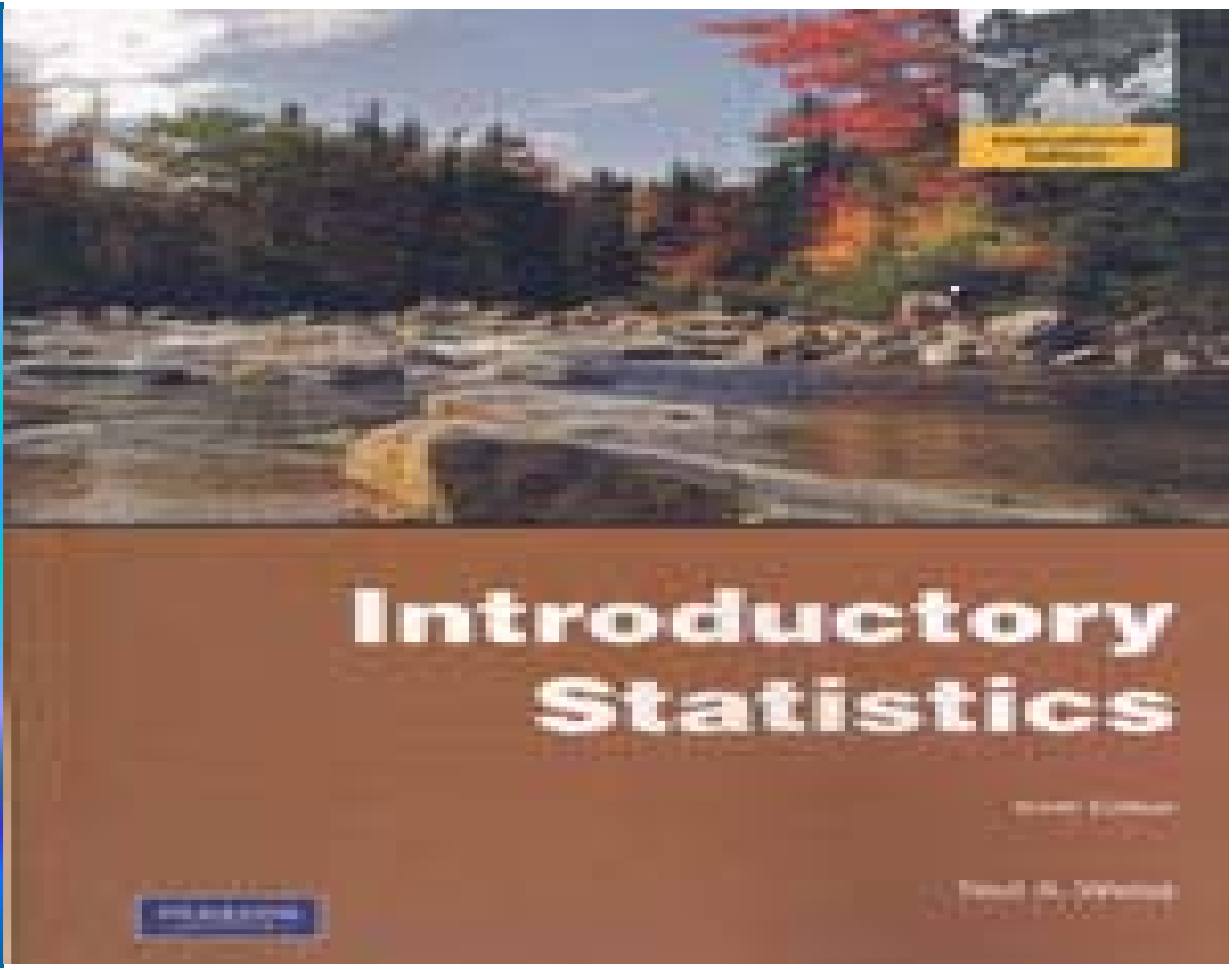

Introductory
Statistics

Chapter 6

The Normal Distribution

PEARSON

Chapter Goals

- Learn about the *normal*, *bell-shaped*, or *Gaussian* distribution
- How probabilities are found
- How probabilities are represented
- How normal distributions are used in the real world

Section 6.1

Introducing Normally Distributed Variables

6.1 ~ Normal Probability Distributions

- The normal probability distribution is the most important distribution in all of statistics

- Many continuous random variables have normal or approximately normal distributions

- Need to learn how to describe a normal probability distribution

Percentage, Proportion & Probability

- Basically the same concepts
- *Percentage* (30%) is usually used when talking about a *proportion* (3/10) of a population
- *Probability* is usually used when talking about the chance that the next individual item will possess a certain property
- *Area* is the graphic representation of all three when we draw a picture to illustrate the situation

Key Fact 6.1

Basic Properties of Density Curves

Property 1: A density curve is always on or above the horizontal axis.

Property 2: The total area under a density curve (and above the horizontal axis) equals 1.

Key Fact 6.2

Variables and Their Density Curves

For a variable with a density curve, the percentage of all possible observations of the variable that lie within any specified range equals (at least approximately) the corresponding area under the density curve, expressed as a percentage.

Normal Probability Distribution

1. A continuous random variable
2. Description involves two functions:
   a. A function to determine the ordinates of the graph picturing the distribution
   b. A function to determine probabilities
3. Normal probability distribution function:

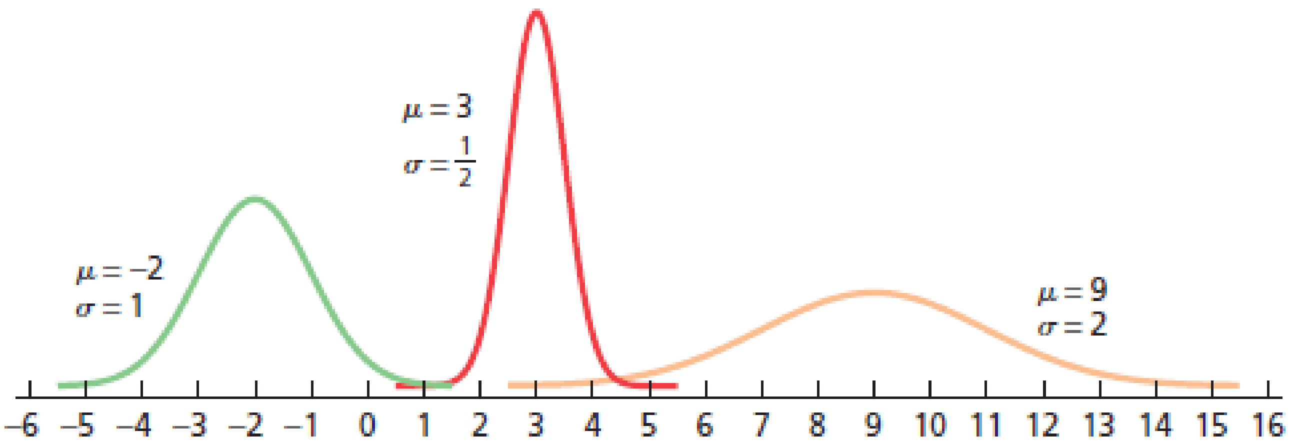
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}$$

   This is the function for the normal (bell-shaped) curve

4. The probability that *x* lies in some interval is the area under the curve

Figure 6.2

Three normal distributions



EXAMPLE 6.1 A Normally Distributed Variable

- Heights of Female College Students: A mid-western college has an enrollment of 3264 female students. Records show that the mean height of these students is 64.4 inches and that the standard deviation is 2.4 inches. Here the variable is height, and the population consists of the 3264 female students attending the college. Frequency and relative-frequency distributions for these heights appear in Table 6.1. The table shows, for instance, that 7.35% (0.0735) of the students are between 67 and 68 inches tall.

Table 6.1

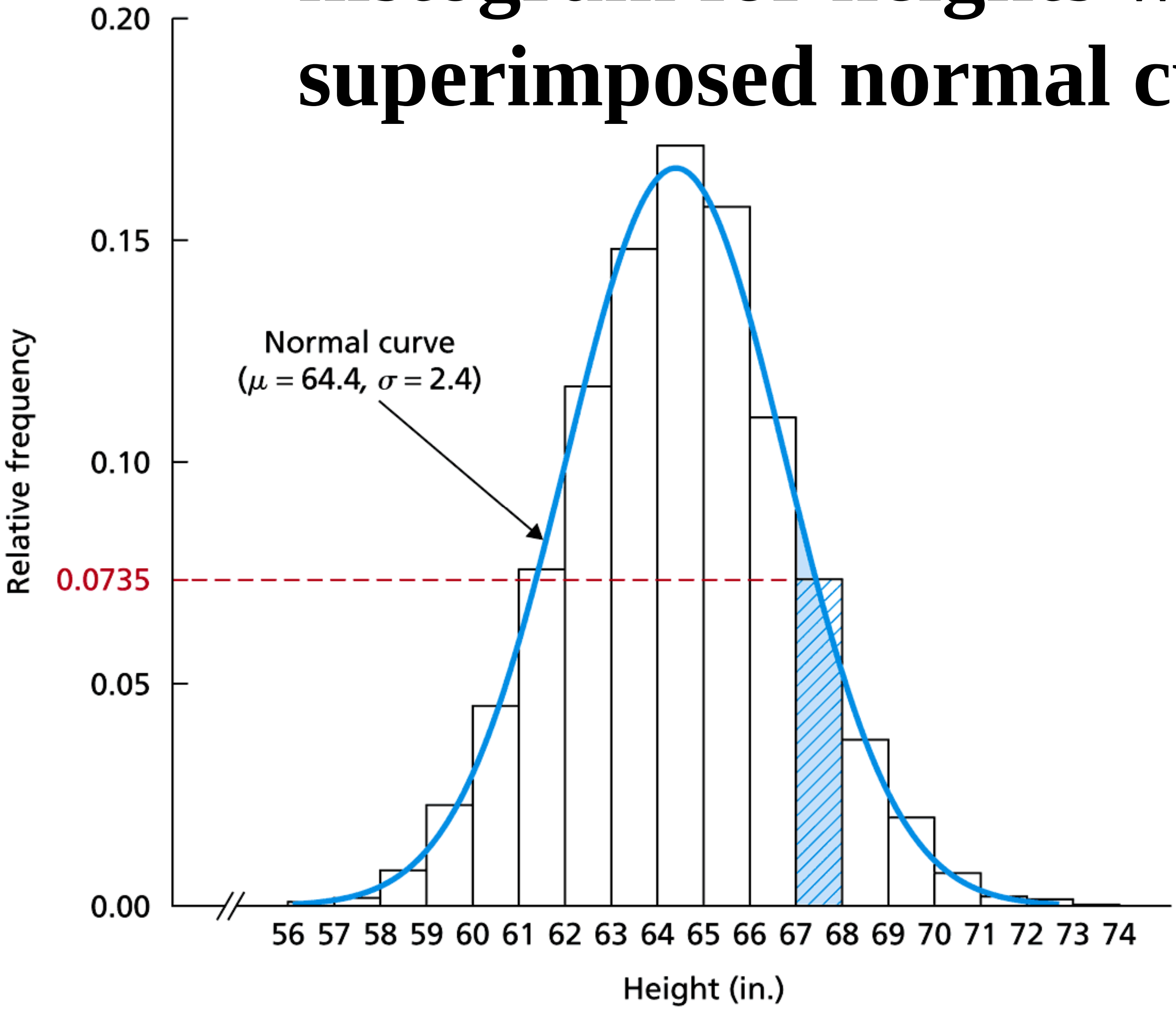
Frequency and relative-frequency distributions for heights

How to describe or present the data on the heights of 3264 females at certain college?

