more about pdf’s in Chapter 8.

## 2.2 STANDARD NORMAL CALCULATIONS

### The standard normal distribution

As the 68–95–99.7 rule suggests, all normal distributions share many common properties. In fact, all normal distributions are the same if we measure in units

of size  $\sigma$  about the mean  $\mu$  as center. Changing to these units is called *standardizing*. To standardize a value, subtract the mean of the distribution and then divide by the standard deviation.

#### STANDARDIZING AND Z-SCORES

If  $x$  is an observation from a distribution that has mean  $\mu$  and standard deviation  $\sigma$ , the **standardized value** of  $x$  is

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

A standardized value is often called a **z-score**.

A z-score tells us how many standard deviations the original observation falls away from the mean, and in which direction. Observations larger than the mean are positive when standardized, and observations smaller than the mean are negative.

#### EXAMPLE 2.4 STANDARDIZING WOMEN'S HEIGHTS

The heights of young women are approximately normal with  $\mu = 64.5$  inches and  $\sigma = 2.5$  inches. The standardized height is

$$z = \frac{\text{height} - 64.5}{2.5}$$

A woman's standardized height is the number of standard deviations by which her height differs from the mean height of all young women. A woman 68 inches tall, for example, has standardized height

$$z = \frac{68 - 64.5}{2.5} = 1.4$$

or 1.4 standard deviations above the mean. Similarly, a woman 5 feet (60 inches) tall has standardized height

$$z = \frac{60 - 64.5}{2.5} = -1.8$$

or 1.8 standard deviations less than the mean height.

If the variable we standardize has a normal distribution, standardizing does more than give a common scale. It makes all normal distributions into a

single distribution, and this distribution is still normal. Standardizing a variable that has any normal distribution produces a new variable that has the *standard normal distribution*.

### STANDARD NORMAL DISTRIBUTION

The **standard normal distribution** is the normal distribution  $N(0, 1)$  with mean 0 and standard deviation 1 (Figure 2.14).

If a variable  $x$  has any normal distribution  $N(\mu, \sigma)$  with mean  $\mu$  and standard deviation  $\sigma$ , then the standardized variable

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

has the standard normal distribution.

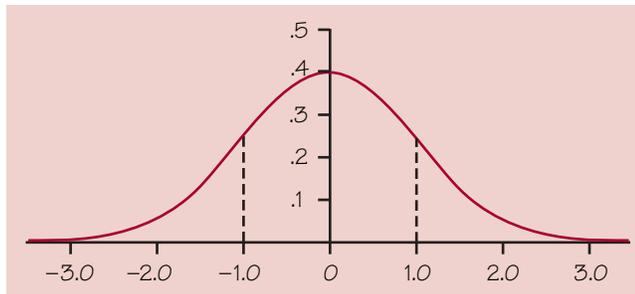


FIGURE 2.14 Standard normal distribution.

### EXERCISES

**2.19 SAT VERSUS ACT** Eleanor scores 680 on the mathematics part of the SAT. The distribution of SAT scores in a reference population is normal, with mean 500 and standard deviation 100. Gerald takes the American College Testing (ACT) mathematics test and scores 27. ACT scores are normally distributed with mean 18 and standard deviation 6. Find the standardized scores for both students. Assuming that both tests measure the same kind of ability, who has the higher score?

**2.20 COMPARING BATTING AVERAGES** Three landmarks of baseball achievement are Ty Cobb's batting average of .420 in 1911, Ted Williams's .406 in 1941, and George Brett's .390 in 1980. These batting averages cannot be compared directly because the distribution of major league batting averages has changed over the years. The distributions are quite symmetric and (except for outliers such as Cobb, Williams, and Brett) reasonably normal. While the mean batting average has been held roughly constant by rule changes and the balance between hitting and pitching, the standard deviation has dropped over time. Here are the facts:

Decade	Mean	Std. dev.
1910s	.266	.0371
1940s	.267	.0326
1970s	.261	.0317

Compute the standardized batting averages for Cobb, Williams, and Brett to compare how far each stood above his peers.<sup>2</sup>

### Normal distribution calculations

An area under a density curve is a proportion of the observations in a distribution. Any question about what proportion of observations lie in some range of values can be answered by finding an area under the curve. Because all normal distributions are the same when we standardize, we can find areas under any normal curve from a single table, a table that gives areas under the curve for the standard normal distribution.

#### THE STANDARD NORMAL TABLE

Table A is a table of areas under the standard normal curve. The table entry for each value  $z$  is the area under the curve to the left of  $z$ .

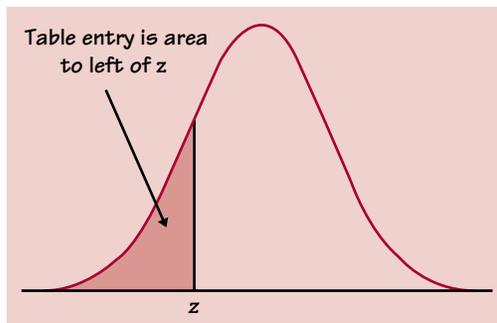
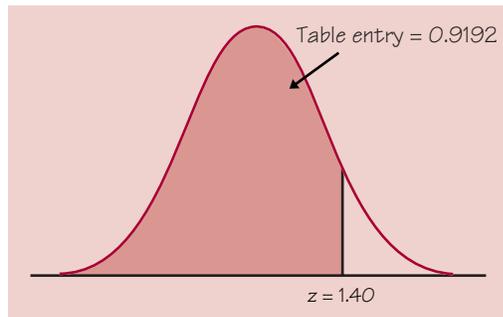


Table A, inside the front cover, gives areas under the standard normal curve. The next two examples show how to use the table.

#### EXAMPLE 2.5 USING THE $z$ TABLE

*Problem:* Find the proportion of observations from the standard normal distribution that are less than 1.4.

*Solution:* To find the area to the left of 1.40, locate 1.4 in the left-hand column of Table A, then locate the remaining digit 0 as .00 in the top row. The entry opposite 1.4 and under .00 is 0.9192. This is the area we seek. Figure 2.15 illustrates the relationship between the value  $z = 1.40$  and the area 0.9192.

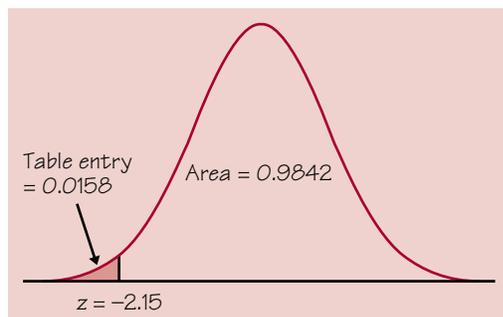


**FIGURE 2.15** The area under a standard normal curve to the left of the point  $z = 1.4$  is 0.9192. Table A gives areas under the standard normal curve.

### EXAMPLE 2.6 MORE ON USING THE $z$ TABLE

*Problem:* Find the proportion of observations from the standard normal distribution that are greater than  $-2.15$ .

