5 | CONTINUOUS RANDOM VARIABLES

Figure 5.1 The heights of these radish plants are continuous random variables. (Credit: Rev Stan)

Introduction

Chapter Objectives

By the end of this chapter, the student should be able to:

- Recognize and understand continuous probability density functions in general.
- Recognize the uniform probability distribution and apply it appropriately.
- Recognize the exponential probability distribution and apply it appropriately.

Continuous random variables have many applications. Baseball batting averages, IQ scores, the length of time a long distance telephone call lasts, the amount of money a person carries, the length of time a computer chip lasts, and SAT scores are just a few. The field of reliability depends on a variety of continuous random variables.

NOTE

The values of discrete and continuous random variables can be ambiguous. For example, if *X* is equal to the number of miles (to the nearest mile) you drive to work, then *X* is a discrete random variable. You count the miles. If *X* is the distance you drive to work, then you measure values of *X* and *X* is a continuous random variable. For a second example, if *X* is equal to the number of books in a backpack, then *X* is a discrete random variable. If *X* is the weight of a book, then *X* is a continuous random variable because weights are measured. How the random variable is defined is very important.

Properties of Continuous Probability Distributions

The graph of a continuous probability distribution is a curve. Probability is represented by area under the curve.

The curve is called the **probability density function** (abbreviated as **pdf**). We use the symbol $f(x)$ to represent the curve. $f(x)$ is the function that corresponds to the graph; we use the density function $f(x)$ to draw the graph of the probability distribution.

Area under the curve is given by a different function called the **cumulative distribution function** (abbreviated as **cdf**). The cumulative distribution function is used to evaluate probability as area.

- The outcomes are measured, not counted.
- The entire area under the curve and above the x-axis is equal to one.
- Probability is found for intervals of *x* values rather than for individual *x* values.
- $P(c < x < d)$ is the probability that the random variable *X* is in the interval between the values *c* and *d*. $P(c < x < d)$ is the area under the curve, above the *x*-axis, to the right of *c* and the left of *d*.
- $P(x = c) = 0$ The probability that *x* takes on any single individual value is zero. The area below the curve, above the *x*-axis, and between $x = c$ and $x = c$ has no width, and therefore no area (area = 0). Since the probability is equal to the area, the probability is also zero.
- $P(c < x < d)$ is the same as $P(c \le x \le d)$ because probability is equal to area.

We will find the area that represents probability by using geometry, formulas, technology, or probability tables. In general, calculus is needed to find the area under the curve for many probability density functions. When we use formulas to find the area in this textbook, the formulas were found by using the techniques of integral calculus. However, because most students taking this course have not studied calculus, we will not be using calculus in this textbook.

There are many continuous probability distributions. When using a continuous probability distribution to model probability, the distribution used is selected to model and fit the particular situation in the best way.

In this chapter and the next, we will study the uniform distribution, the exponential distribution, and the normal distribution. The following graphs illustrate these distributions.

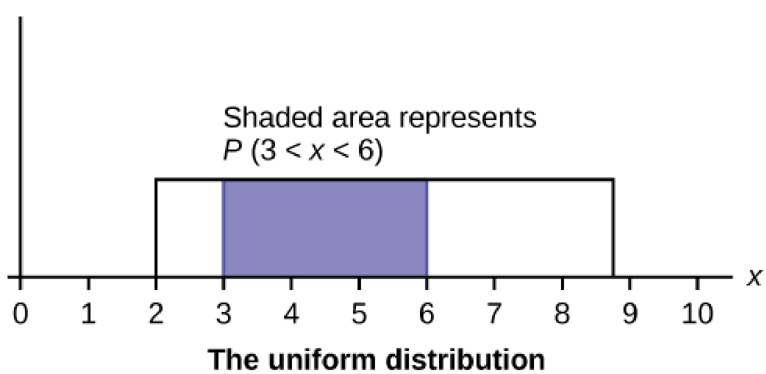

Figure 5.2 The graph shows a Uniform Distribution with the area between $x = 3$ and $x = 6$ shaded to represent the probability that the value of the random variable *X* is in the interval between three and six.

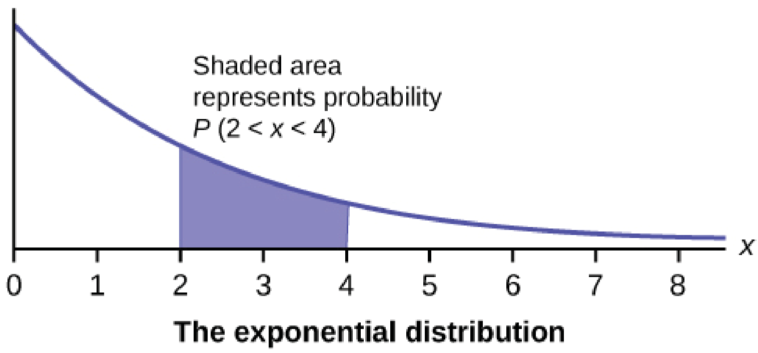

Figure 5.3 The graph shows an Exponential Distribution with the area between $x = 2$ and $x = 4$ shaded to represent the probability that the value of the random variable X is in the interval between two and four.