Height (in.)	Frequency f	Relative frequency
56–under 57	3	0.0009
57–under 58	6	0.0018
58–under 59	26	0.0080
59–under 60	74	0.0227
60–under 61	147	0.0450
61–under 62	247	0.0757
62–under 63	382	0.1170
63–under 64	483	0.1480
64–under 65	559	0.1713
65–under 66	514	0.1575
66–under 67	359	0.1100
67–under 68	240	0.0735
68–under 69	122	0.0374
69–under 70	65	0.0199
70–under 71	24	0.0074
71–under 72	7	0.0021
72–under 73	5	0.0015
73–under 74	1	0.0003
	3264	1.0000

Figure 6.4

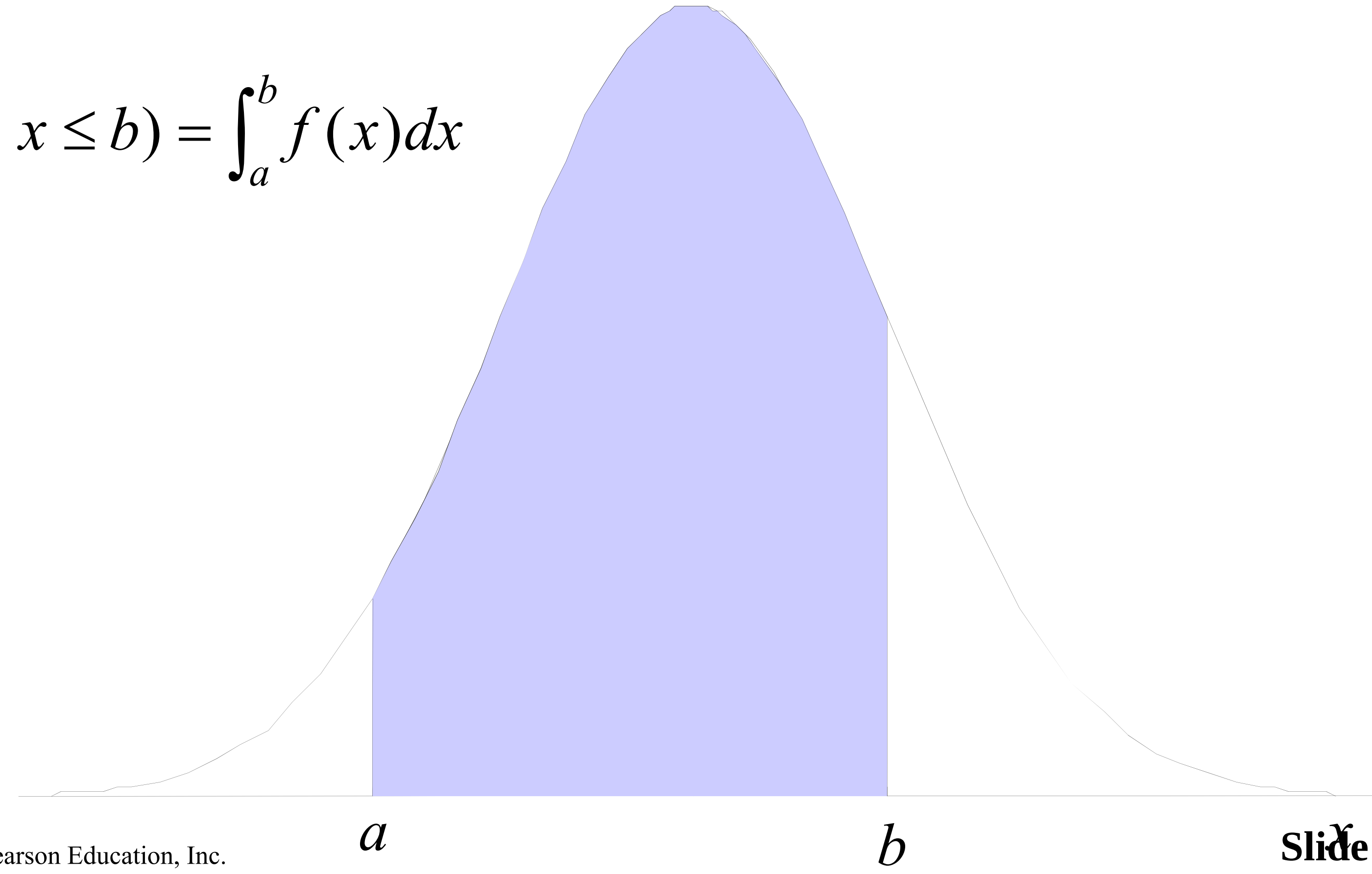
Relative-frequency histogram for heights with superimposed normal curve



Probabilities for a Normal Distribution

- Illustration

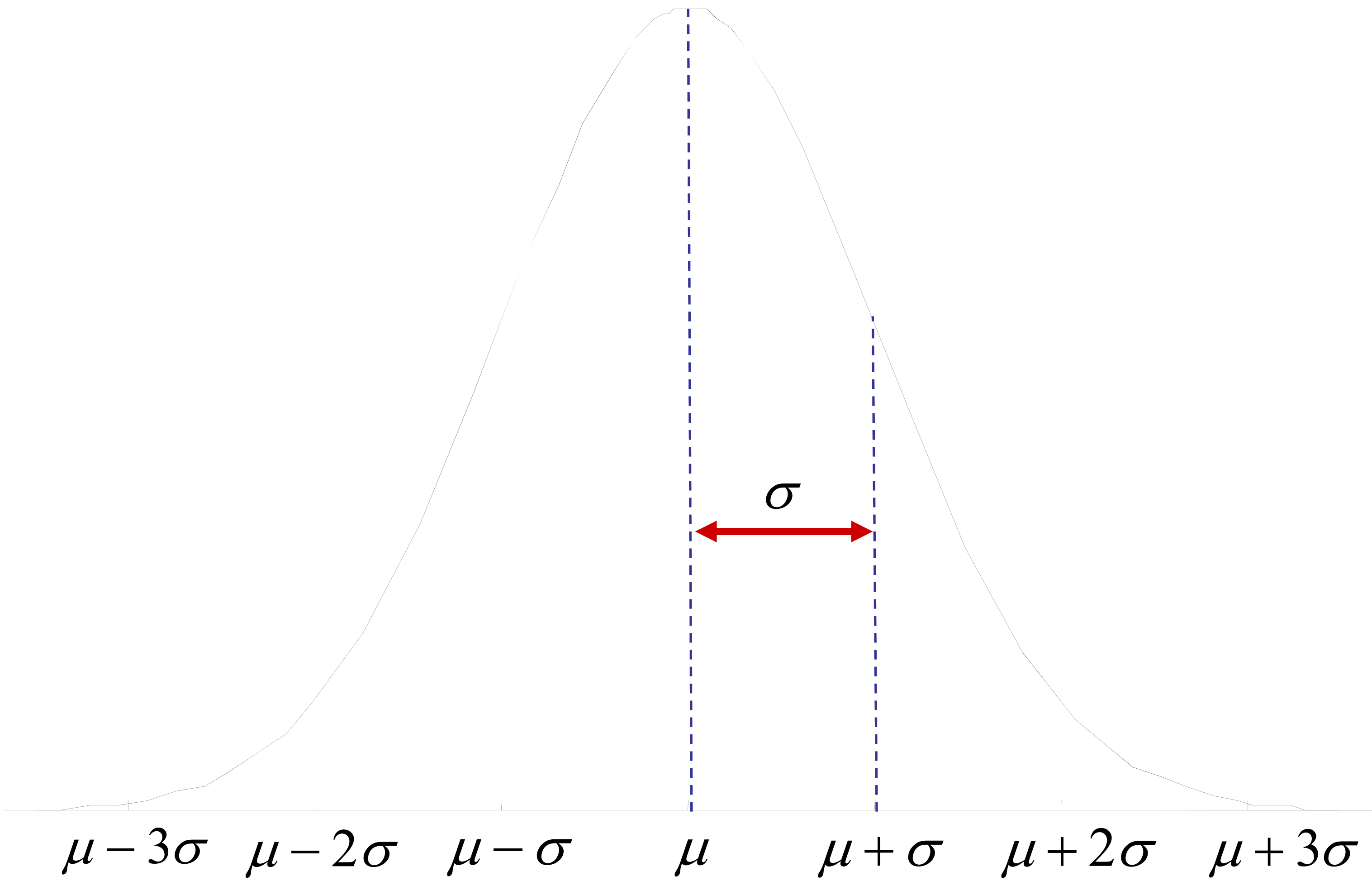
$$P(a \le x \le b) = \int_a^b f(x)\,dx$$



Notes

- The definite integral is a calculus topic
- We will use a table to find probabilities for normal distributions
- We will learn how to compute probabilities for one special normal distribution: the standard normal distribution
- Transform all other normal probability questions to this special distribution
- Recall the empirical rule: the percentages that lie within certain intervals about the mean come from the normal probability distribution
- We need to refine the empirical rule to be able to find the percentage that lies between any two numbers

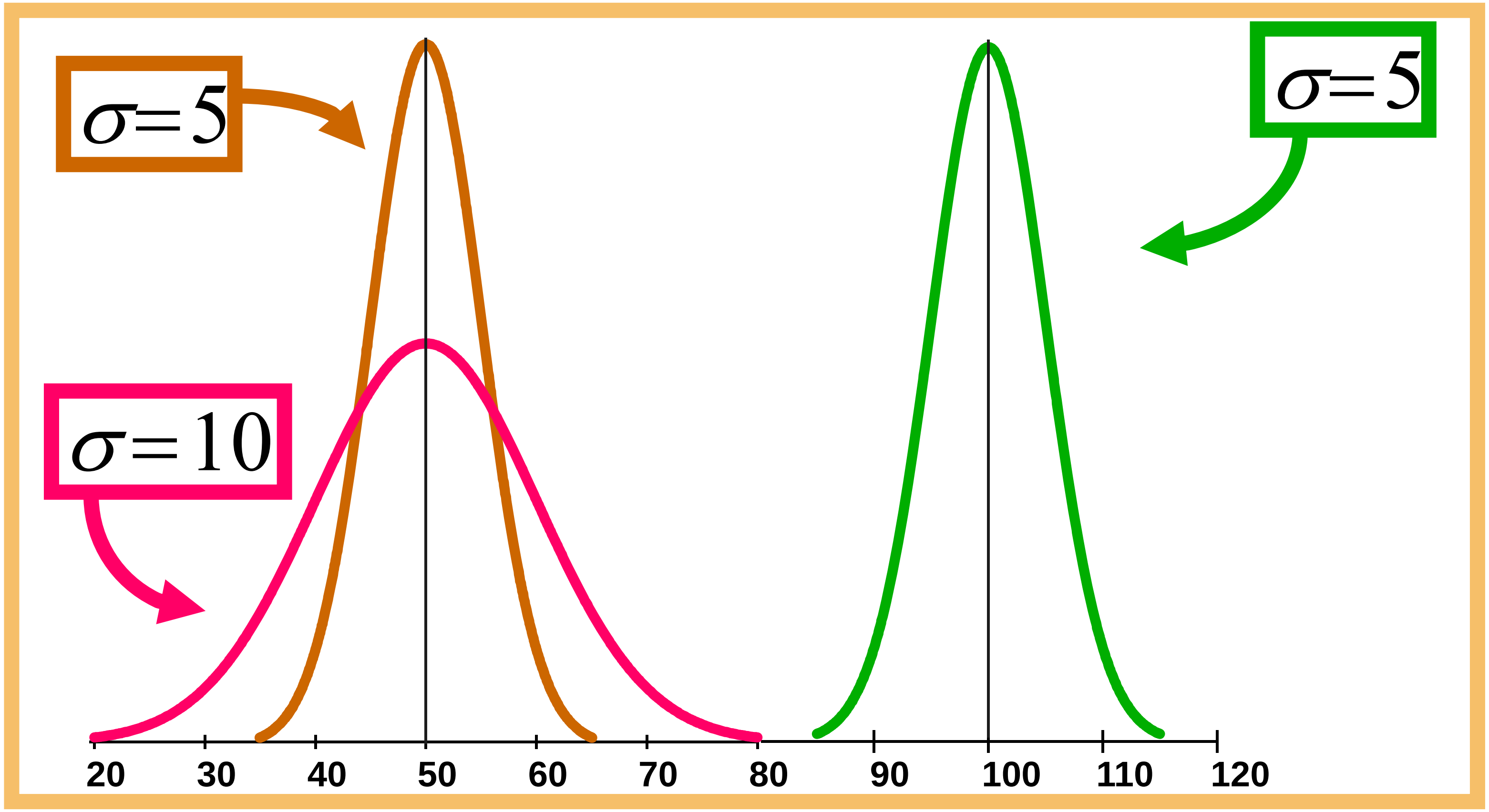
The Normal Probability Distribution



The Standard Normal Distribution

- There are infinitely many normal probability distributions
- They are all related to the *standard normal distribution*
- The standard normal distribution is the normal distribution of the standard variable z (the z-score)

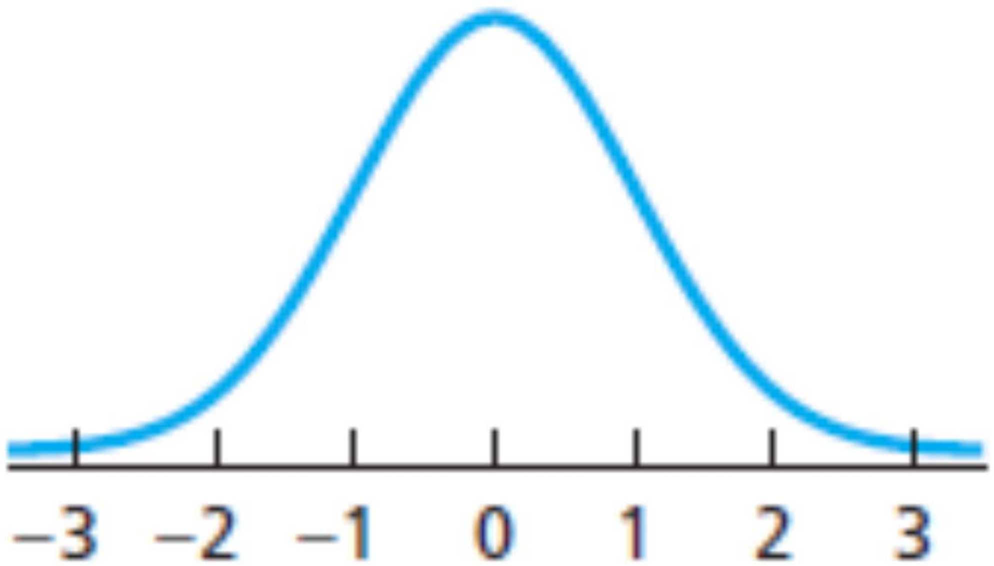
Normal Curves for Different Means and Standard Deviations



Definition 6.2

Standard Normal Distribution; Standard Normal Curve

A normally distributed variable having mean 0 and standard deviation 1 is said to have the **standard normal distribution.** Its associated normal curve is called the **standard normal curve,** which is shown in Fig. 6.5.



Key Fact 6.4

Standardized Normally Distributed Variable

The standardized version of a normally distributed variable x,

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

has the standard normal distribution.