*Solution:* Enter Table A under  $z = -2.15$ . That is, find  $-2.1$  in the left-hand column and  $.05$  in the top row. The table entry is 0.0158. This is the area to the *left* of  $-2.15$ . Because the total area under the curve is 1, the area lying to the *right* of  $-2.15$  is  $1 - 0.0158 = 0.9842$ . Figure 2.16 illustrates these areas.



**FIGURE 2.16** Areas under the standard normal curve to the right and left of  $z = -2.15$ . Table A gives only areas to the left.

**Caution!** A common student mistake is to look up a  $z$ -value in Table A and report the entry corresponding to that  $z$ -value, regardless of whether the problem asks for the area to the left or to the right of that  $z$ -value. Always sketch the standard normal curve, mark the  $z$ -value, and shade the area of interest. And before you finish, make sure your answer is reasonable in the context of the problem.

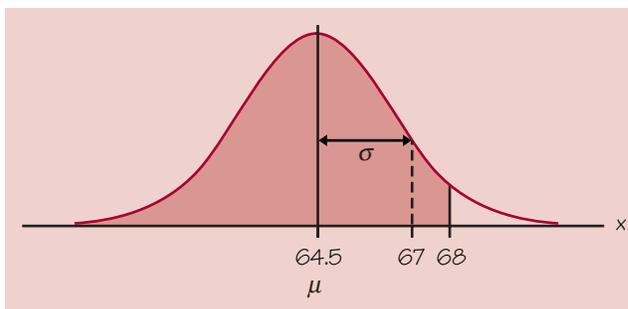
The value of the  $z$  table is that we can use it to answer any question about proportions of observations in a normal distribution by standardizing and then using the standard normal table.

### EXAMPLE 2.7 USING THE STANDARD NORMAL DISTRIBUTION

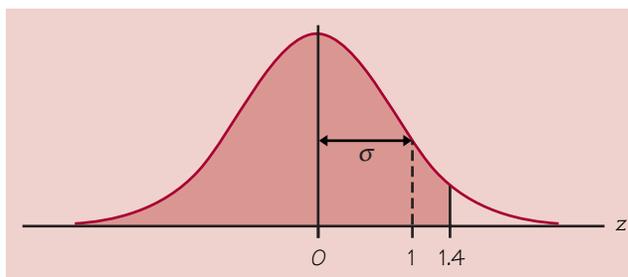
What proportion of all young women are less than 68 inches tall? This proportion is the area under the  $N(64.5, 2.5)$  curve to the left of the point 68. Figure 2.17(a) shows this area. The standardized height corresponding to 68 inches is

$$z = \frac{x - \mu}{\sigma} = \frac{68 - 64.5}{2.5} = 1.4$$

The area to the left of  $z = 1.4$  in Figure 2.17(b) under the standard normal curve is the same as the area to the left of  $x = 68$  in Figure 2.17(a).



**FIGURE 2.17(a)** The area under the  $N(64.5, 2.5)$  curve to the left of  $x = 68$ .



**FIGURE 2.17(b)** The area to the left of  $z = 1.4$  under the standard normal curve  $N(0,1)$ . This area is the same as the shaded area in Figure 2.17(a).

In Example 2.5, we found this area to be 0.9192. Our conclusion is that 91.92% of all young women are less than 68 inches tall.

Here is an outline of the method for finding the proportion of the distribution in any region.

#### FINDING NORMAL PROPORTIONS

**Step 1:** State the problem in terms of the observed variable  $x$ . Draw a picture of the distribution and shade the area of interest under the curve.

**Step 2:** Standardize  $x$  to restate the problem in terms of a standard normal variable  $z$ . Draw a picture to show the area of interest under the standard normal curve.

**Step 3:** Find the required area under the standard normal curve, using Table A and the fact that the total area under the curve is 1.

**Step 4:** Write your conclusion in the context of the problem.

#### EXAMPLE 2.8 IS CHOLESTEROL A PROBLEM FOR YOUNG BOYS?

The level of cholesterol in the blood is important because high cholesterol levels may increase the risk of heart disease. The distribution of blood cholesterol levels in a large population of people of the same age and sex is roughly normal. For 14-year-old boys, the mean is  $\mu = 170$  milligrams of cholesterol per deciliter of blood (mg/dl) and the standard deviation is  $\sigma = 30$  mg/dl.<sup>3</sup> Levels above 240 mg/dl may require medical attention. What percent of 14-year-old boys have more than 240 mg/dl of cholesterol?

**Step 1:** *State the problem.* Call the level of cholesterol in the blood  $x$ . The variable  $x$  has the  $N(170,30)$  distribution. We want the proportion of boys with cholesterol level  $x > 240$ . Sketch the distribution, mark the important points on the horizontal axis, and shade the area of interest. See Figure 2.18(a).

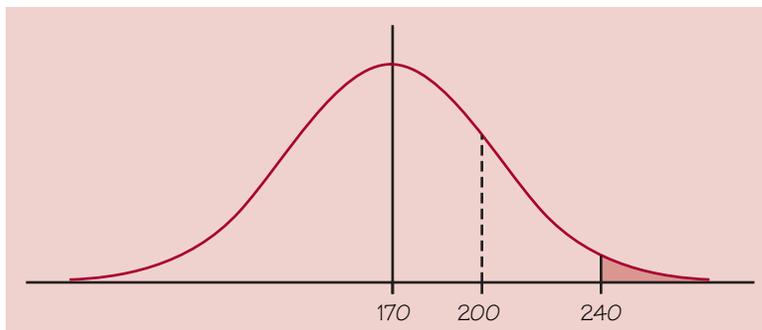
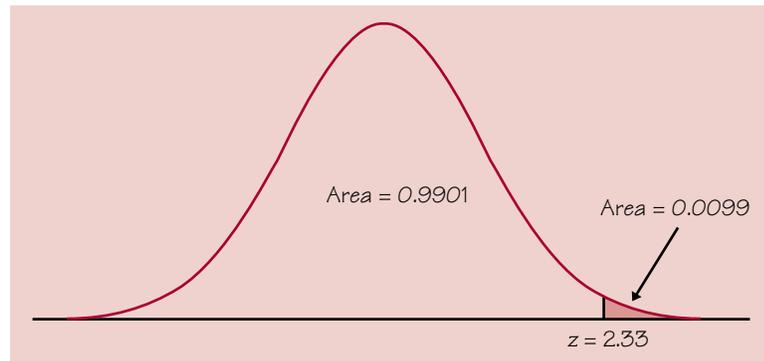


FIGURE 2.18(a) Cholesterol levels for 14-year-old boys who may require medical attention.



**FIGURE 2.18(b)** Areas under the standard normal curve.

**Step 2:** *Standardize  $x$  and draw a picture.* On both sides of the inequality, subtract the mean, then divide by the standard deviation, to turn  $x$  into a standard normal  $z$ :

$$\begin{aligned} x &> 240 \\ \frac{x-170}{30} &> \frac{240-170}{30} \\ z &> 2.33 \end{aligned}$$

Sketch a standard normal curve, and shade the area of interest. See Figure 2.18(b).

**Step 3:** *Use the table.* From Table A, we see that the proportion of observations less than 2.33 is 0.9901. About 99% of boys have cholesterol levels less than 240. The area to the right of 2.33 is therefore  $1 - 0.9901 = 0.0099$ . This is about 0.01, or 1%.

