Chapter 6: Probability

Introduction to Probability

- The role of inferential statistics is to use the sample data as the basis for answering questions about the population.
- To accomplish this goal, inferential procedures are typically built around the concept of probability.
- Specifically, the relationships between samples and populations are usually defined in terms of probability.

Probability Definition

- <u>Definition</u>: For a situation in which several different outcomes are possible, the probability for any specific outcome is defined as a fraction or a proportion of all the possible outcomes.
  - In other words, if the possible outcomes are identified as A, B, C, D, and so on, then:

$$\text{probability of } A = \frac{\text{number of outcomes classified as } A}{\text{total number of possible outcomes}}$$

- To simplify the discussion of probability, we use a notation system that eliminates a lot of the words.

Probability Definition cont.

- The probability of a specific outcome is expressed with a *p* (for probability) followed by the specific outcome in parentheses.
  - For example, the probability of selecting a king from a deck of cards is written as p(king).
  - The probability of obtaining heads for a coin toss is written as p(heads).
- <u>Note that probability is defined as a proportion, or a part of the whole.</u>
  - In other words, this definition makes it possible to restate any probability problem as a proportion problem.

Probability Definition cont.

- By convention, probability values most often are expressed as decimal values.
- But you should realize that any of these three forms is acceptable.
  - Fraction
  - Proportion
  - Percentage

Random Sampling

- For the preceding definition of probability to be accurate, it is necessary that the outcomes be obtained by a process called *random sampling.*
- Definition: A *random sample* requires that each individual in the population has an equal chance of being selected.
  - A second requirement, necessary for many statistical formulas, states that the probabilities must stay constant from one selection to the next if more than one individual is selected.
  - To keep the probabilities from changing from one selection to the next, it is necessary to return each individual to the population before you make the next selection.

Random Sampling cont.

- This process is called *sampling with replacement.*
  - The second requirement for random samples (constant probability) demands that you sample with replacement.

Probability and the Normal Distribution

- An example of a normal distribution is shown in Figure 6.3.
- <u>Note that the normal distribution is symmetrical, with the highest frequency in the middle and frequencies tapering off as you move toward either extreme.</u>
- Although the exact shape for the normal distribution is defined by an equation (see Figure 6.3), the normal shape can also be described by the proportions of area contained in each section of the distribution.
- Statisticians often identify sections of a normal distribution by using *z*-scores.