Figure 6.6

Standardizing normal distributions

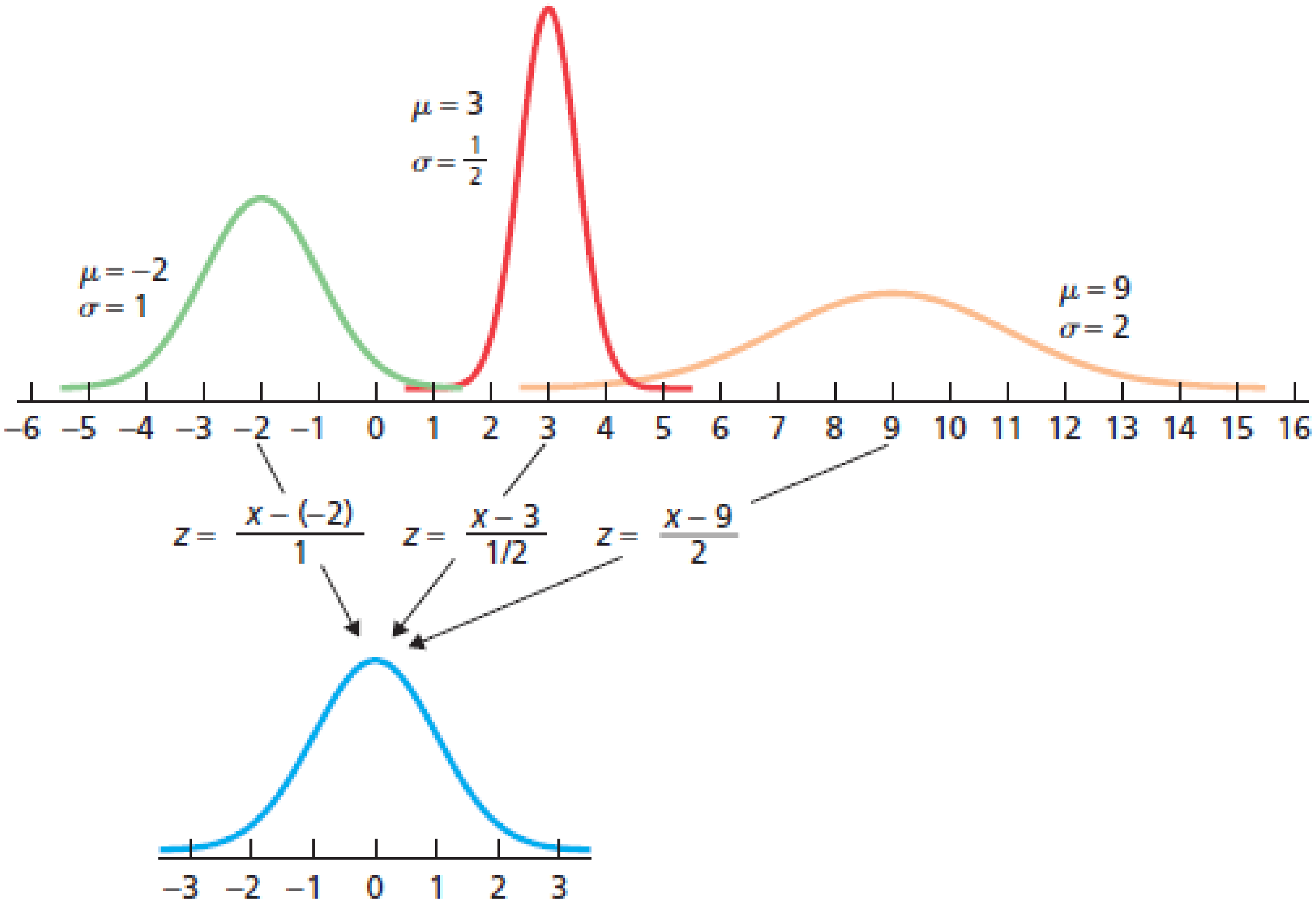
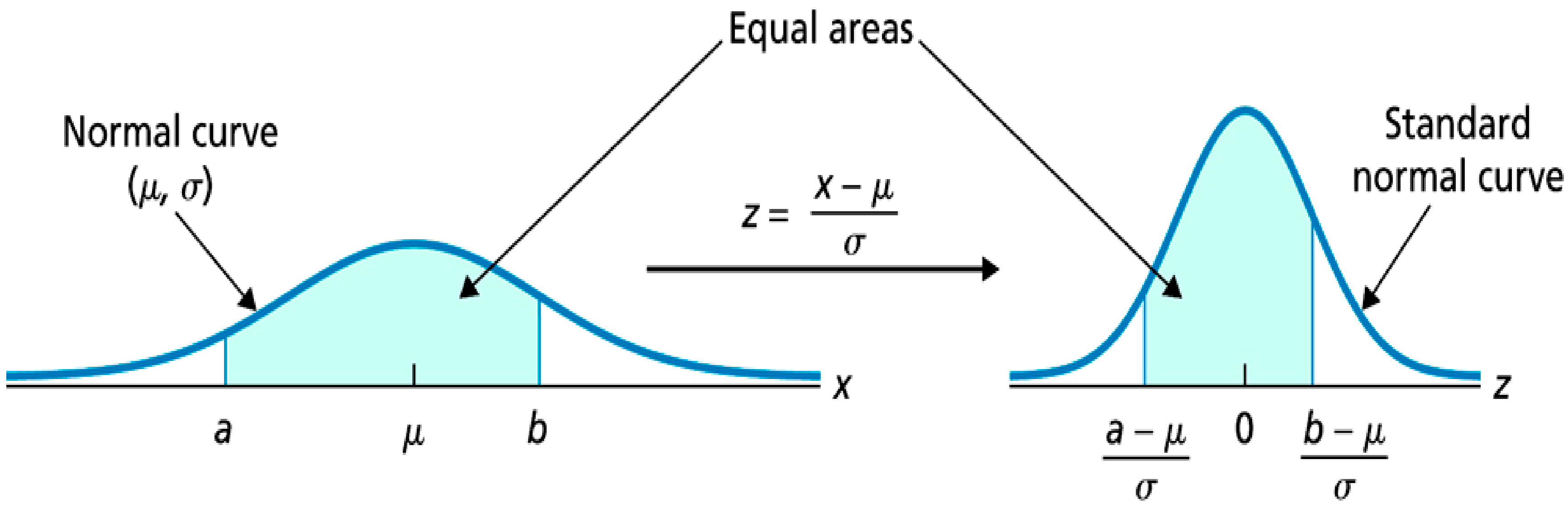


Figure 6.7

Finding percentages for a normally distributed variable from areas under the standard normal curve



Section 6.2

Areas Under the Standard Normal Curve

Key Fact 6.5

Basic Properties of the Standard Normal Curve

Property 1: The total area under the standard normal curve is 1.

Property 2: The standard normal curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.

Property 3: The standard normal curve is symmetric about 0; that is, the part of the curve to the left of the dashed line in Fig. 6.8 is the mirror image of the part of the curve to the right of it.

Property 4: Almost all the area under the standard normal curve lies between -3 and 3.

Standard Normal Distribution

Properties:

- The total area under the normal curve is equal to 1
- The distribution is mounded and symmetric; it extends indefinitely in both directions, approaching but never touching the horizontal axis
- The distribution has a mean of 0 and a standard deviation of 1
- The mean divides the area in half, 0.50 on each side
- Nearly all the area is between $z = -3.00$ and $z = 3.00$

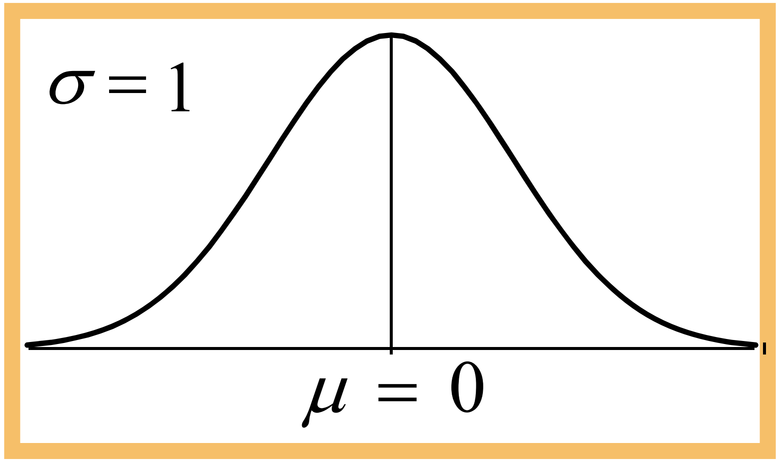
STANDARD NORMAL DISTRIBUTION

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

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

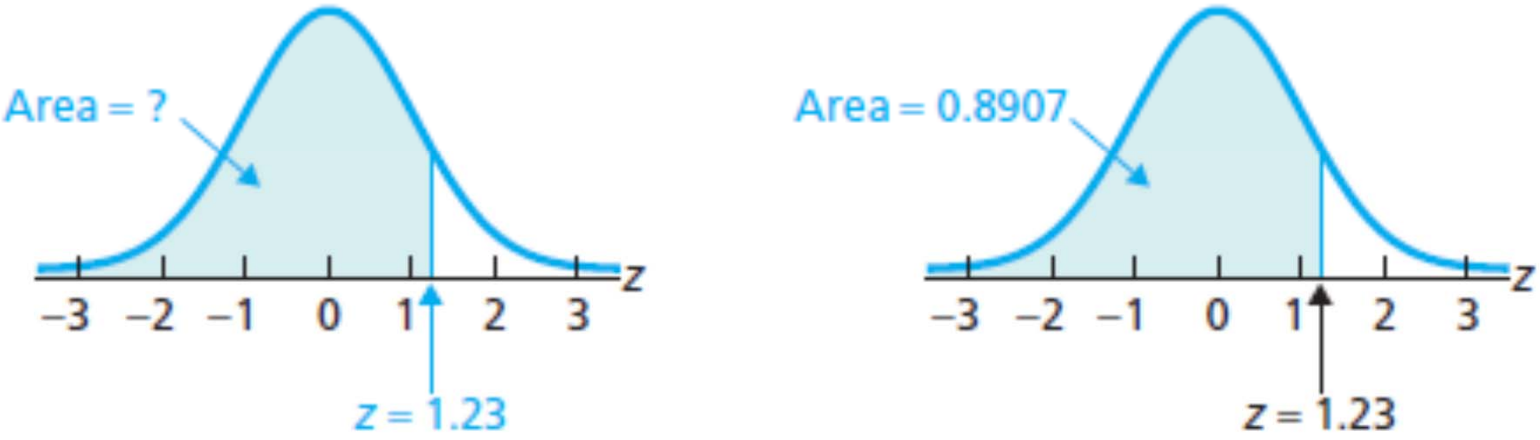
$$z = \frac{x - \mu}{\sigma}$$

has the standard Normal distribution.



EXAMPLE 6.3 Finding the Area to the Left of a Specified *z*-Score

Determine the area under the standard normal curve that lies to the left of 1.23, as shown in Fig. 6.9(a).



(a) (b)

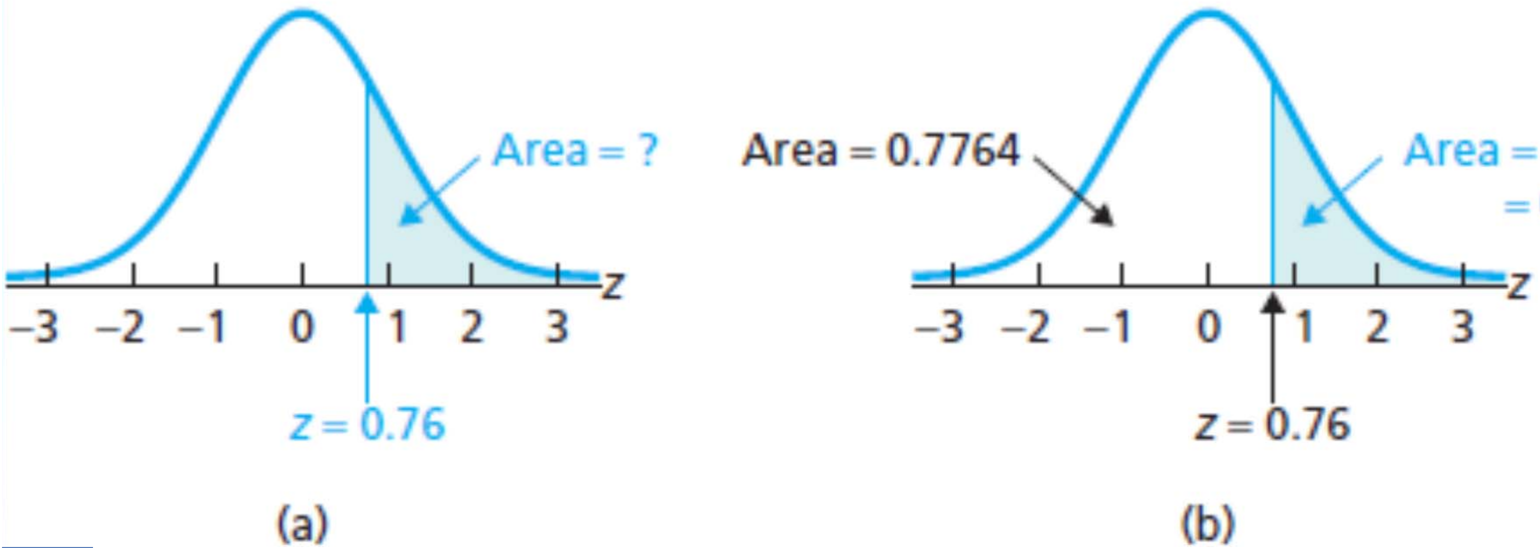
Table 6.2

Areas under the standard normal curve

Second decimal place in z										
0.09	*0.08*	*0.07*	*0.06*	*0.05*	*0.04*	*0.03*	*0.02*	*0.01*	*0.00*	z
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

EXAMPLE 6.4 Finding the Area to the Right of a Specified *z*-Score

Determine the area under the standard normal curve that lies to the right of 0.76, as shown in Fig. 6.10(a).