**Step 4:** *Write your conclusion in the context of the problem.* Only about 1% of boys have high cholesterol.

In a normal distribution, the proportion of observations with  $x > 240$  is the same as the proportion with  $x \geq 240$ . There is no area under the curve and exactly over 240, so the areas under the curve with  $x > 240$  and  $x \geq 240$  are the same. This isn't true of the actual data. There may be a boy with exactly 240 mg/dl of blood cholesterol. The normal distribution is just an easy-to-use approximation, not a description of every detail in the actual data.

The key to doing a normal calculation is to sketch the area you want, then match that area with the areas that the table gives you. Here is another example.

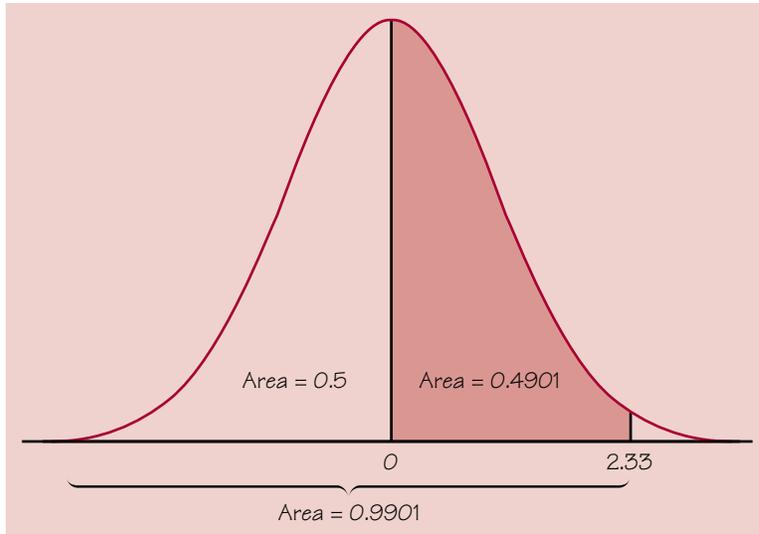
### EXAMPLE 2.9 WORKING WITH AN INTERVAL

What percent of 14-year-old boys have blood cholesterol between 170 and 240 mg/dl?

**Step 1:** *State the problem.* We want the proportion of boys with  $170 \leq x \leq 240$ .

**Step 2:** *Standardize and draw a picture.*

Sketch a standard normal curve, and shade the area of interest. See Figure 2.19.



**FIGURE 2.19** Areas under the standard normal curve.

$$\begin{aligned} 170 &\leq x \leq 240 \\ \frac{170-170}{30} &\leq \frac{x-170}{30} \leq \frac{240-170}{30} \\ 0 &\leq z \leq 2.33 \end{aligned}$$

**Step 3:** Use the table. The area between 2.33 and 0 is the area below 2.33 minus the area below 0. Look at Figure 2.19 to check this. From Table A,

$$\begin{aligned} \text{area between 0 and 2.33} &= \text{area below 2.33} - \text{area below 0.00} \\ &= 0.9901 - 0.5000 = 0.4901 \end{aligned}$$

**Step 4:** State your conclusion in context. About 49% of boys have cholesterol levels between 170 and 240 mg/dl.

What if we meet a  $z$  that falls outside the range covered by Table A? For example, the area to the left of  $z = -4$  does not appear in the table. But since  $-4$  is less than  $-3.4$ , this area is smaller than the entry for  $z = -3.40$ , which is 0.0003. There is very little area under the standard normal curve outside the range covered by Table A. You can take this area to be zero with little loss of accuracy.

### Finding a value given a proportion

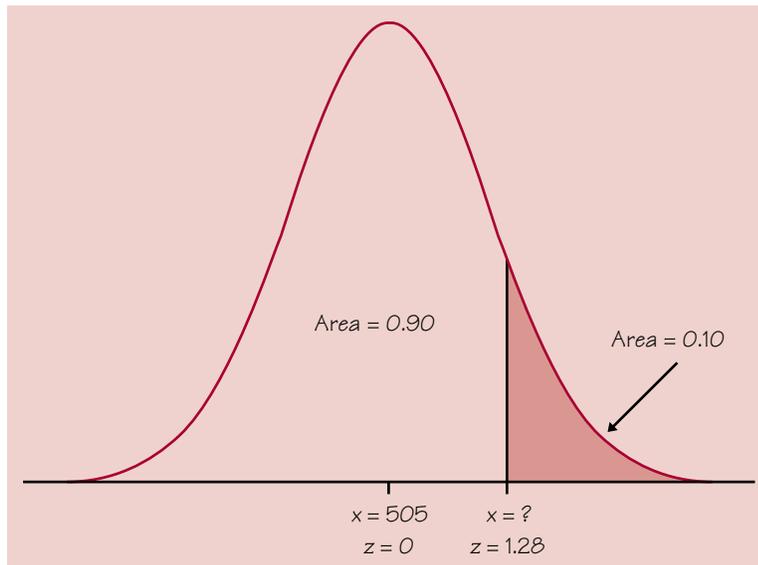
Examples 2.8 and 2.9 illustrate the use of Table A to find what proportion of the observations satisfies some condition, such as “blood cholesterol between 170 mg/dl and 240 mg/dl.” We may instead want to find the observed value with a given proportion of the observations above or below it. To do this, use

Table A backward. Find the given proportion in the body of the table, read the corresponding  $z$  from the left column and top row, then “unstandardize” to get the observed value. Here is an example.

### EXAMPLE 2.10 SAT VERBAL SCORES

Scores on the SAT Verbal test in recent years follow approximately the  $N(505, 110)$  distribution. How high must a student score in order to place in the top 10% of all students taking the SAT?

**Step 1:** *State the problem and draw a sketch.* We want to find the SAT score  $x$  with area 0.1 to its *right* under the normal curve with mean  $\mu = 505$  and standard deviation  $\sigma = 110$ . That’s the same as finding the SAT score  $x$  with area 0.9 to its *left*. Figure 2.20 poses the question in graphical form.



**FIGURE 2.20** Locating the point on a normal curve with area 0.10 to its right.

Because Table A gives the areas to the left of  $z$ -values, always state the problem in terms of the area to the left of  $x$ .

**Step 2:** *Use the table.* Look in the body of Table A for the entry closest to 0.9. It is 0.8997. This is the entry corresponding to  $z = 1.28$ . So  $z = 1.28$  is the standardized value with area 0.9 to its left.

**Step 3:** *Unstandardize* to transform the solution from the  $z$  back to the original  $x$  scale. We know that the standardized value of the unknown  $x$  is  $z = 1.28$ . So  $x$  itself satisfies

$$\frac{x - 505}{110} = 1.28$$

Solving this equation for  $x$  gives

$$x = 505 + (1.28)(110) = 645.8$$

This equation should make sense: it finds the  $x$  that lies 1.28 standard deviations above the mean on this particular normal curve. That is the “unstandardized” meaning of  $z = 1.28$ . We see that a student must score at least 646 to place in the highest 10%.