$$Y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(X-\mu)^2/2\sigma^2}$$

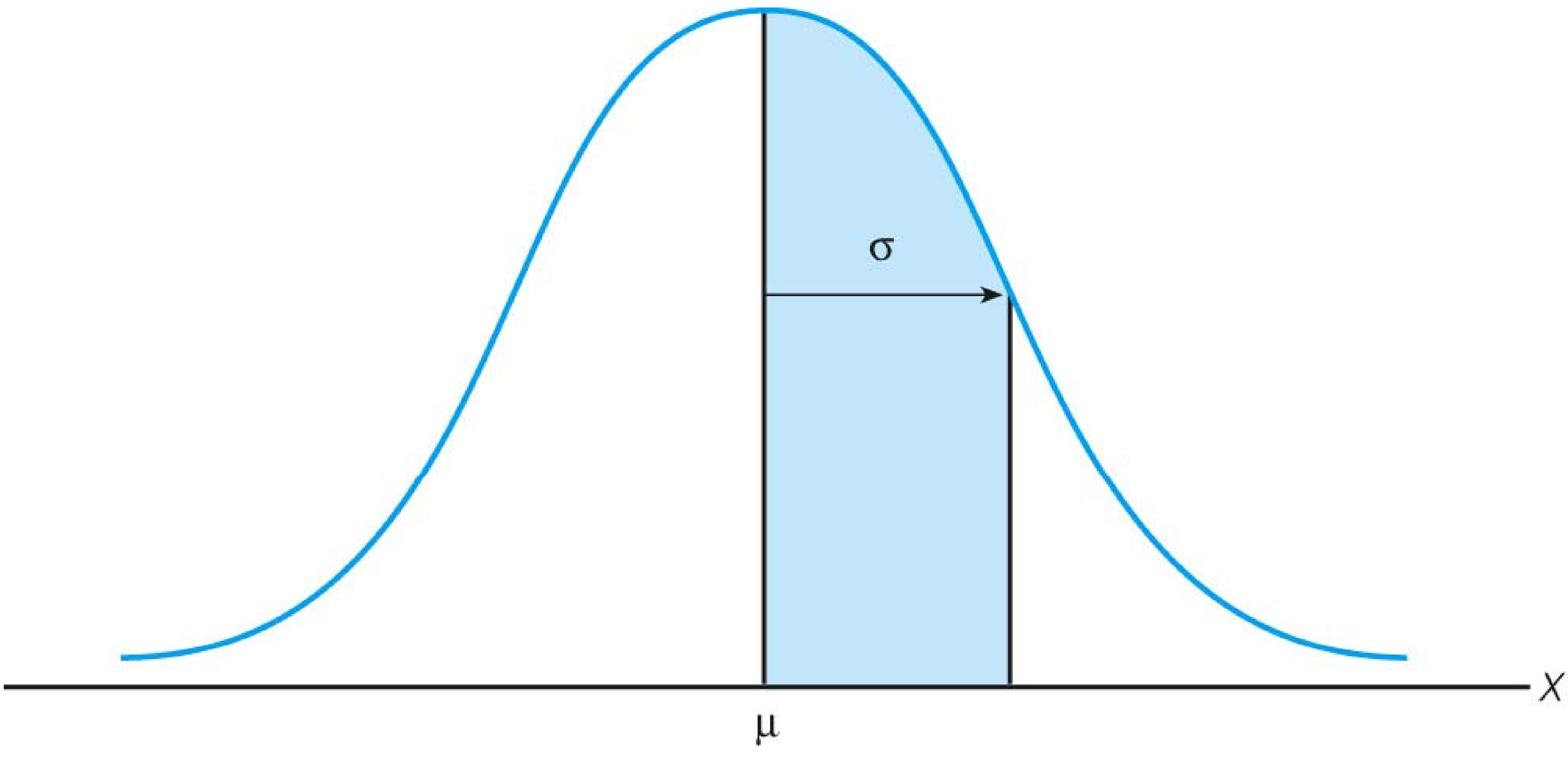


Fig. 6-3, p. 170

Probability and the Normal Distribution cont.

- Figure 6.4 shows a normal distribution with several sections marked in *z*-score units.
- You should recall that *z*-scores measure positions in a distribution in terms of standard deviations from the mean.
- There are two additional points to be made about the distribution shown in Figure 6.4.
  - First, you should realize that the sections on the left side of the distribution have exactly the same areas as the corresponding sections on the right side because the normal distribution is symmetrical.

Probability and the Normal Distribution cont.

- Second, because the locations in the distribution are identified by *z*-scores, the percentages shown in the figure apply to any normal distribution regardless of the values for the mean and the standard deviation.
- Remember: When any distribution is transformed into *z*-scores, the mean becomes zero and the standard deviation becomes one.

- Because the normal distribution is a good model for many naturally occurring distributions and because this shape is guaranteed in some circumstances (as you will see in Chapter 7), we will devote considerable attention to this particular distribution.

FIGURE 6.4

The normal distribution following a z-score transformation.