EXAMPLE 6.5 Finding the Area between Two Specified *z*-Scores

Determine the area under the standard normal curve that lies between –0.68 and 1.82, as shown in Fig. 6.11(a).

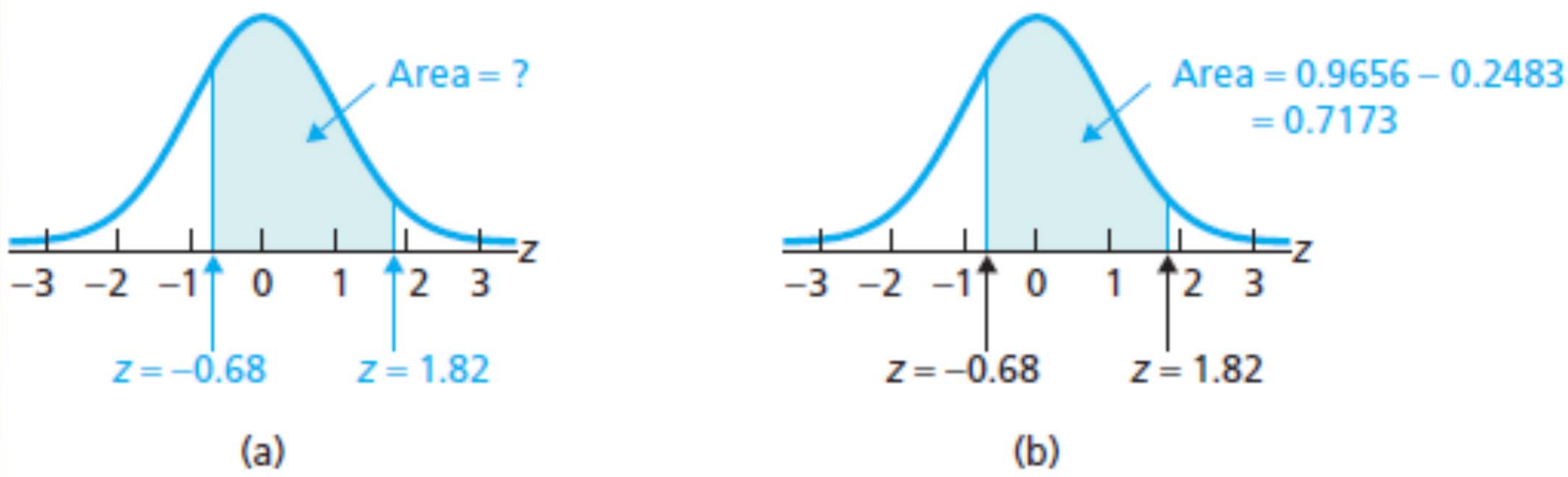
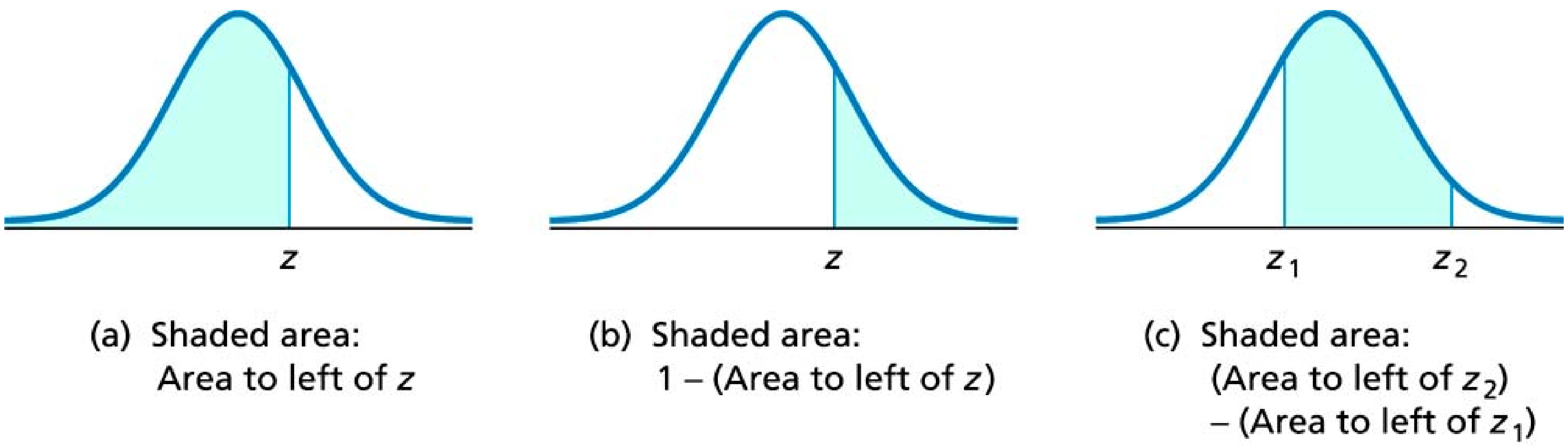


Figure 6.12

Using Table II to find the area under the standard normal curve that lies (a) to the left of a specified z-score, (b) to the right of a specified z-score, and (c) between two specified z-scores



EXAMPLE

The weekly starting salaries of part-time students follow the normal distribution with a mean of $2,000 and a standard deviation of $200. What is the *z*-value for a salary of $2,200?

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

$$= \frac{\$2{,}200 - \$2{,}000}{\$200} = 1.00$$

EXAMPLE *continued*

What is the z-value of $1,700.

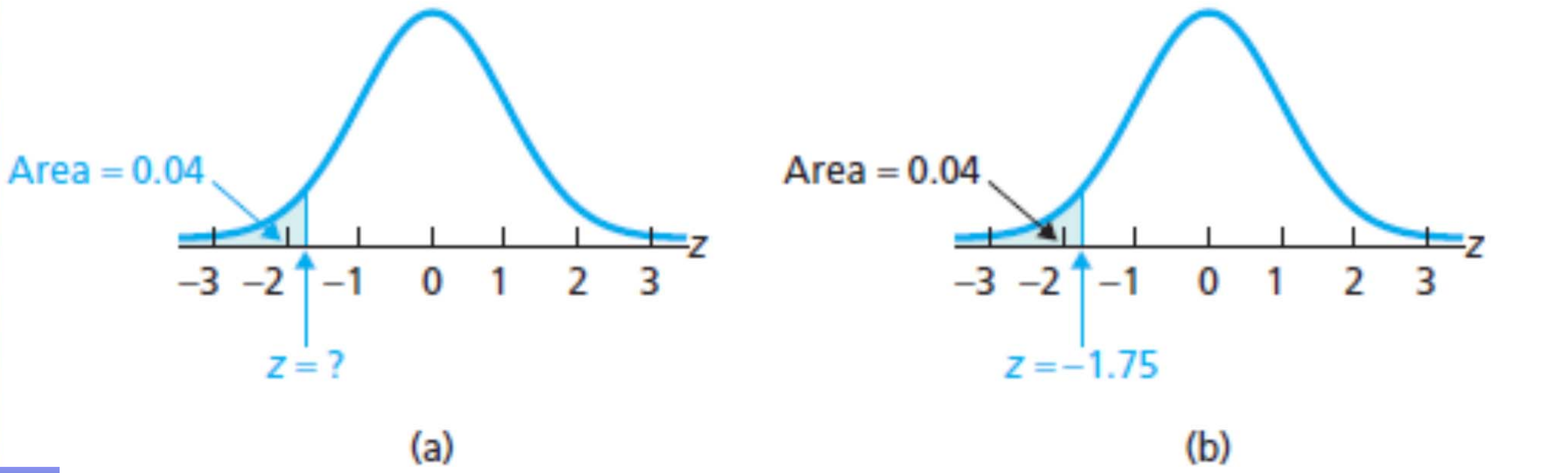
$$z = \frac{X - \mu}{\sigma}$$

$$= \frac{\$1,700 - \$2,000}{\$200} = -1.50$$

A *z-v*alue of 1 indicates that the value of $2,200 is one standard deviation above the mean of $2,000. A *z-v*alue of –1.50 indicates that $1,700 is 1.5 standard deviation below the mean of $2000.

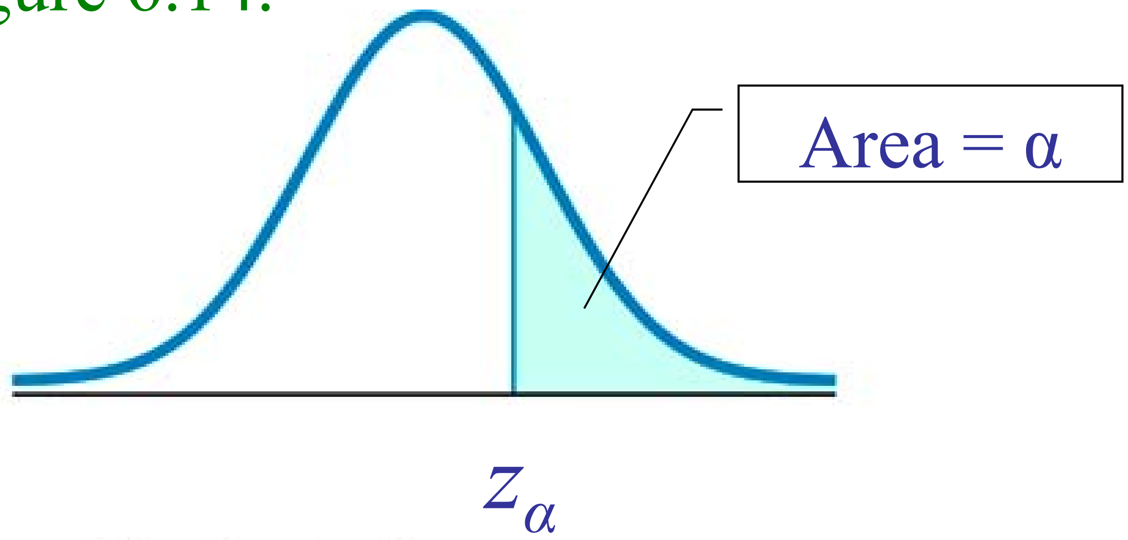
EXAMPLE 6.6 Finding the *z*-Score Having a Specified Area to Its Left

Determine the z-score having an area of 0.04 to its left under the standard normal curve, as shown in Fig. 6.13(a).



DEFINITION 6.3 The Z_α Notation

The symbol z_α is used to denote the z-score that has an area of α to its right under the standard normal curve, as illustrated in Figure 6.14.

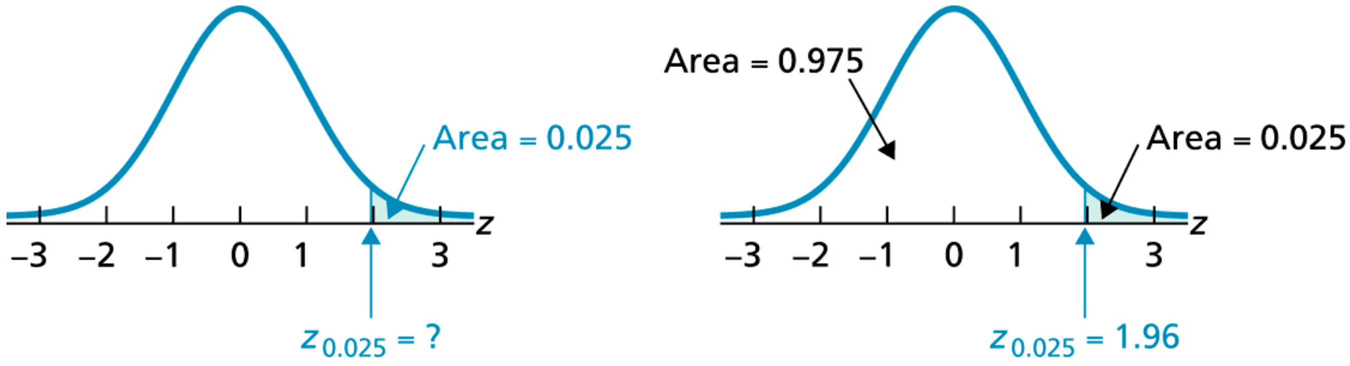


(b) Shaded area:
1 – (Area to left of z)

Figures 6.15 & 6.16

Finding z

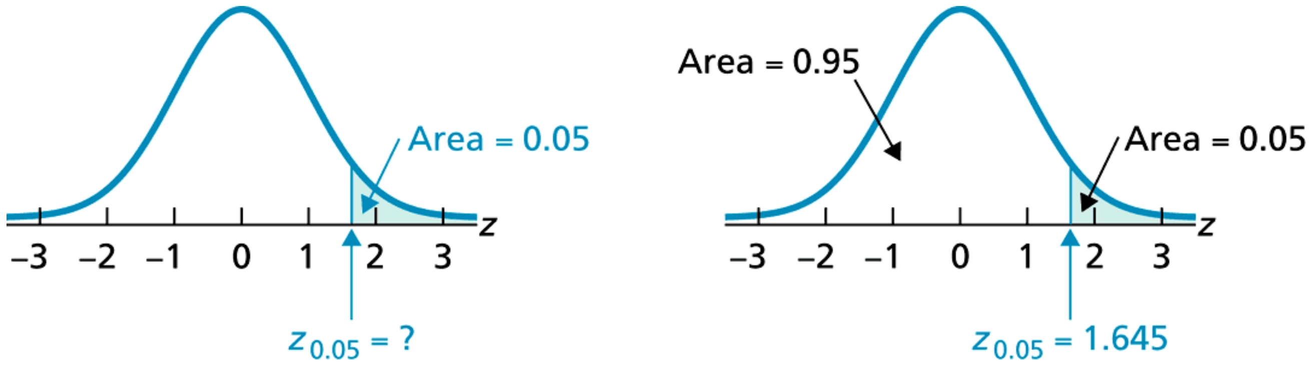
0.025



(b)