## EXERCISES

**2.21 TABLE A PRACTICE** Use Table A to find the proportion of observations from a standard normal distribution that satisfies each of the following statements. In each case, sketch a standard normal curve and shade the area under the curve that is the answer to the question.

- (a)  $z < 2.85$
- (b)  $z > 2.85$
- (c)  $z > -1.66$
- (d)  $-1.66 < z < 2.85$

**2.22 MORE TABLE A PRACTICE** Use Table A to find the value  $z$  of a standard normal variable that satisfies each of the following conditions. (Use the value of  $z$  from Table A that comes closest to satisfying the condition.) In each case, sketch a standard normal curve with your value of  $z$  marked on the axis.

- (a) The point  $z$  with 25% of the observations falling below it.
- (b) The point  $z$  with 40% of the observations falling above it.

**2.23 HEIGHTS OF AMERICAN MEN** The distribution of heights of adult American men is approximately normal with mean 69 inches and standard deviation 2.5 inches.

- (a) What percent of men are at least 6 feet (72 inches) tall?
- (b) What percent of men are between 5 feet (60 inches) and 6 feet tall?
- (c) How tall must a man be to be in the tallest 10% of all adult men?

**2.24 IQ TEST SCORES** Scores on the Wechsler Adult Intelligence Scale (a standard “IQ test”) for the 20 to 34 age group are approximately normally distributed with  $\mu = 110$  and  $\sigma = 25$ .

- (a) What percent of people age 20 to 34 have IQ scores above 100?
- (b) What percent have scores above 150?
- (c) How high an IQ score is needed to be in the highest 25%?

**2.25 HOW HARD DO LOCOMOTIVES PULL?** An important measure of the performance of a locomotive is its “adhesion,” which is the locomotive’s pulling force as a multiple of its weight. The adhesion of one 4400-horsepower diesel locomotive model varies in actual use according to a normal distribution with mean  $\mu = 0.37$  and standard deviation  $\sigma = 0.04$ .

- (a) What proportion of adhesions measured in use are higher than 0.40?

- (b) What proportion of adhesions are between 0.40 and 0.50?
- (c) Improvements in the locomotive’s computer controls change the distribution of adhesion to a normal distribution with mean  $\mu = 0.41$  and standard deviation  $\sigma = 0.02$ . Find the proportions in (a) and (b) after this improvement.

### Assessing normality

In the latter part of this course we will want to invoke various tests of significance to try to answer questions that are important to us. These tests involve sampling people or objects and inspecting them carefully to gain insights into the populations from which they come. Many of these procedures are based on the assumption that the host population is approximately normally distributed. Consequently, we need to develop methods for assessing normality.

**Method 1** Construct a frequency histogram or a stemplot. See if the graph is approximately bell-shaped and symmetric about the mean.

A histogram or stemplot can reveal distinctly nonnormal features of a distribution, such as outliers, pronounced skewness, or gaps and clusters. You can improve the effectiveness of these plots for assessing whether a distribution is normal by marking the points  $\bar{x}$ ,  $\bar{x} \pm s$ , and  $\bar{x} \pm 2s$  on the  $x$  axis. This gives the scale natural to normal distributions. Then compare the count of observations in each interval with the 68–95–99.7 rule.

#### EXAMPLE 2.11 ASSESSING NORMALITY OF THE GARY VOCABULARY SCORES

The histogram in Figure 2.2 (page 78) suggests that the distribution of the 947 Gary vocabulary scores is close to normal. It is hard to assess by eye how close to normal a histogram is. Let’s use the 68–95–99.7 rule to check more closely. We enter the scores into a statistical computing system and ask for the mean and standard deviation. The computer replies,

MEAN = 6.8585  
STDEV = 1.5952

Now that we know that  $\bar{x} = 6.8585$  and  $s = 1.5952$ , we check the 68–95–99.7 rule by finding the actual counts of Gary vocabulary scores in intervals of length  $s$  about the mean  $\bar{x}$ . The computer will also do this for us. Here are the counts:

1	21	129	331	318	125	21	1
2.07	3.67	5.26	6.86	8.45	10.05	11.64	
$\bar{x} - 3s$	$\bar{x} - 2s$	$\bar{x} - s$	$\bar{x}$	$\bar{x} + s$	$\bar{x} + 2s$	$\bar{x} + 3s$	

The distribution is very close to symmetric. It also follows the 68–95–99.7 rule closely: there are 68.5% of the scores (649 out of 947) within one standard deviation of the mean, 95.4% (903 of 947) within two standard deviations, and 99.8% (945 of the 947 scores) within three. These counts confirm that the normal distribution with  $\mu = 6.86$  and  $\sigma = 1.595$  fits these data well.

Smaller data sets rarely fit the 68–95–99.7 rule as well as the Gary vocabulary scores. This is true even of observations taken from a larger population that really has a normal distribution. There is more chance variation in small data sets.

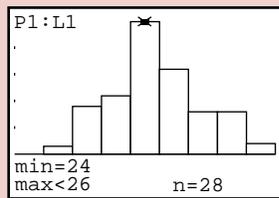
**Method 2** Construct a *normal probability plot*. A normal probability plot provides a good assessment of the adequacy of the normal model for a set of data. Most statistics utilities, including Minitab and Data Desk, can construct normal probability plots from entered data. The TI-83/89 will also do normal probability plots. You will need to be able to produce a normal probability plot (either with a calculator or with computer software) and interpret it. We will do this part first, and then we will describe the steps the calculator goes through to produce the plot.

*normal probability plot*

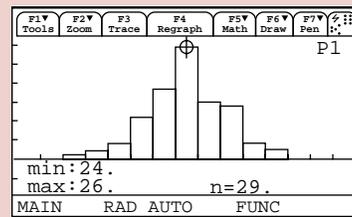
**TECHNOLOGY TOOLBOX** *Normal probability plots on the TI-83/89*

If you ran the program FLIP50 in Exercise 2.17, and you still have the data (100 numbers mostly in the 20s) in  $L_1$ /list1, then use these data. If you have not entered the program and run it, take a few minutes to do that now. Duplicate this example with *your* data. Here is the histogram that was generated at the end of one run of this simulation on each calculator.

TI-83



TI-89

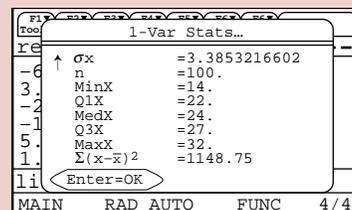
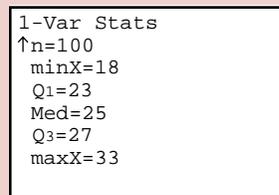
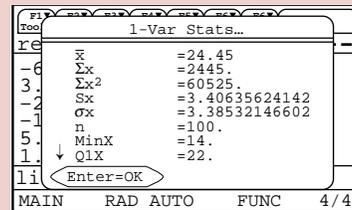
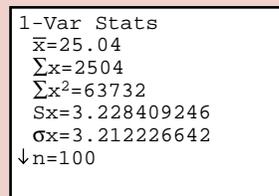


Ask for one-variable statistics:

- Press **[STAT]**, choose CALC, then 1:1-Var Stats and **[2nd]** **[1]** ( $L_1$ )

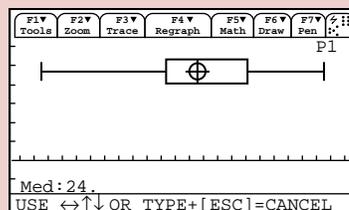
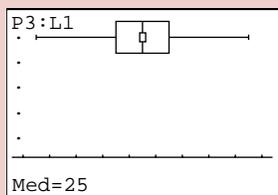
- In the Statistics/List Editor, press **[F4]** (Calc) and choose 1:1-Var Stats for list1.