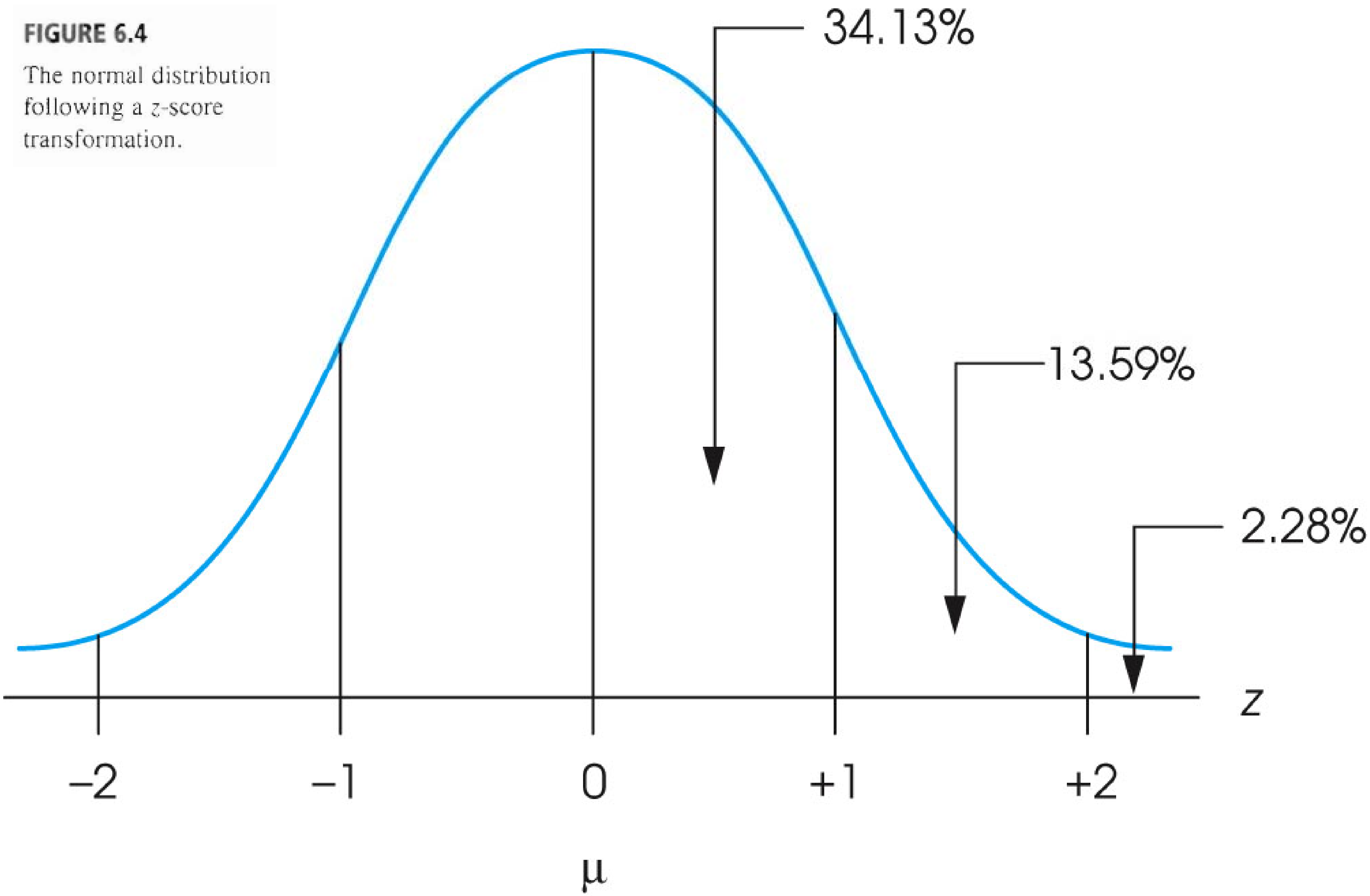


Fig. 6-4, p. 171

Probability and the Normal Distribution cont.

- The process of answering probability questions about a normal distribution is introduced in the following example.
- Example 6.2
  - Assume that the population of adult heights forms a normal-shaped distribution with a mean of μ = 68 inches and a standard deviation of σ = 6 inches.
  - Given this information about the population and the known proportions for a normal distribution (see Figure 6.4), we can determine the probabilities associated with specific samples.

Probability and the Normal Distribution cont.

- For example, what is the probability of randomly selecting an individual from this population who is taller than 6 feet 8 inches (X = 80 inches)?
- Restating this question in probability notation, we get:

$$p(X > 80) = ?$$

- We will follow a step-by-step process to find the answer to this question.
  - 1. First, the probability question is translated into a proportion question: Out of all possible adult heights, what proportion is greater than 80 inches?

Probability and the Normal Distribution cont.

- 2. The set of "all possible adult heights" is simply the population distribution.
- This population is shown in Figure 6.5(a).
- The mean is μ = 68, so the score X = 80 is to the right of the mean.
- Because we are interested in all heights greater than 80, we shade in the area to the right of 80. This area represents the proportion we are trying to determine.
- 3. Identify the exact position of X = 80 by computing a z-score. For this example,

$$z = \frac{X - \mu}{\sigma} = \frac{80 - 68}{6} = \frac{12}{6} = 2.00$$

FIGURE 6.5

The distribution for Example 6.2.