Finding z

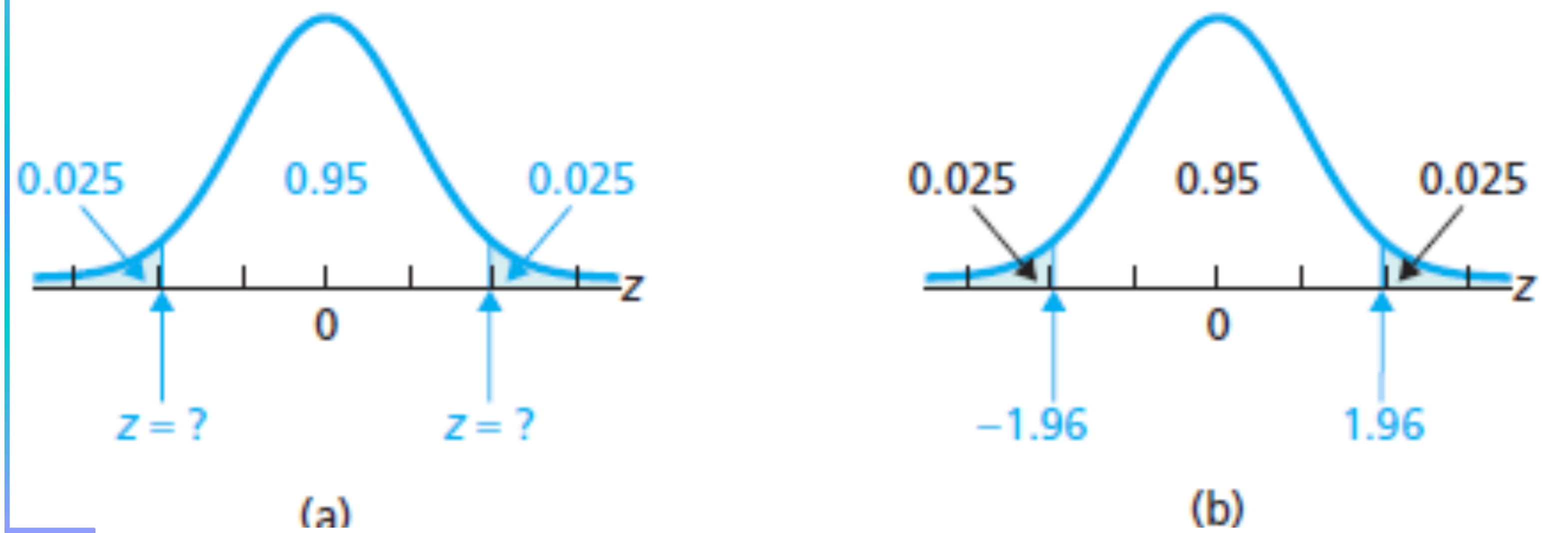
0.05



(a) (b)

EXAMPLE 6.8 Finding the z-Scores for a Specified Area

Find the two z-scores that divide the area under the standard normal curve into a middle 0.95 area and two outside 0.025 areas, as shown in Fig. 6.17(a).



Section 6.3

Working with Normally Distributed Variables

Procedure 6.1

To Determine a Percentage or Probability for a Normally Distributed Variable

Step 1 Sketch the normal curve associated with the variable.

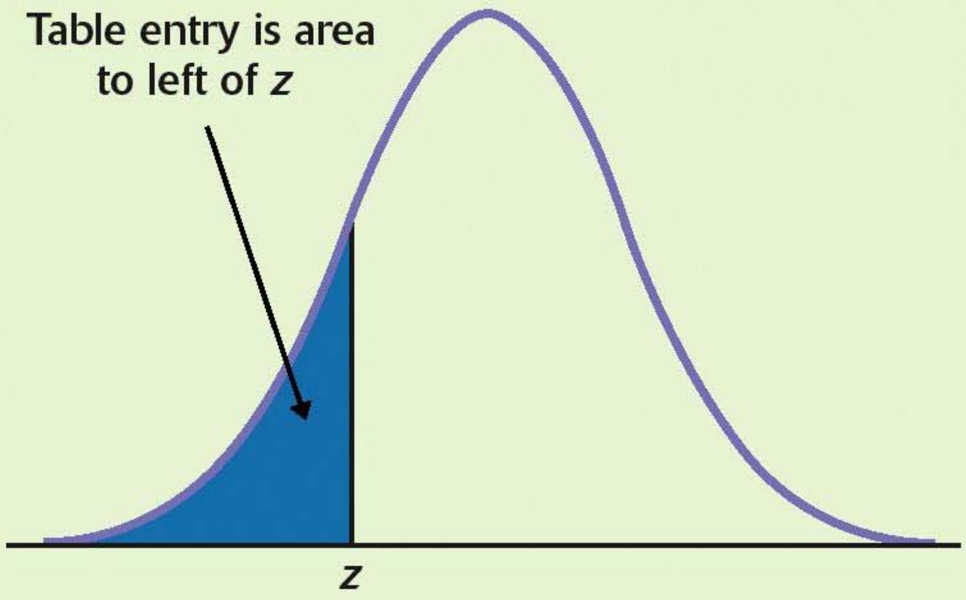
Step 2 Shade the region of interest and mark its delimiting x-value(s).

Step 3 Find the z-score(s) for the delimiting x-value(s) found in Step 2.

Step 4 Use Table II to find the area under the standard normal curve delimited by the z-score(s) found in Step 3.

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.



FINDING NORMAL PROPORTIONS

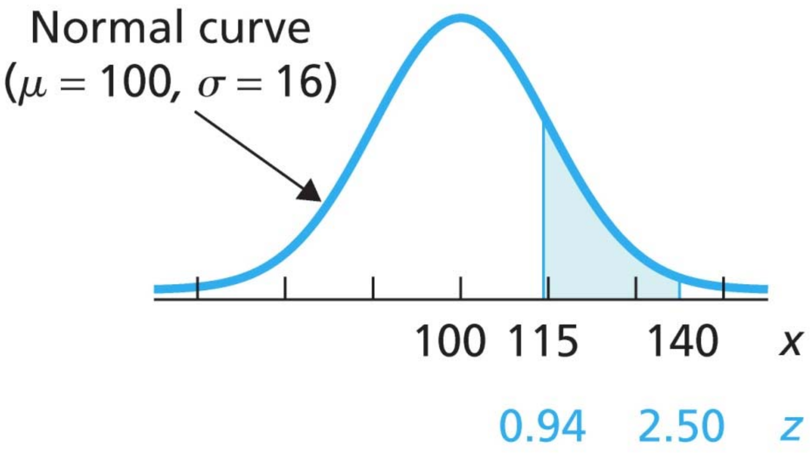
1. State the problem in terms of the observed variable x.
2. Standardize x to restate the problem in terms of a standard Normal variable z. Draw a picture to show the area under the standard Normal curve.
3. Find the required area under the standard Normal curve using Table A and the fact that the total area under the curve is 1.

EXAMPLE 6.9 Percentages for a Normally Distributed Variable

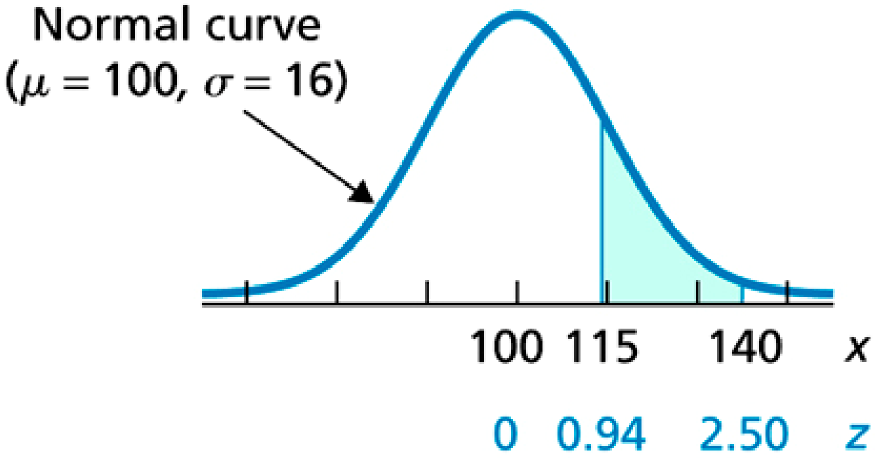
- Intelligence Quotients: Intelligence quotients (IQs) measured on the Stanford Revision of the Binet–Simon Intelligence Scale are normally distributed with a mean of 100 and a standard deviation of 16. Determine the percentage of people who have IQs between 115 and 140.

Figure 6.19

Determination of the percentage of people having IQs between 115 and 140



Example 6.9 Figure 6.19



z-score computations: Area to the left of z:

$$x = 115 \longrightarrow z = \frac{115 - 100}{16} = 0.94 \qquad 0.8264$$

$$x = 140 \longrightarrow z = \frac{140 - 100}{16} = 2.50 \qquad 0.9938$$

Shaded area $= 0.9938 - 0.8264 = 0.1674$

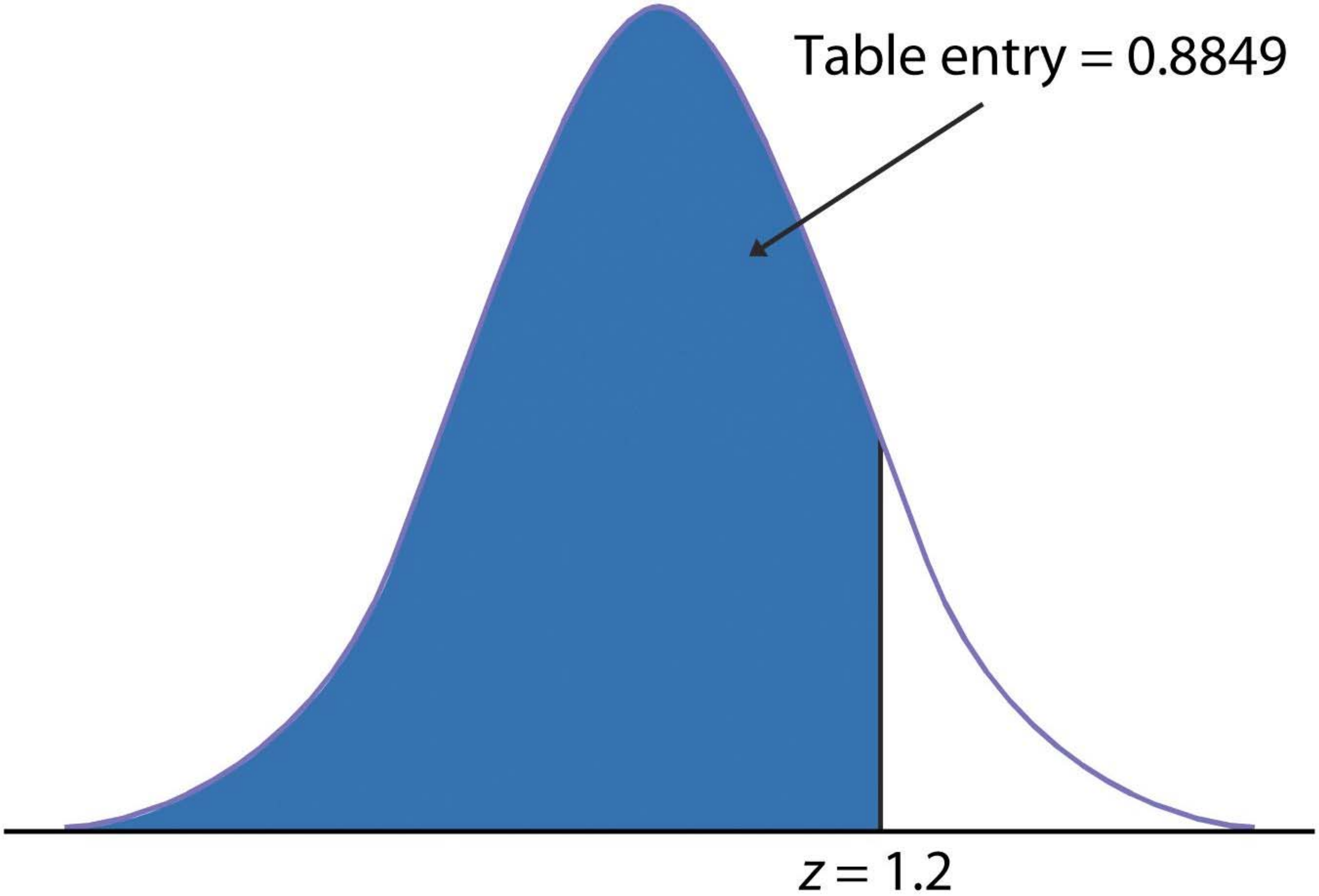
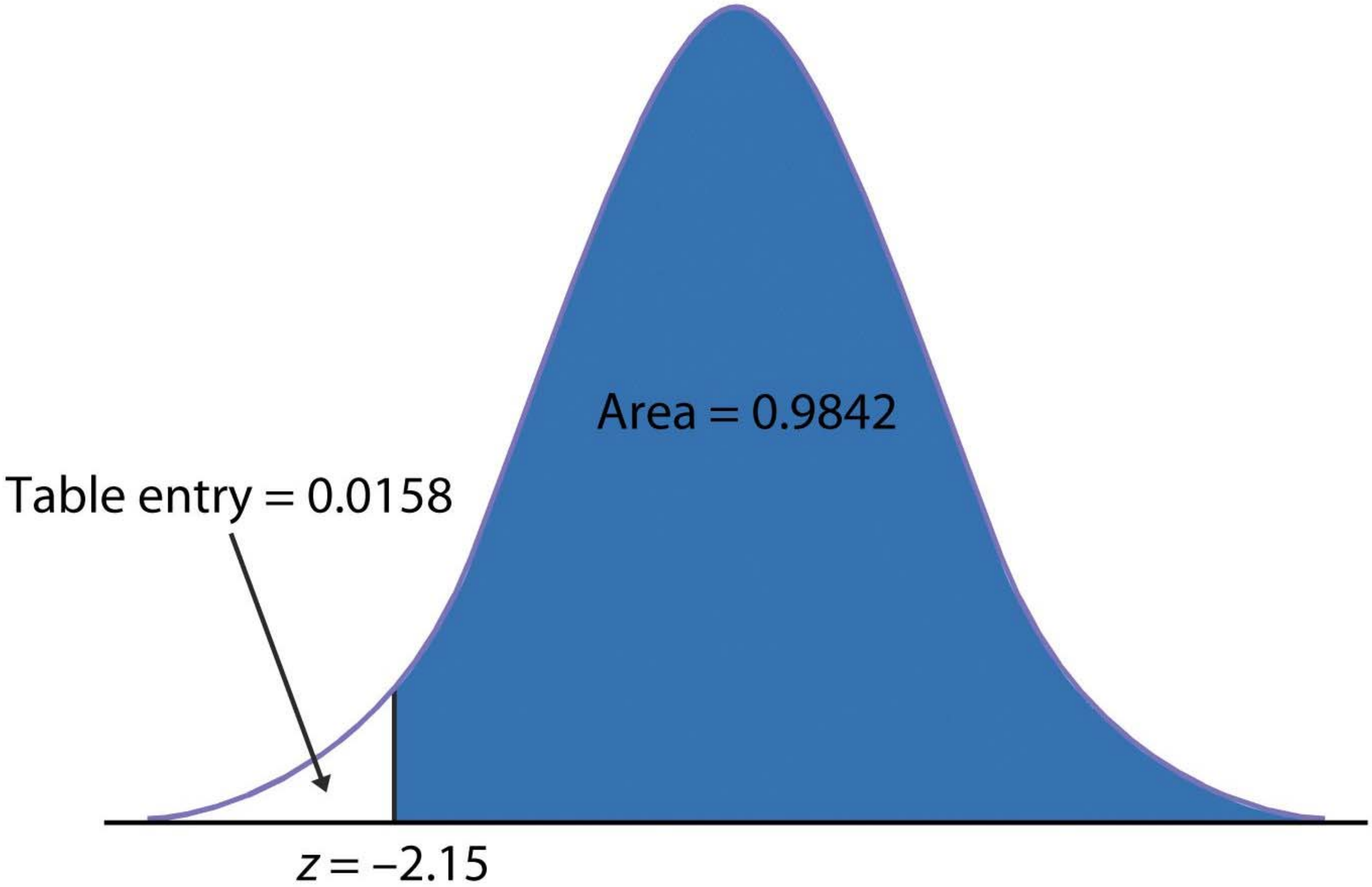