This will give us the following:

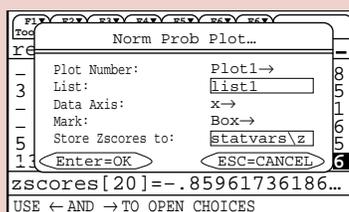
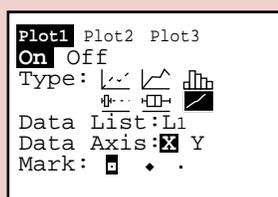


**TECHNOLOGY TOOLBOX** Normal probability plots (continued)

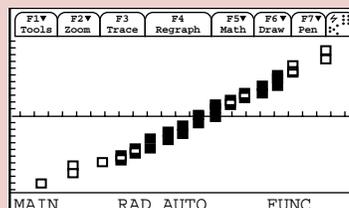
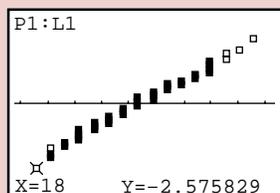
- Comparing the means and medians ( $\bar{x} = 25.04$  vs.  $M = 25$  on the TI-83 and  $\bar{x} = 24.45$  vs.  $M = 24$  on the TI-89) suggests that the distributions are fairly symmetric. Boxplots confirm the roughly symmetric shape.



- To construct a normal probability plot of the data, define Plot 1 like this:



- Use ZoomStat (ZoomData on the TI-89) to see the finished graph.

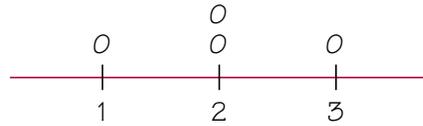


*Interpretation:* If the data distribution is close to a normal distribution, the plotted points will lie close to a straight line. Conversely, nonnormal data will show a nonlinear trend. Outliers appear as points that are far away from the overall pattern of the plot. Since the above plot is quite linear, our conclusion is that it is reasonable to believe that the data are from a normal distribution.

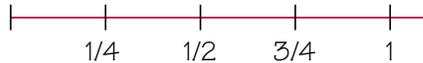
The next example uses a very simple data set to illustrate how a normal probability plot is constructed.

**EXAMPLE 2.12** HOW NORMAL PROBABILITY PLOTS ARE CONSTRUCTED

To show how the calculator constructs a normal probability plot, let's look at a very simple data set: {1,2,2,3}. Here,  $n = 4$  and a dotplot shows that the distribution is perfectly symmetric if not exactly bell-shaped.



**Step 1:** Order the observations from smallest to largest. In this case, the points are already ordered. Since  $n = 4$ , divide the interval  $[0,1]$  on the horizontal axis into four subintervals.

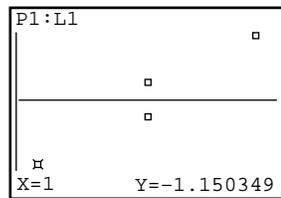


Mark the midpoint of each subinterval:  $1/8$ ,  $3/8$ ,  $5/8$ , and  $7/8$ . In the general case, we would mark the points corresponding to  $1/2n$ ,  $3/2n$ ,  $5/2n$ ,  $\dots$ ,  $(2n - 1)/2n$ .

**Step 2:** For the first midpoint,  $1/8$ , find the  $z$ -value that has area  $1/8 = 0.125$  lying to the left of it. The closest value in the body of the table is 0.1251, and the corresponding  $z$ -value is  $-1.15$ . Do the same for the other midpoints. Here is a table of our results:

$x$	Midpoint	$y$
1	$1/8 = 0.1250$	$-1.15$
2	$3/8 = 0.3750$	$-0.319$
3	$5/8 = 0.6250$	$0.319$
4	$7/8 = 0.8750$	$1.15$

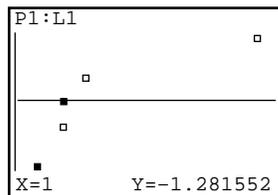
**Step 3:** Plot the points  $(x,y)$ . This is the normal probability plot for our simple data set.



If an outlier were added, say 10, then the table would look like this:

$x$	Midpoint	$y$
1	0.1	$-1.28$
2	0.3	$-0.52$
2	0.5	0
3	0.7	0.52
10	0.9	1.28

and the normal probability plot becomes



This last picture shows a normal probability plot for a data set that is clearly not approximately normally distributed.

Any normal distribution produces a straight line on the plot because standardizing is a transformation that can change the slope and intercept of the line in our plot but cannot change a line into a curved pattern.

### EXERCISES

**2.26 CAVENDISH AND THE DENSITY OF THE EARTH** Repeated careful measurements of the same physical quantity often have a distribution that is close to normal. Here are Henry Cavendish’s 29 measurements of the density of the earth, made in 1798. (The data give the density of the earth as a multiple of the density of water.)

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

- (a) Construct a stemplot to show that the data are reasonably symmetric.
- (b) Now check how closely they follow the 68–95–99.7 rule. Find  $\bar{x}$  and  $s$ , then count the number of observations that fall between  $\bar{x} - s$  and  $\bar{x} + s$ , between  $\bar{x} - 2s$  and  $\bar{x} + 2s$ , and between  $\bar{x} - 3s$  and  $\bar{x} + 3s$ . Compare the percents of the 29 observations in each of these intervals with the 68–95–99.7 rule.
- (c) Use your calculator to construct a normal probability plot for Cavendish’s density of the earth data, and write a brief statement about the normality of the data. Does the normal probability plot reinforce your findings in (a)?

We expect that when we have only a few observations from a normal distribution, the percents will show some deviation from 68, 95, and 99.7. Cavendish’s measurements are in fact close to normal.

**2.27 GREAT WHITE SHARKS** Here are the lengths in feet of 44 great white sharks:

18.7	12.3	18.6	16.4	15.7	18.3	14.6	15.8	14.9	17.6	12.1
16.4	16.7	17.8	16.2	12.6	17.8	13.8	12.2	15.2	14.7	12.4
13.2	15.8	14.3	16.6	9.4	18.2	13.2	13.6	15.3	16.1	13.5
19.1	16.2	22.8	16.8	13.6	13.2	15.7	19.7	18.7	13.2	16.8

- (a) Use the methods of Chapter 1 to describe the distribution of these lengths.
- (b) Compare the mean with the median. Does this comparison support your assessment of the shape of the distribution in (a)? Explain.