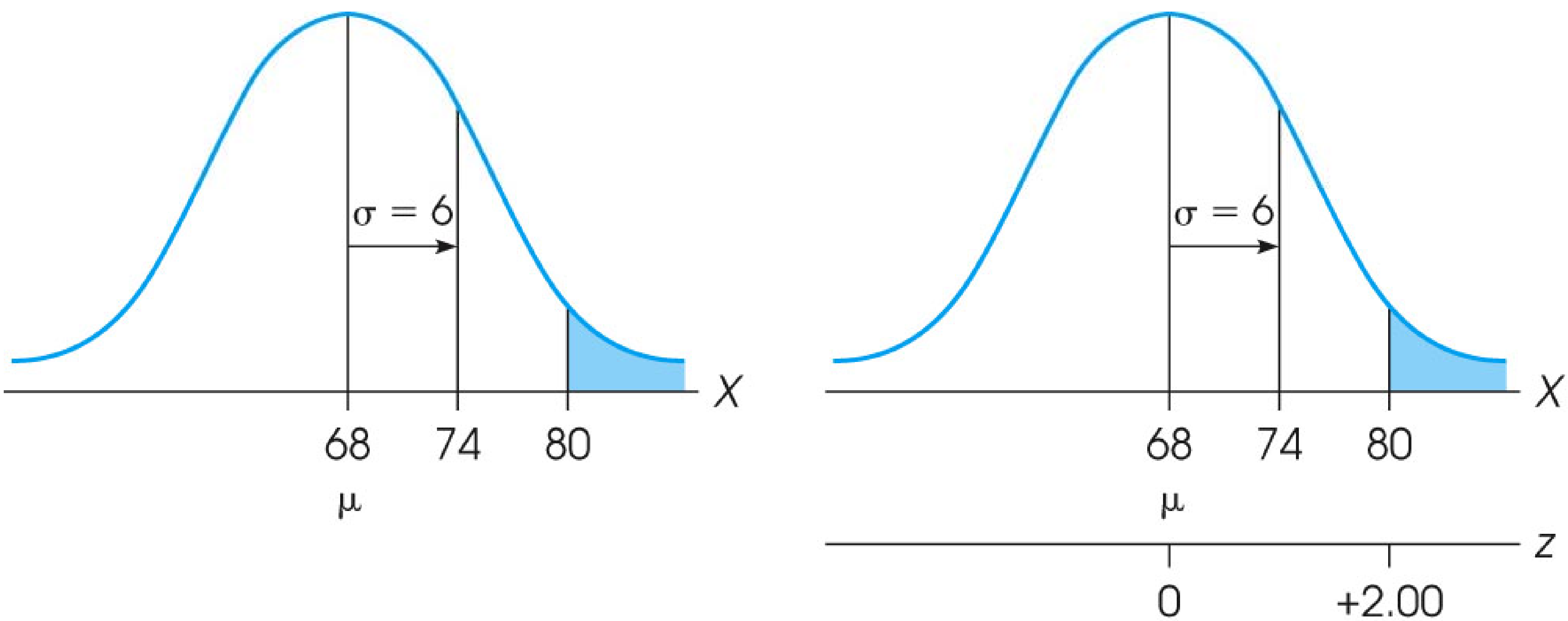


Fig. 6-5, p. 172

Probability and the Normal Distribution cont.

- That is, a height of X = 80 inches is exactly 2 standard deviations above the mean and corresponds to a *z*-score of *z* = *+2.00 [see Figure 6.5(b)].*
- 4. The proportion we are trying to determine may now be expressed in terms of its *z*-score:

$$p(z > 2.00) = ?$$

- According to the proportions shown in Figure 6.4, all normal distributions, regardless of the values for μ and σ , will have 2.28% of the scores in the tail beyond *z* = *+2.00.*

Probability and the Normal Distribution cont.

- Thus, for the population of adult heights,

$$p(X > 80) = p(z > +2.00) = 2.28\%$$

The Unit Normal Table

- To make full use of the unit normal table, there are a few facts to keep in mind:
  - The body always corresponds to the larger part of the distribution whether it is on the right-hand side or the left-hand side.
    - Similarly, the tail is always the smaller section whether it is on the right or the left.
  - Because the normal distribution is symmetrical, the proportions on the right-hand side are exactly the same as the corresponding proportions on the left-hand side.
  -

The Unit Normal Table cont.

- Although the z-score values change signs (+ and -) from one side to the other, the proportions are always positive.
- Thus, column C in the table always lists the proportion in the tail whether it is the right-hand tail or the left-hand tail.

Probabilities, Proportions, and z-Scores

- The unit normal table lists relationships between z-score locations and proportions in a normal distribution.
- For any z-score location, you can use the table to look up the corresponding proportions.
- Similarly, if you know the proportions, you can use the table to look up the specific z-score location.
- Because we have defined probability as equivalent to proportion, you can also use the unit normal table to look up probabilities for normal distributions.

Probabilities, Proportions, and z-Scores cont.

- In the preceding section, we used the unit normal table to find probabilities and proportions corresponding to specific z-score values.
- In most situations, however, it will be necessary to find probabilities for specific X values.
- Consider the following example:
  – It is known that IQ scores form a normal distribution with $\mu = 100$ and $\sigma = 15$.
  – Given this information, what is the probability of randomly selecting an individual with an IQ score less than 130?

Probabilities, Proportions, and *z*-Scores cont.