Table entry = 0.8849
$z = 1.2$


Area = 0.9842
Table entry = 0.0158
z = −2.15

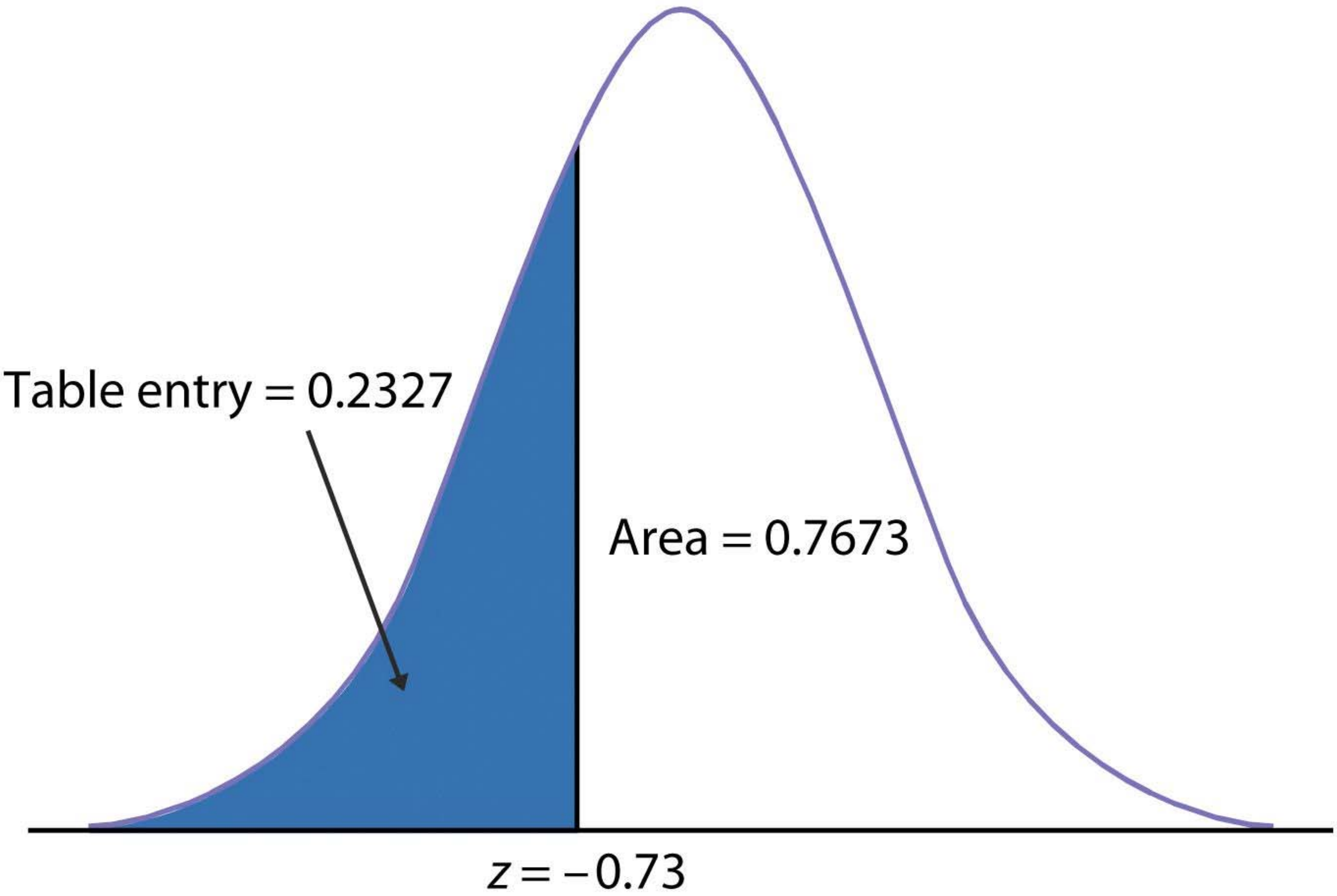
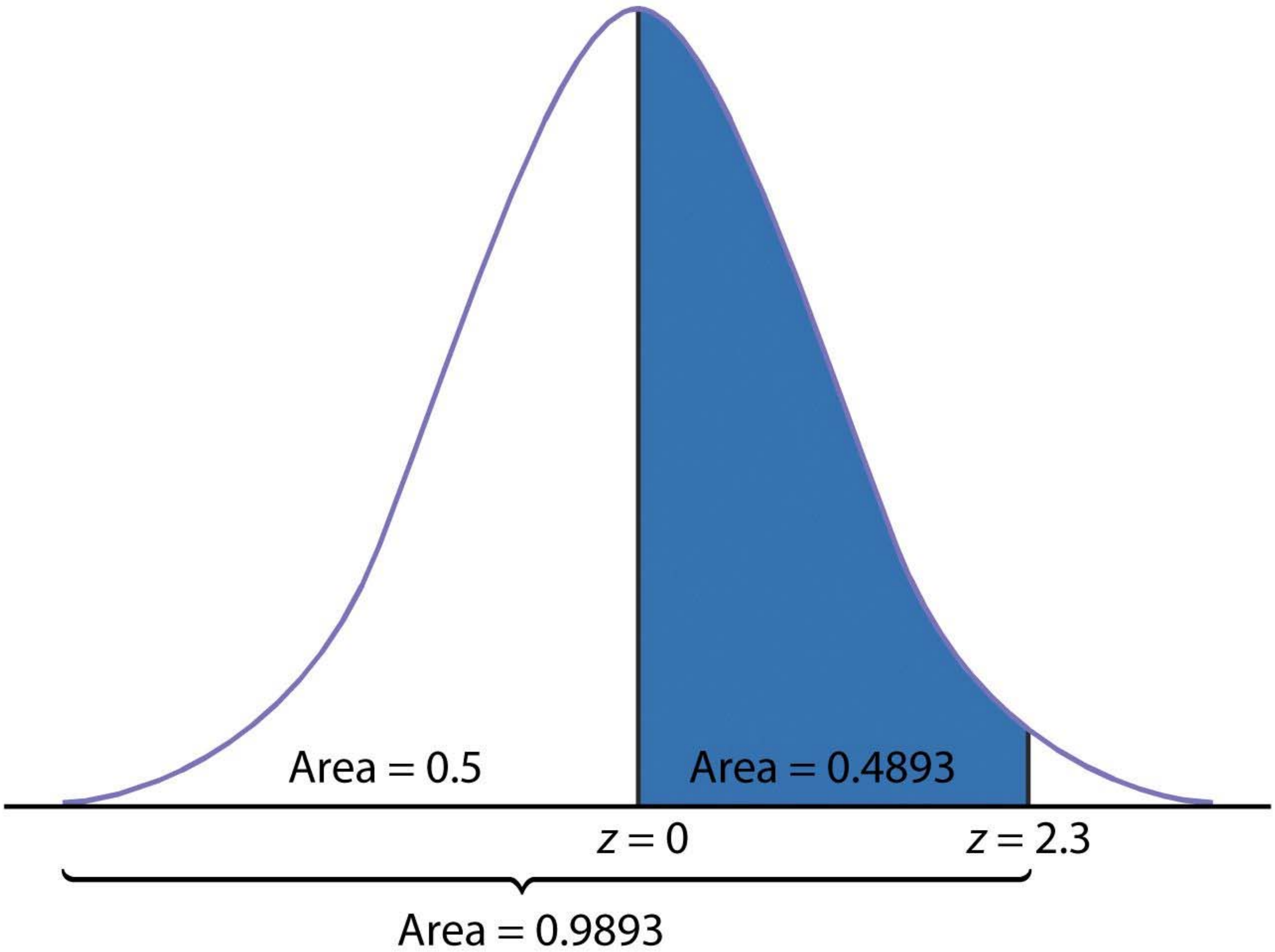


Table entry = 0.2327
Area = 0.7673
$z = -0.73$



Area = 0.5
Area = 0.4893
z = 0
z = 2.3
Area = 0.9893

Key Fact 6.6 & Figure 6.20

The 68.26-95.44-99.74 Rule

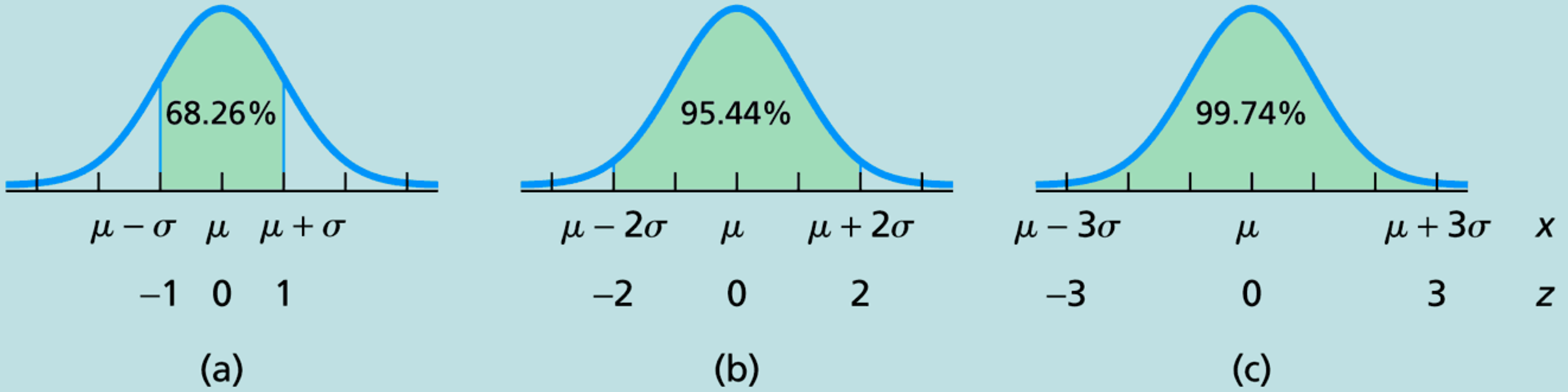
Any normally distributed variable has the following properties.

Property 1: 68.26% of all possible observations lie within one standard deviation to either side of the mean, that is, between $\mu - \sigma$ and $\mu + \sigma$.