- (c) Is the distribution approximately normal? If you haven't done this already, enter the data into your calculator, and reorder them from smallest to largest. Then calculate the percent of the data that lies within one standard deviation of the mean. Within two standard deviations of the mean. Within three standard deviations of the mean.
- (d) Use your calculator to construct a normal probability plot. Interpret this plot.
- (e) Having inspected the data from several different perspectives, do you think these data are approximately normal? Write a brief summary of your assessment that combines your findings from (a) to (d).

## SUMMARY

To **standardize** any observation  $x$ , subtract the mean of the distribution and then divide by the standard deviation. The resulting  **$z$ -score**

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

says how many standard deviations  $x$  lies from the distribution mean.

All normal distributions are the same when measurements are transformed to the standardized scale. If  $x$  has the  $N(\mu, \sigma)$  distribution, then the **standardized variable**  $z = (x - \mu)/\sigma$  has the **standard normal distribution**  $N(0, 1)$  with mean 0 and standard deviation 1. Table A gives the proportions of standard normal observations that are less than  $z$  for many values of  $z$ . By standardizing, we can use Table A for any normal distribution.

In order to perform certain inference procedures in later chapters, we will need to know that the data come from populations that are approximately normally distributed. To assess normality, one can observe the shape of histograms, stemplots, and boxplots and see how well the data fit the 68–95–99.7 rule for normal distributions. Another good method for assessing normality is to construct a **normal probability plot**.

## SECTION 2.2 EXERCISES

**2.28 TABLE A PRACTICE** Use Table A to find the proportion of observations from a standard normal distribution that falls in each of the following regions. In each case, sketch a standard normal curve and shade the area representing the region.

- (a)  $z \leq -2.25$   
 (b)  $z \geq -2.25$   
 (c)  $z > 1.77$   
 (d)  $-2.25 < z < 1.77$

**2.29 MORE TABLE A PRACTICE** Use Table A to find the value  $z$  of a standard normal variable that satisfies each of the following conditions. (Use the value of  $z$  from Table A

that comes closest to satisfying the condition.) In each case, sketch a standard normal curve with your value of  $z$  marked on the axis.

- (a) The point  $z$  with 70% of the observations falling below it.
- (b) The point  $z$  with 85% of the observations falling above it.
- (c) Find the number  $z$  such that the proportion of observations that are less than  $z$  is 0.8.
- (d) Find the number  $z$  such that 90% of all observations are greater than  $z$ .

**2.30 THE STOCK MARKET** The annual rate of return on stock indexes (which combine many individual stocks) is approximately normal. Since 1945, the Standard & Poor's 500 Index has had a mean yearly return of 12%, with a standard deviation of 16.5%. Take this normal distribution to be the distribution of yearly returns over a long period.

- (a) In what range do the middle 95% of all yearly returns lie?
- (b) The market is down for the year if the return on the index is less than zero. In what proportion of years is the market down?
- (c) In what proportion of years does the index gain 25% or more?

**2.31 GESTATION PERIOD** The length of human pregnancies from conception to birth varies according to a distribution that is approximately normal with mean 266 days and standard deviation 16 days.

- (a) What percent of pregnancies last less than 240 days (that's about 8 months)?
- (b) What percent of pregnancies last between 240 and 270 days (roughly between 8 months and 9 months)?
- (c) How long do the longest 20% of pregnancies last?

**2.32 ARE WE GETTING SMARTER?** When the Stanford-Binet "IQ test" came into use in 1932, it was adjusted so that scores for each age group of children followed roughly the normal distribution with mean  $\mu = 100$  and standard deviation  $\sigma = 15$ . The test is readjusted from time to time to keep the mean at 100. If present-day American children took the 1932 Stanford-Binet test, their mean score would be about 120. The reasons for the increase in IQ over time are not known but probably include better childhood nutrition and more experience in taking tests.<sup>4</sup>

- (a) IQ scores above 130 are often called "very superior." What percent of children had very superior scores in 1932?
- (b) If present-day children took the 1932 test, what percent would have very superior scores? (Assume that the standard deviation  $\sigma = 15$  does not change.)

**2.33 QUANTILES** The quartiles of any density curve are the points with area 0.25 and 0.75 to their left under the curve.

- (a) What are the quartiles of a standard normal distribution?
- (b) How many standard deviations away from the mean do the quartiles lie in any normal distribution? What are the quartiles for the lengths of human pregnancies? (Use the distribution in Exercise 2.31.)

**2.34 DECILES** The *deciles* of any distribution are the points that mark off the lowest 10% and the highest 10%. The deciles of a density curve are therefore the points with area 0.1 and 0.9 to their left under the curve.

- (a) What are the deciles of the standard normal distribution?
- (b) The heights of young women are approximately normal with mean 64.5 inches and standard deviation 2.5 inches. What are the deciles of this distribution?

**2.35 LACTIC ACID IN CHEESE** The taste of mature cheese is related to the concentration of lactic acid in the cheese. The concentrations of lactic acid in 30 samples of cheddar cheese are given in the Technology Toolbox on page 15.

- (a) Enter the data into your calculator. Make a histogram and overlay a boxplot. Sketch the results on your paper. Compare the mean with the median. Describe the distribution of these data in a sentence.
- (b) Calculate the percent of the data that lies within one, two, and three standard deviations of the mean.
- (c) Use your calculator to construct a normal probability plot. Sketch this plot on your paper.
- (d) Having inspected the data from several different perspectives, do you think these data are approximately normal? Write a brief statement of your assessment that combines your findings from (a) to (c).

**2.36 ARE THE PRESIDENTS' AGES NORMAL?** The histogram for the ages of the 43 presidents was very symmetric (see Figure 1.7, page 20). Use the list that we named PREZ to construct a normal probability plot for this data set, and confirm the linear trend. Write a statement about your assessment of normality of the presidents' ages.

**2.37 STANDARDIZED VALUES BY CALCULATOR** This exercise uses the TI-83/89 to calculate standardized values for a familiar data set and then calculates the mean and standard deviation for these transformed values. Without knowing the data set, can you guess the mean and standard deviation?

Set up your Statistics/List Editor so that the list PREZ (the presidents' ages from Exercise 2.36) is the first list:

- TI-83: Press  $\boxed{\text{STAT}}$ , choose 5:SETUPEDITOR, then press  $\boxed{2\text{nd}}\boxed{\text{STAT}}$  (LIST), choose PREZ, and press  $\boxed{\text{ENTER}}$ .
- TI-89: Press  $\boxed{\text{CATALOG}}\boxed{\text{F3}}$  (Flash Apps), choose setupEd(, then type prez) and press  $\boxed{\text{ENTER}}$ .

In the Statistics/List Editor, move your cursor to the header of the next (blank) list and name it STDSC (for standardized scores). With the name of this list highlighted, define the list by carefully entering  $(\text{PREZ}-\text{mean}(\text{PREZ}))/\text{stdDev}(\text{PREZ})$ . The mean and stdDev commands are found under the LIST/MATH menu.

Scroll through the list STDSC to verify that the values range from about  $-3$  to  $3$ . Then construct a histogram of STDSC, and calculate one-variable statistics for STDSC. What are the mean and standard deviation?