- Specifically, what is the probability of randomly selecting an individual with an IQ score less than 130?

$$p(X < 130) = ?$$

- Restated in terms of proportions, we want to find the proportion of the IQ distribution that corresponds to scores less than 130.
- The distribution is drawn in Figure 6.9, and the portion we want has been shaded.

Probabilities, Proportions, and z-Scores cont.

- The first step is to change the X values into z-scores.
  - In particular, the score of X = 130 is changed to

$$z = \frac{X - \mu}{\sigma} = \frac{130 - 100}{15} = \frac{30}{15} = 2.00$$

- Thus, an IQ score of X = 130 corresponds to a z-score of z =+2.00, and IQ scores less than 130 correspond to z-scores less than 2.00.
- Next, look up the z-score value in the unit normal table.

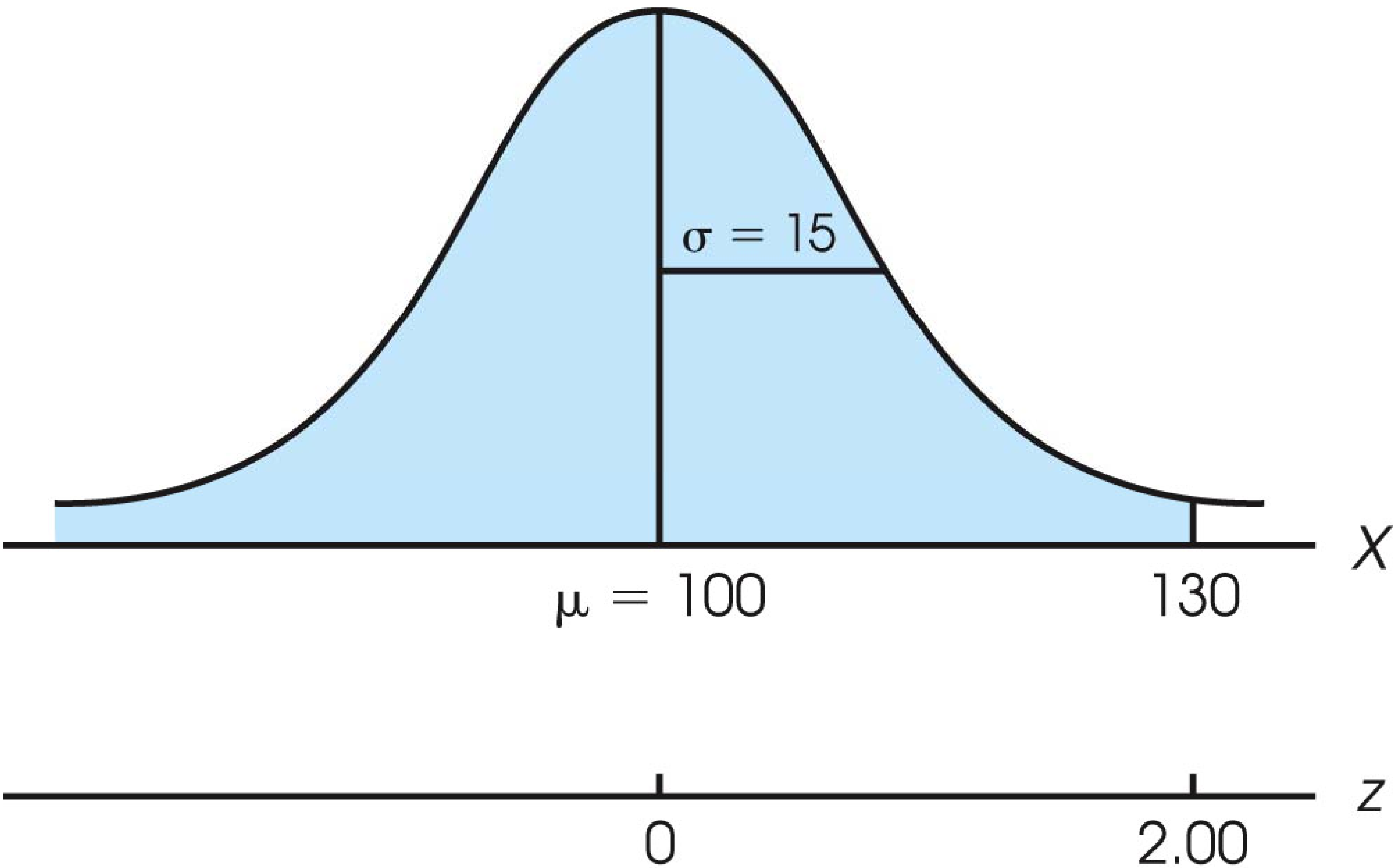


Fig. 6-9, p. 178

Probabilities, Proportions, and z-Scores cont.