Property 2: 95.44% of all possible observations lie within two standard deviations to either side of the mean, that is, between $\mu - 2\sigma$ and $\mu + 2\sigma$.

Property 3: 99.74% of all possible observations lie within three standard deviations to either side of the mean, that is, between $\mu - 3\sigma$ and $\mu + 3\sigma$.

These properties are illustrated in Fig. 6.20.



Procedure 6.2

To Determine the Observations Corresponding to a Specified Percentage or Probability for a Normally Distributed Variable

Step 1 Sketch the normal curve associated with the variable.

Step 2 Shade the region of interest.

Step 3 Use Table II to determine the z-score(s) delimiting the region found in Step 2.

Step 4 Find the x-value(s) having the z-score(s) found in Step 3.

Section 6.4

Assessing Normality; Normal Probability Plots

6.4 Assessing Normality; Normal Probability plots

How do we know if a normal distribution is a reasonable model for data?

Probability plotting is a graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data.

Probability plotting typically uses special graph paper, known as **probability paper,** that has been designed for the hypothesized distribution. Probability paper is widely available for the normal, and various distributions.

Key Fact 6.7

Guidelines for Assessing Normality Using a Normal Probability Plot

To assess the normality of a variable using sample data, construct a normal probability plot.

- If the plot is roughly linear, you can assume that the variable is approximately normally distributed.
- If the plot is not roughly linear, you can assume that the variable is not approximately normally distributed.

These guidelines should be interpreted loosely for small samples but usually strictly for large samples.

EXAMPLE 6.14 Normal Probability Plots

- Adjusted Gross Incomes: The Internal Revenue Service publishes data on federal individual income tax returns in Statistics of Income, Individual Income Tax Returns. A simple random sample of 12 returns from last year revealed the adjusted gross incomes, in thousands of dollars, shown in Table 6.3. Construct a normal probability plot for these data, and use the plot to assess the normality of adjusted gross incomes.

TABLE 6.3

Adjusted gross incomes ($1000s)

9.7	93.1	33.0	21.2
81.4	51.1	43.5	10.6
12.8	7.8	18.1	12.7

Table 6.4

Ordered data and normal scores

Adjusted gross income	Normal score
7.8	−1.64
9.7	−1.11
10.6	−0.79
12.7	−0.53
12.8	−0.31
18.1	−0.10
21.2	0.10
33.0	0.31
43.5	0.53
51.1	0.79
81.4	1.11
93.1	1.64

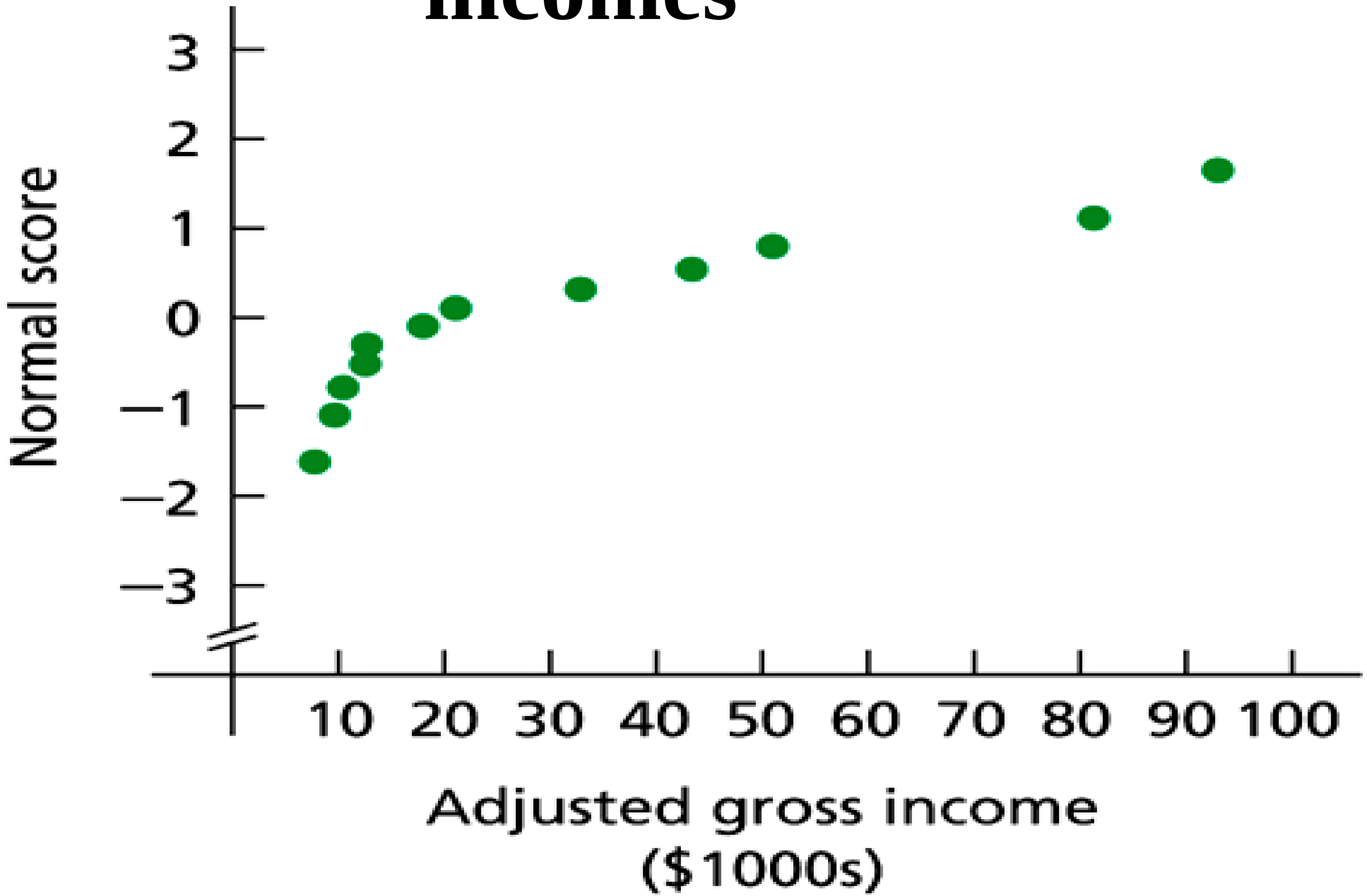
Table 6.4

First arrange the data in increasing order

Adjusted gross income	Normal score
7.8	−1.64
9.7	−1.11
10.6	−0.79
12.7	−0.53
12.8	−0.31
18.1	−0.10
21.2	0.10
33.0	0.31
43.5	0.53
51.1	0.79
81.4	1.11
93.1	1.64

We then obtain the normal scores from table III in Appendix A.

Figure 6.23 **Normal probability plot for the sample of adjusted gross incomes**

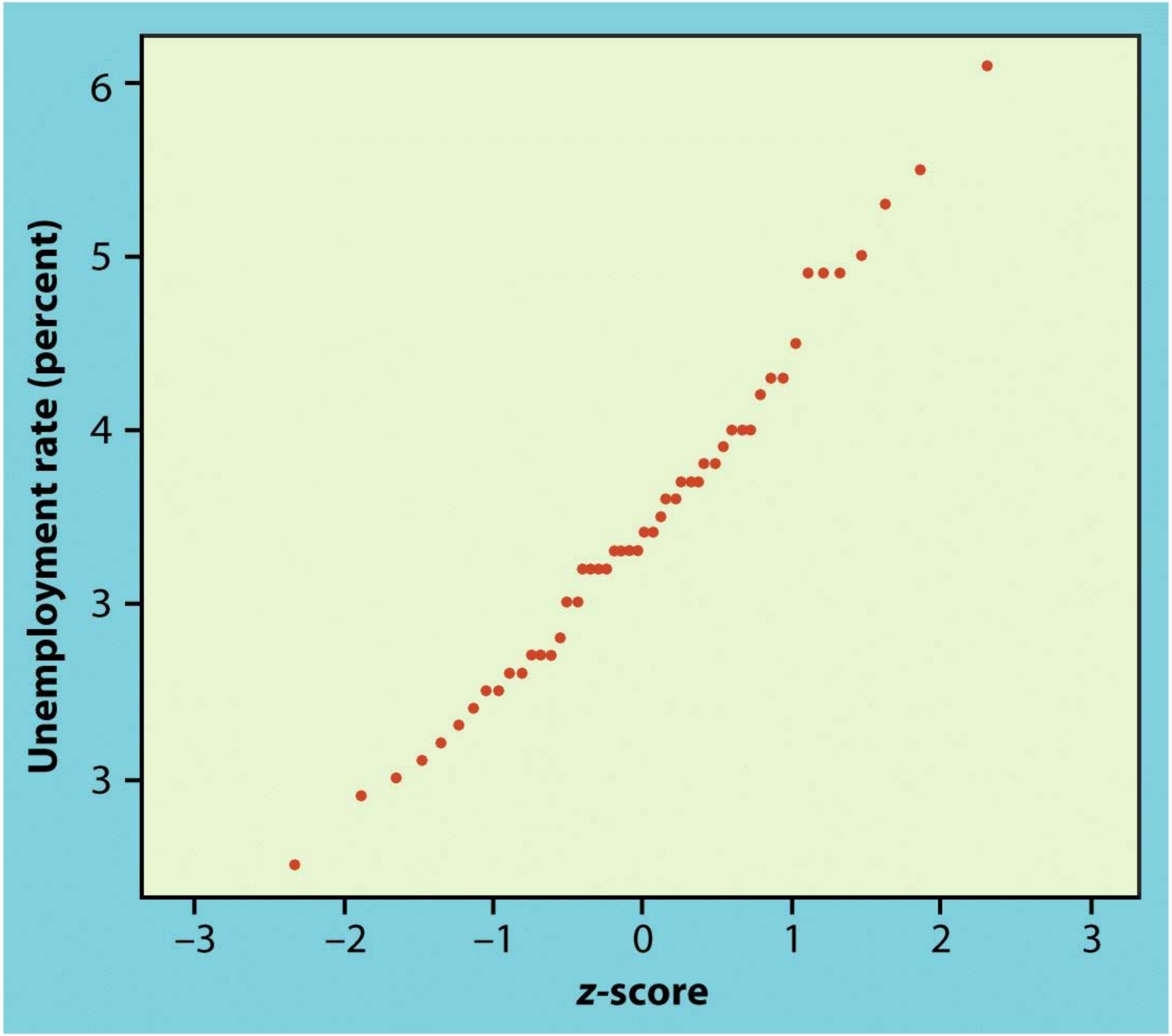


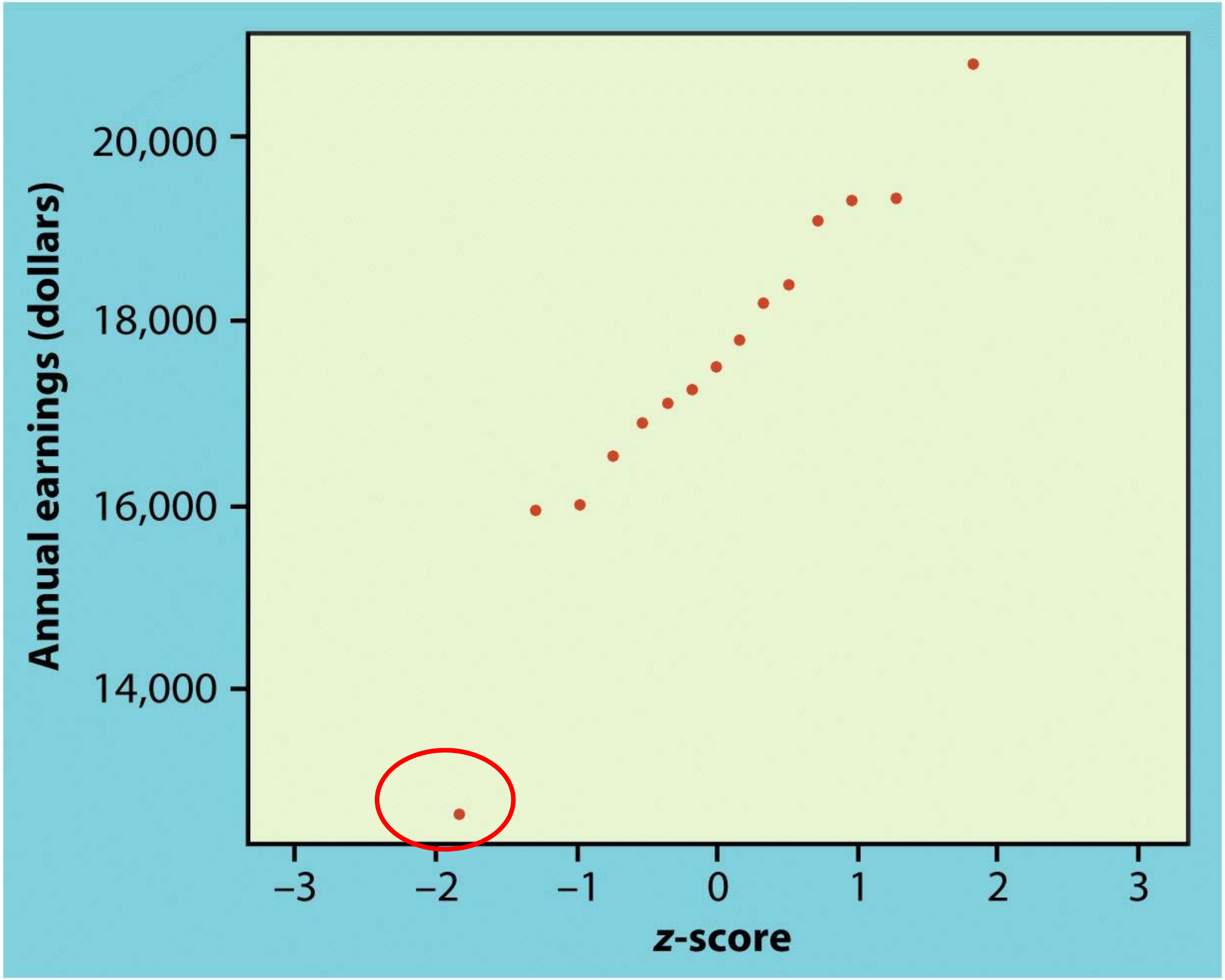
Guidelines for assessing Normality Using a Normal Probability Plot