## CHAPTER REVIEW

Here is a review list of the most important skills you should have acquired from your study of this chapter.

### A. DENSITY CURVES

1. Know that areas under a density curve represent proportions of all observations and that the total area under a density curve is 1.
2. Approximately locate the median (equal-areas point) and the mean (balance point) on a density curve.
3. Know that the mean and median both lie at the center of a symmetric density curve and that the mean moves farther toward the long tail of a skewed curve.

### B. NORMAL DISTRIBUTIONS

1. Recognize the shape of normal curves and be able to estimate both the mean and standard deviation from such a curve.
2. Use the 68–95–99.7 rule and symmetry to state what percent of the observations from a normal distribution fall between two points when both points lie at the mean or one, two, or three standard deviations on either side of the mean.
3. Find the standardized value ( $z$ -score) of an observation. Interpret  $z$ -scores and understand that any normal distribution becomes standard normal  $N(0, 1)$  when standardized.
4. Given that a variable has the normal distribution with a stated mean  $\mu$  and standard deviation  $\sigma$ , use Table A and your calculator to calculate the proportion of values above a stated number, below a stated number, or between two stated numbers.
5. Given that a variable has the normal distribution with a stated mean  $\mu$  and standard deviation  $\sigma$ , calculate the point having a stated proportion of all values above it. Also calculate the point having a stated proportion of all values below it.

### C. ASSESSING NORMALITY

1. Plot a histogram, stemplot, and/or boxplot to determine if a distribution is bell-shaped.
2. Determine the proportion of observations within one, two, and three standard deviations of the mean, and compare with the 68–95–99.7 rule for normal distributions.
3. Construct and interpret normal probability plots.

## CHAPTER 2 REVIEW EXERCISES

**2.38** A certain density curve consists of a straight-line segment that begins at the origin,  $(0, 0)$ , and has slope 1.

- Sketch the density curve. What are the coordinates of the right endpoint of the segment? (*Note:* The right endpoint should be fixed so that the total area under the curve is 1. This is required for a valid density curve.)
- Determine the median, the first quartile ( $Q_1$ ), and the third quartile ( $Q_3$ ).
- Relative to the median, where would you expect the mean of the distribution?
- What percent of the observations lie below 0.5? Above 1.5?

**2.39** A certain density curve looks like an inverted letter “V.” The first segment goes from the point  $(0, 0.6)$  to the point  $(0.5, 1.4)$ . The second segment goes from  $(0.5, 1.4)$  to  $(1, 0.6)$ .

- Sketch the curve. Verify that the area under the curve is 1, so that it is a valid density curve.
- Determine the median. Mark the median and the approximate locations of the quartiles  $Q_1$  and  $Q_3$  on your sketch.
- What percent of the observations lie below 0.3?
- What percent of the observations lie between 0.3 and 0.7?

**2.40 STANDARDIZED TEST SCORES AS PERCENTILES** Joey received a report that he scored in the 97th percentile on a national standardized reading test but in the 72nd percentile on the math portion of the test. Explain to Joey’s grandmother, who knows no statistics, what these numbers mean.

**2.41 TABLE A PRACTICE** Use Table A to find the proportion of observations from a standard normal distribution that falls in each of the following regions. In each case, sketch a standard normal curve and shade the area representing the region.

- $z < 1.28$
- $z > -0.42$
- $-0.42 < z < 1.28$
- $z < 0.42$

### 2.42 WORKING BACKWARD, FINDING $z$ -VALUES

- Find the number  $z$  such that the proportion of observations that are less than  $z$  in a standard normal distribution is 0.98.
- Find the number  $z$  such that 22% of all observations from a standard normal distribution are greater than  $z$ .

**2.43 QUANTILES FROM A NORMAL DISTRIBUTION** Find the quartiles for the distribution of blood cholesterol levels for 14-year-old boys (see Example 2.8, page 99). This distribution is  $N(170 \text{ mg/dl}, 30 \text{ mg/dl})$ .

**2.44 ARE YOU A GOOD JUDGE OF PEOPLE?** The Chapin Social Insight Test evaluates how accurately the subject appraises other people. In the reference population used to develop the test, scores are approximately normally distributed with mean 25 and standard deviation 5. The range of possible scores is 0 to 41.

- (a) What proportion of the population has scores below 20 on the Chapin test?
- (b) What proportion has scores below 10?
- (c) What proportion has scores above 35?
- (d) How high a score must you have in order to be in the top quarter of the population in social insight?

**2.45 IQ SCORES FOR CHILDREN** The scores of a reference population on the Wechsler Intelligence Scale for Children (WISC) are normally distributed with  $\mu = 100$  and  $\sigma = 15$ . A school district classified children as “gifted” if their WISC score exceeds 135. There are 1300 sixth-graders in the school district. About how many of them are gifted?

**2.46 CULTURE SHOCK** The Acculturation Rating Scale for Mexican Americans (ARSMA) is a psychological test that measures the degree to which Mexican Americans are adapted to Mexican/Spanish versus Anglo/English culture. The range of possible scores is 1.0 to 5.0, with higher scores showing more Anglo/English acculturation. The distribution of ARSMA scores in a population used to develop the test is approximately normal with mean 3.0 and standard deviation 0.8. A researcher believes that Mexicans will have an average score near 1.7 and that first-generation Mexican Americans will average about 2.1 on the ARSMA scale. What proportion of the population used to develop the test has scores below 1.7? Between 1.7 and 2.1?

**2.47 HELMET SIZES** The army reports that the distribution of head circumference among soldiers is approximately normal with mean 22.8 inches and standard deviation 1.1 inches. Helmets are mass-produced for all except the smallest 5% and the largest 5% of head sizes. Soldiers in the smallest or largest 5% get custom-made helmets. What head sizes get custom-made helmets?

**2.48 ADAPTING CULTURALLY** The ARSMA test is described in Exercise 2.46. How high a score on this test must a Mexican American obtain to be among the 30% of the population used to develop the test who are most Anglo/English in cultural orientation? What scores make up the 30% who are most Mexican/Spanish in their acculturation?

**2.49 PROFESSOR MOORE'S DRIVING TIMES** Exercise 1.29 (page 36) shows driving times between home and college for Professor Moore.

- (a) Make a histogram of these drive times. Is the distribution roughly symmetric, clearly skewed, or neither? Are there any clear outliers?
- (b) The data show three unusual situations: the day after Thanksgiving (no traffic on campus); a delay due to an accident; and a day with icy roads. Identify and remove these three observations. Are the remaining observations reasonably close to having a normal distribution? Write a short statement that describes your analyses and your conclusions.

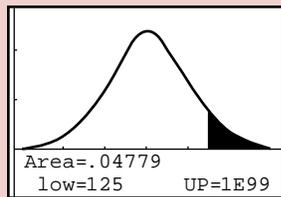
**2.50 CORN-FED CHICKS** Exercise 1.57 (page 63) presents data on the weight gains of chicks fed two types of corn. The researchers use  $\bar{x}$  and  $s$  to summarize each of the two distributions. Make a normal probability plot for each group and report your findings. Is the use of  $\bar{x}$  and  $s$  justified?