- Because we want the proportion of the distribution in the body to the left of X = 130 (see Figure 6.9), the answer will be found in column B.
- Consulting the table, we see that a z-score of 2.00 corresponds to a proportion of 0.9772.
- The probability of randomly selecting an individual with an IQ less than 130 is 0.9772:

$$p(X < 130) = p(z < 2.00) = 0.9772 \text{ (or } 97.72\%\text{)}$$

- Finally, notice that we phrased this question in terms of a probability.
- Specifically, we asked, "What is the probability of selecting an individual with an IQ less than 130?"

Probabilities, Proportions, and z-Scores cont.

- However, the same question can be phrased in terms of a proportion: "What proportion of the individuals in the population have IQ scores less than 130?"
- Both versions ask exactly the same question and produce exactly the same answer.

Finding Proportions/Probabilities Located Between Two Scores

- The highway department conducted a study measuring driving speeds on a local section of interstate highway.
- They found an average speed of μ. = 58 miles per hour with a standard deviation of $\sigma = 10$.
- The distribution was approximately normal.
- Given this information, what proportion of the cars are traveling between 55 and 65 miles per hour?
- Using probability notation, we can express the problem as

$$p(55 < X < 65) = ?$$

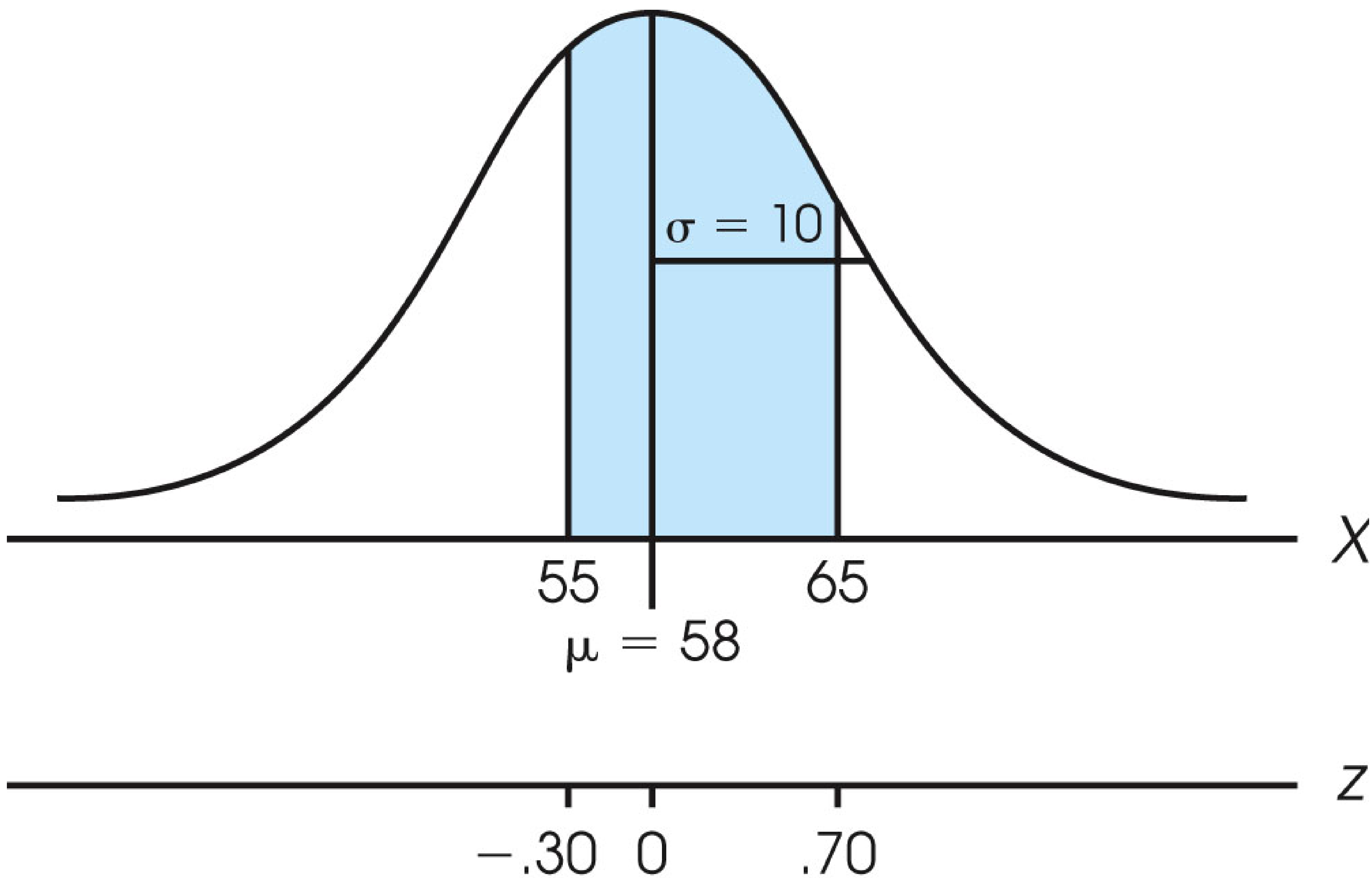


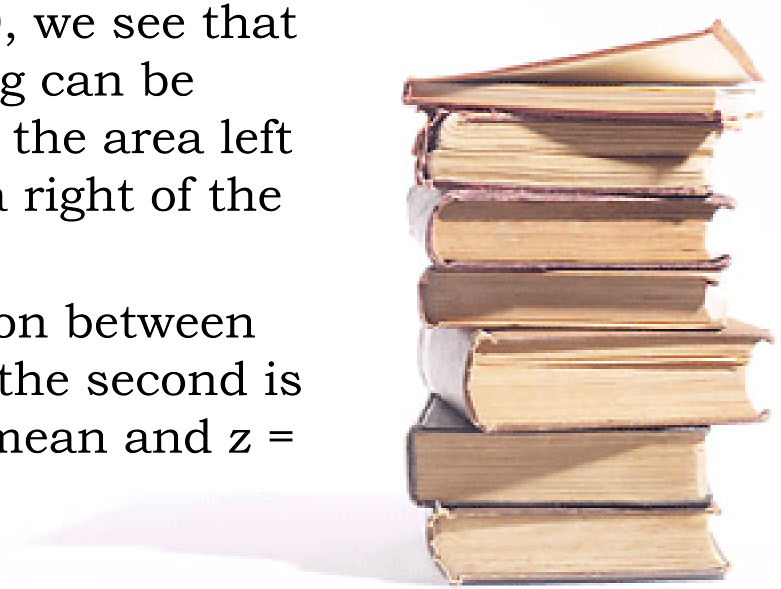
Fig. 6-10, p. 179

Finding Proportions/Probabilities Located Between Two Scores cont.

$$\text{For } X = 55: \quad z = \frac{X - \mu}{\sigma} = \frac{55 - 58}{10} = \frac{-3}{10} = -0.30$$

$$\text{For } X = 65: \quad z = \frac{X - \mu}{\sigma} = \frac{65 - 58}{10} = \frac{7}{10} = 0.70$$