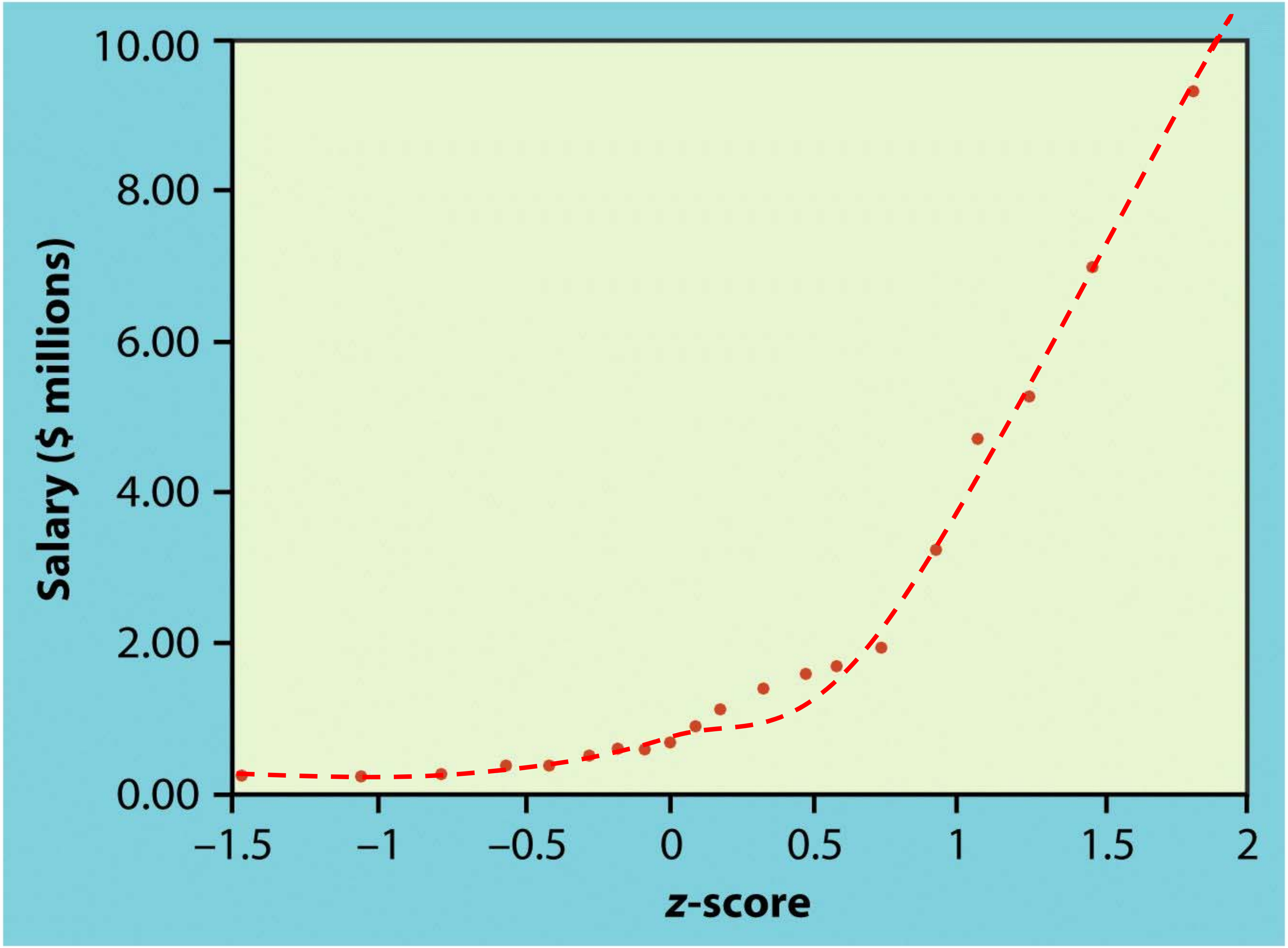
To assess the normality of a variable using sample data, construct a normal probability plot.

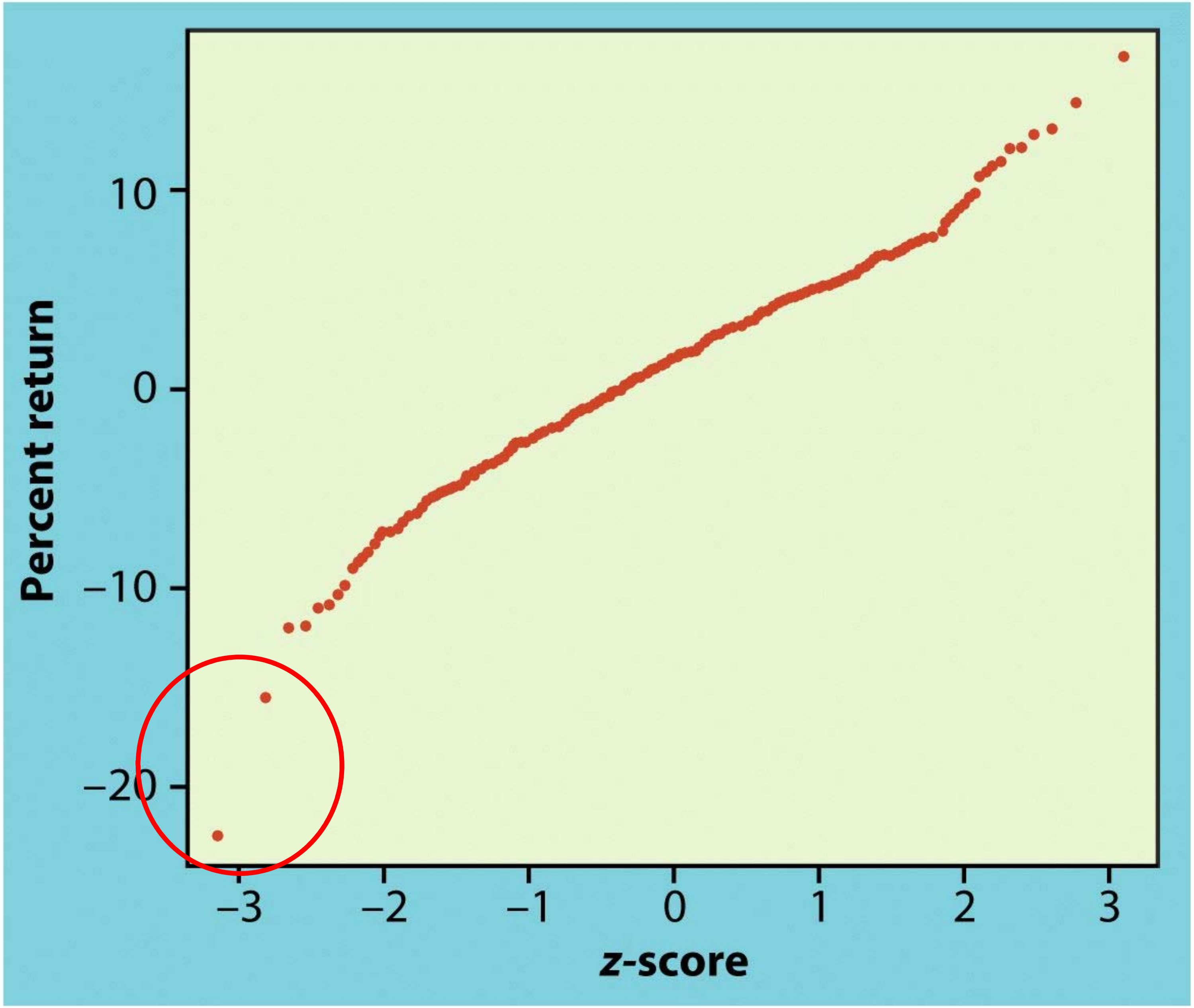
- If the plot is roughly linear, you can assume that the variable is approximately normally distributed.
- If the plot is not roughly linear, you can assume that the variable is not approximately normally distributed.

There guidelines should be interpreted loosely for small samples, but usually strictly for large samples.


6
5
4
3
3
Unemployment rate (percent)
−3
−2
−1
0
1
2
3
z-score


20,000
18,000
16,000
14,000
Annual earnings (dollars)
−3
−2
−1
0
1
2
3
z-score


10.00
8.00
6.00
4.00
2.00
0.00
Salary ($ millions)
−1.5
−1
−0.5
0
0.5
1
1.5
2
z-score


10
0
−10
−20
Percent return
−3
−2
−1
0
1
2
3
z-score

Section 6.5

Normal Approximation to the Binomial Distribution

Procedure 6.3

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

To Approximate Binomial Probabilities by Normal-Curve Areas

Step 1 **Find n, the number of trials, and p, the success probability.**

Step 2 **Continue only if both np and $n(1 - p)$ are 5 or greater.**

Step 3 **Find μ and σ, using the formulas $\mu = np$ and $\sigma = \sqrt{np(1 - p)}$.**

Step 4 **Make the correction for continuity, and find the required area under the normal curve with parameters μ and σ.**

Figure 6.25

Probability histogram for X with superimposed normal curve

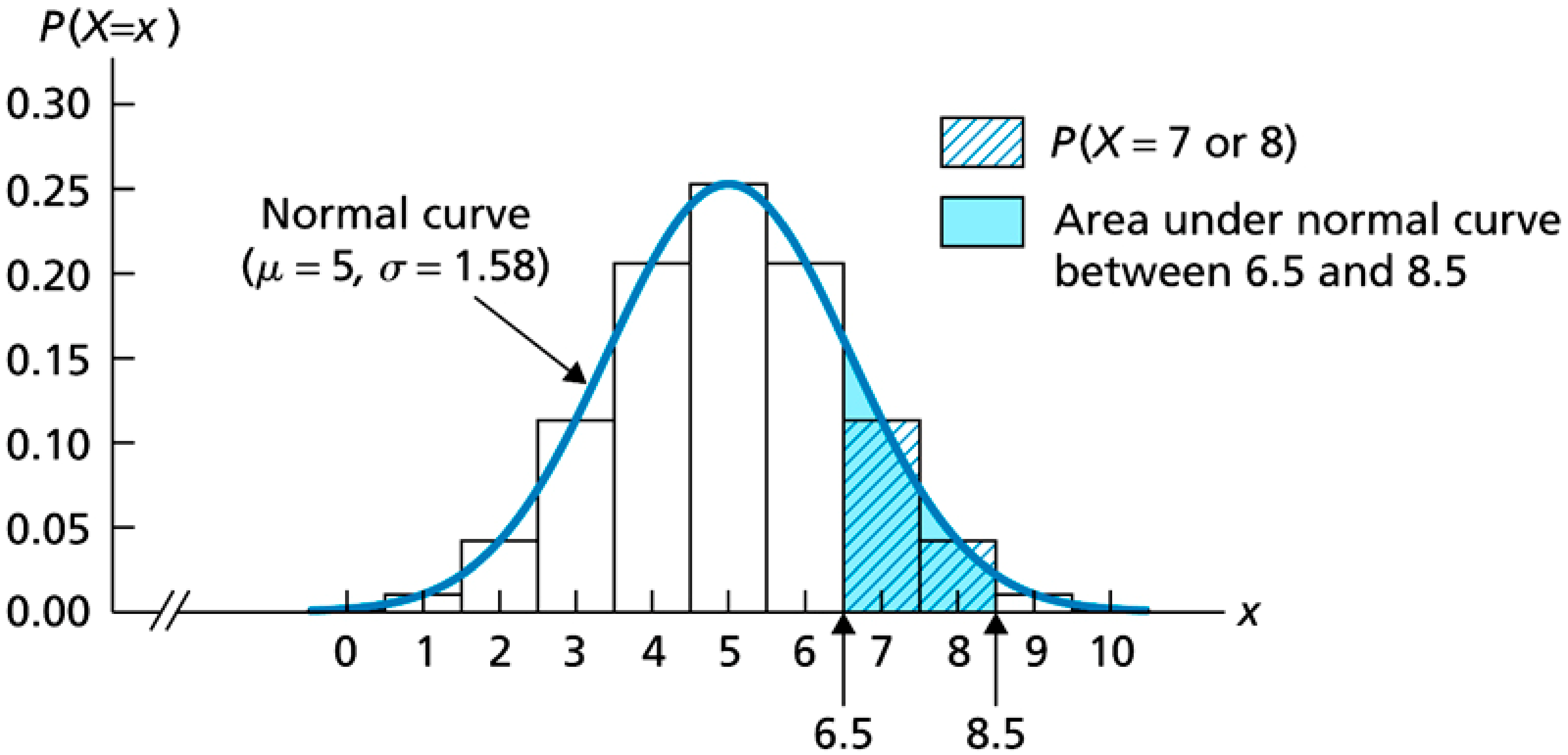


Figure 6.27

Nine different binomial distributions