### TECHNOLOGY TOOLBOX Finding areas with ShadeNorm

The TI-83/89 can be used to find the area to the left or right of a point or above an interval without referring to a standard normal table. Consider the WISC scores for children of Exercise 2.45. This distribution is  $N(100, 15)$ . Suppose we want to find the percent of children whose WISC scores are above 125. Begin by specifying a viewing window as follows:  $X[55, 145]_{15}$  and  $Y[-0.008, 0.028]_{.01}$ . You will generally need to experiment with the  $y$  settings to get a good graph.

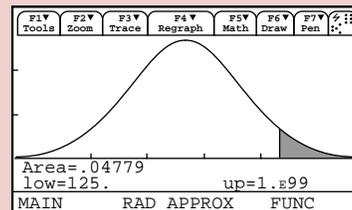
#### TI-83

- Press  $\boxed{2\text{nd}} \boxed{\text{VAR}} \boxed{\text{(DISTR)}}$ , then choose DRAW and 1: ShadeNorm(.).
- Complete the command ShadeNorm(125, 1E99, 100, 15) and press  $\boxed{\text{ENTER}}$ .



#### TI-89

- Press  $\boxed{\text{CATALOG}} \boxed{\text{F3}}$  (Flash Apps) and choose shadNorm(.).
- Complete the command tistat.shadNorm(125, 1E99, 100, 15) and press  $\boxed{\text{ENTER}}$ .



You must always specify an interval. An area in the right tail of the distribution would theoretically be the interval  $(125, \infty)$ . The calculator limitation dictates that we use a *number* that is at least 5 or 10 standard deviations to the right of the mean. To find the area to the left of 85, you would specify ShadeNorm( $-1E99, 85, 100, 15$ ). Or we could specify ShadeNorm( $0, 85, 100, 15$ ) since WISC scores can't be negative. Both yield at least four-decimal-place accuracy. If you're using standard normal values, then you need only specify the endpoints of the interval; the mean 0 and standard deviation 1 will be understood. For example, use ShadeNorm(1, 2) to find the area above the interval  $z = 1$  to  $z = 2$ .

**2.51 MADE IN THE SHADE** Use the calculator's ShadeNorm feature to find the following areas correct to four-decimal-place accuracy. Then write your findings in a sentence.

- The relative frequency of scores greater than 110.
- The relative frequency of scores lower than 85.
- Show two ways to find the relative frequency of scores within two standard deviations of the mean.

**TECHNOLOGY TOOLBOX** *Finding areas with normalcdf*

The **normalcdf** command on the TI-83/89 can be used to find the area under a normal distribution and above an interval. This method has the advantage over ShadeNorm of being quicker to do, and the disadvantage of not providing a picture of the area it is finding. Here are the keystrokes for the WISC scores of Exercise 2.45:

**TI-83**

- Press  $\boxed{2\text{nd}} \boxed{\text{VARS}}$  (DISTR) and choose 2: normalcdf (.
- Complete the command `normalcdf(125, 1E99, 100, 15)` and press  $\boxed{\text{ENTER}}$ .

```
normalcdf(125,1E99,
100,15)
.0477903304
```

**TI-89**

- Press  $\boxed{\text{CATALOG}} \boxed{\text{F3}}$  (Flash Apps) and choose normCdf (.
- Complete the command `tistat.normCdf(125, 1E99, 100, 15)` and press  $\boxed{\text{ENTER}}$ .

```
tistat.normCdf(125,1.E99)
.04779033036
normCdf(125,1E99,100,15)
MAIN RAD APPROX FUNC 1/30
```

We can say that about 5% of the WISC scores are above 125. If the normal values have already been standardized, then you need only specify the left and right endpoints of the interval. For example, `normalcdf(-1, 1)` returns 0.6827, meaning that the area from  $z = -1$  to  $z = 1$  is approximately 0.6827, correct to four decimal places.

**2.52 AREAS BY CALCULATOR** Use the calculator's **normalcdf** function to verify your answers to Exercises 2.41 (page 113), and 2.46 (page 114).

**2.53 IQ SCORES FOR ADULTS** Wechsler Adult Intelligence Scale (WAIS) scores for young adults are  $N(110, 25)$ . Use your calculator to show that the area under the entire curve is equal to 1. Note that you can't specify the interval  $(-\infty, +\infty)$ , so you'll have to decide on some endpoints that are far enough from the center (110) of the distribution to give at least four-decimal-place accuracy. Record the interval that you use and the area that the calculator reports. Will it suffice to go out four standard deviations on either side of the center? Five standard deviations?

**2.54** Use the calculator's **invNorm** function to verify your answers to Exercises 2.42 (page 113) and 2.47 (page 114). Use the method described in the Technology Toolbox on page 117.

### TECHNOLOGY TOOLBOX Finding $z$ -values with $\text{invNorm}$

The TI-83/89  $\text{invNorm}$  function calculates the raw or standardized normal value corresponding to a known area under a normal distribution or a relative frequency. The following example uses the WISC scores, which have a  $N(100, 15)$  distribution. Here are the keystrokes:

#### TI-83

- Press  $\text{2nd}$   $\text{VAR}$  (DISTR), then choose 3:  $\text{invNorm}$ (.
- Complete the command  $\text{invNorm}(.9, 100, 15)$  and press  $\text{ENTER}$ . Compare this with the command  $\text{invNorm}(.9)$ .

```
invNorm(.9,100,15)
      119.2232735
invNorm(.9)
      1.281551567
```

#### TI-89

- Press  $\text{CATALOG}$   $\text{F3}$  (Flash Apps) and choose  $\text{invNorm}$ (.
- Complete the command  $\text{tistat.invNorm}(.9, 100, 15)$  and press  $\text{ENTER}$ . Compare this with the command  $\text{invNorm}(.9)$ .

F1▼	F2▼	F3▼	F4▼	F5▼	F6▼
Tools	Algebra	Calc	Other	PrgmIO	Clean Up
<ul style="list-style-type: none"> <li>■ <math>\text{tistat.invnorm}(.9,100,15)</math> 119.223</li> <li>■ <math>\text{tistat.invnorm}(.9)</math> 1.28155</li> </ul>					
TISat.invNorm(.9)					
MAIN RAD APPROX FUNC 2/30					

The first command finds that the *raw* WISC score that has 90% of the scores below it from the  $N(100, 15)$  distribution is  $x = 119$ . The second command says that the *standardized* WISC score that has 90% of the scores below it is  $z = 1.28$ .

### NOTES AND DATA SOURCES

1. Data from Gary Community School Corporation, courtesy of Celeste Foster, Department of Education, Purdue University.
2. Data from Stephen Jay Gould, "Entropic homogeneity isn't why no one hits 400 anymore," *Discover*, August 1986, pp. 60–66. Gould does not standardize but gives a speculative discussion instead.
3. Detailed data appear in P. S. Levy et al., "Total serum cholesterol values for youths 12–17 years," *Vital and Health Statistics Series 11*, No. 150 (1975), U.S. National Center for Health Statistics.
4. Ulric Neisser, "Rising scores on intelligence tests," *American Scientist*, September–October 1997, online edition.