- Looking again at Figure 6.10, we see that the proportion we are seeking can be divided into two sections: (1) the area left of the mean, and (2) the area right of the mean.
- The first area is the proportion between the mean and z = -0.30 and the second is the proportion between the mean and z = +0.70.



Finding Proportions/Probabilities Located Between Two Scores cont.

- Using column D of the unit normal table, these two proportions are 0.1179 and 0.2580.
- The total proportion is obtained by adding these two sections:

$$p(55 < X < 65) = p(-0.30 < z < +0.70) = 0.1179 + 0.2580 = 0.3759$$

Finding Scores Corresponding to Specific Proportions or Probabilities

- Scores on the SAT form a normal distribution with $\mu = 500$ and $\sigma = 100$.
- What is the minimum score necessary to be in the top 15% of the SAT distribution? (An alternative form of the same question is presented in Box 6.2.)
- This problem is shown graphically in Figure 6.13.
- In this problem, we begin with a proportion (15% = 0.15), and we are looking for a score.
- According to the map in Figure 6.12, we can move from p (proportion) to X (score) via z-scores.

Finding Scores Corresponding to Specific Proportions or Probabilities cont.

- The first step is to use the unit normal table to find the *z*-score that corresponds to a proportion of 0.15.
- Because the proportion is located beyond *z* in the tail of the distribution, we will look in column C for a proportion of 0.1500.
  - Note that you may not find 0.1500 exactly, but locate the closest value possible.
  - In this case, the closest value in the table is 0.1492, and the *z*-score that corresponds to this proportion is $z = 1.04$.
- The next step is to determine whether the *z*-score is positive or negative.
- Remember that the table does not specify the sign of the *z*-score.

Finding Scores Corresponding to Specific Proportions or Probabilities cont.

- Looking at the graph in Figure 6.13, you should realize that the score we want is above the mean, so the *z*-score is positive, $z = +1.04$.
- Now you are ready for the last stage of the solution-that is, changing the *z*-score into an X value.
- By definition, a z-score of $z = +1.04$ corresponds to an X value that is located above the mean (+) by 1.04 standard deviations.
- One standard deviation is $\sigma = 100$ points, so 1.04 standard deviations is

$$1.04\sigma = 1.04(100) = 104 \text{ points}$$

Finding Scores Corresponding to Specific Proportions or Probabilities cont.

- Thus, our score is located 104 points above the mean.
- With a mean of $\mu = 500$, the score is $X = 500 + 104 = 604$.
- The conclusion for this example is that you must have an SAT score of at least 604 to be in the top 15% of the distribution.

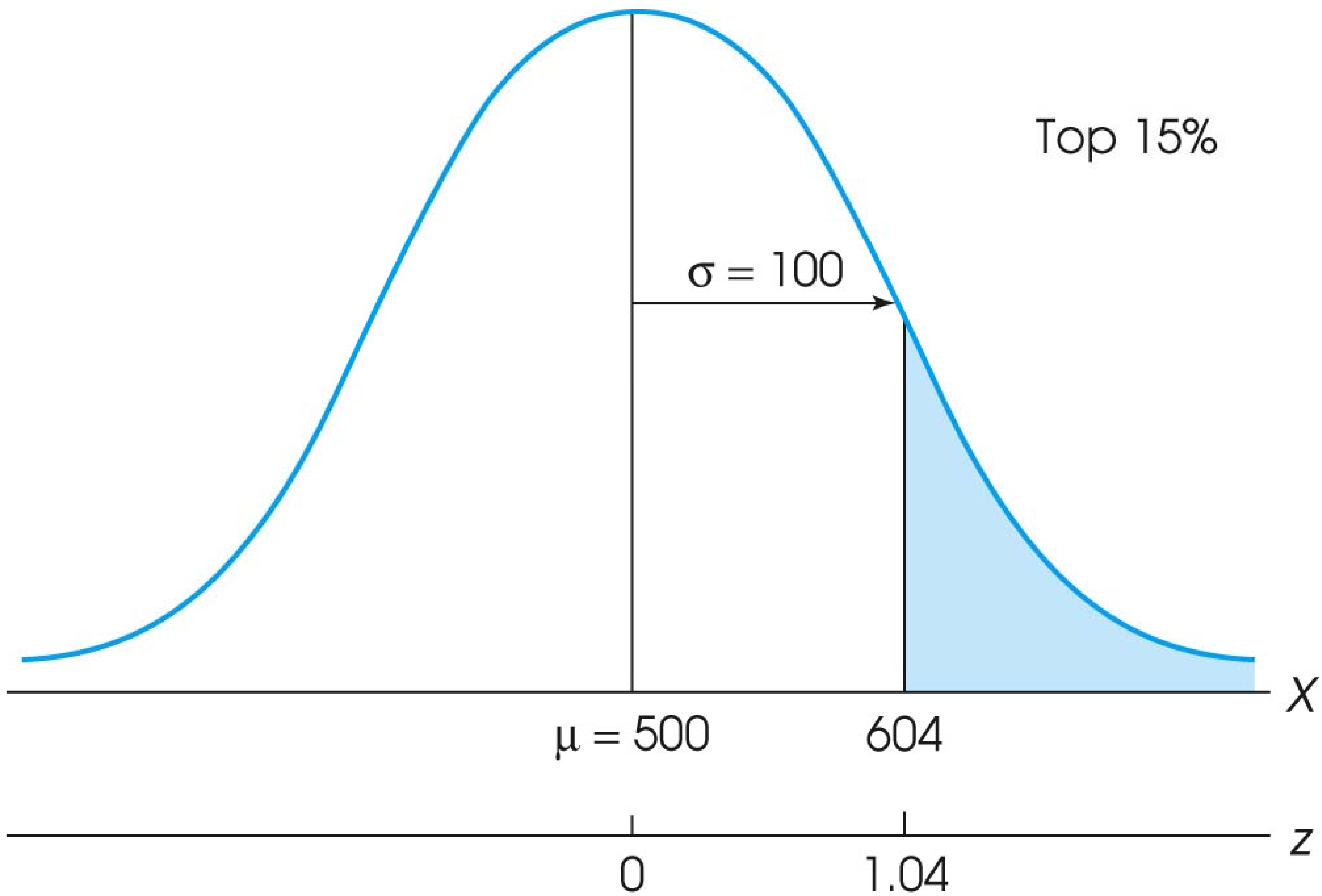


Fig. 6-13, p. 181