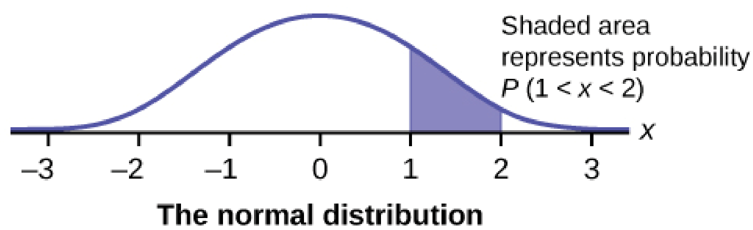

Figure 5.4 The graph shows the Standard Normal Distribution with the area between $x = 1$ and $x = 2$ shaded to represent the probability that the value of the random variable X is in the interval between one and two.

5.1 | Continuous Probability Functions

We begin by defining a continuous probability density function. We use the function notation $f(x)$. Intermediate algebra may have been your first formal introduction to functions. In the study of probability, the functions we study are special. We define the function $f(x)$ so that the area between it and the x-axis is equal to a probability. Since the maximum probability is one, the maximum area is also one. **For continuous probability distributions, PROBABILITY = AREA.**

Example 5.1

Consider the function $f(x) = \frac{1}{20}$ for $0 \le x \le 20$. x = a real number. The graph of $f(x) = \frac{1}{20}$ is a horizontal line. However, since $0 \le x \le 20$, $f(x)$ is restricted to the portion between $x = 0$ and $x = 20$, inclusive.

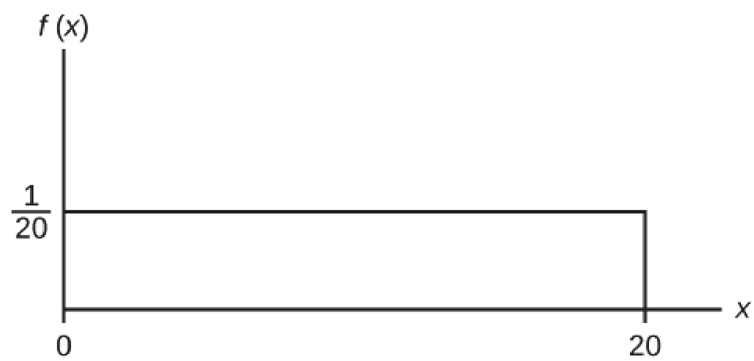

Figure 5.5

$f(x) = \frac{1}{20}$ **for** $0 \le x \le 20$.

The graph of $f(x) = \frac{1}{20}$ is a horizontal line segment when $0 \le x \le 20$.

The area between $f(x) = \frac{1}{20}$ where $0 \le x \le 20$ and the x-axis is the area of a rectangle with base = 20 and height $= \frac{1}{20}$.

$$\text{AREA} = 20\left(\frac{1}{20}\right) = 1$$

Suppose we want to find the area between $f(x) = \frac{1}{20}$ **and the** x**-axis where** $0 < x < 2$**.**

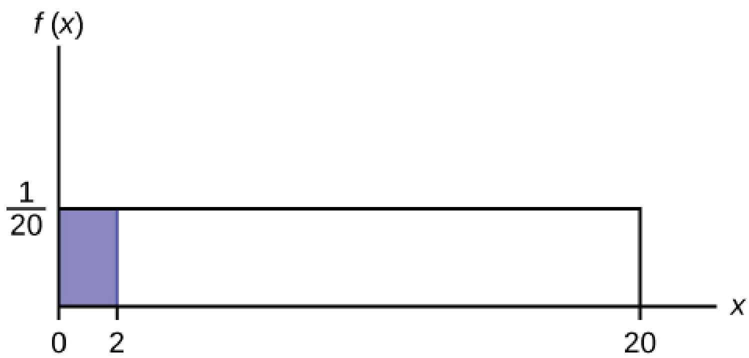

Figure 5.6

$$\text{AREA} = (2 - 0)\left(\frac{1}{20}\right) = 0.1$$

$(2 - 0) = 2 =$ base of a rectangle

REMINDER

area of a rectangle = (base)(height).

The area corresponds to a probability. The probability that x is between zero and two is 0.1, which can be written mathematically as $P(0 < x < 2) = P(x < 2) = 0.1$.

Suppose we want to find the area between $f(x) = \frac{1}{20}$ **and the** x**-axis where** $4 < x < 15$**.**

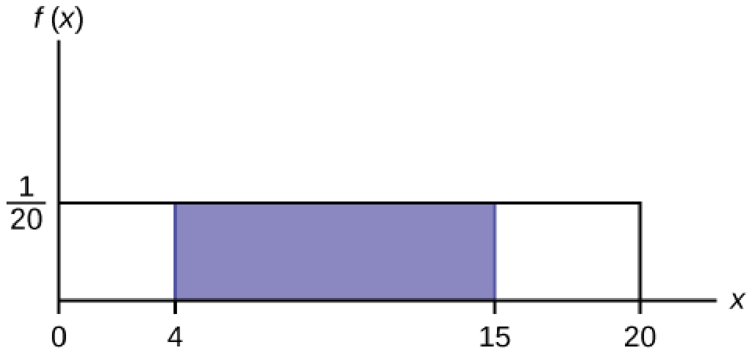

Figure 5.7

$$\text{AREA} = (15 - 4)\left(\frac{1}{20}\right) = 0.55$$

$$\text{AREA} = (15 - 4)\left(\frac{1}{20}\right) = 0.55$$

$$(15 - 4) = 11 = \text{the base of a rectangle}$$

The area corresponds to the probability $P(4 < x < 15) = 0.55$.

Suppose we want to find $P(x = 15)$. On an x-y graph, $x = 15$ is a vertical line. A vertical line has no width (or zero width). Therefore, $P(x = 15) = (\text{base})(\text{height}) = (0)\left(\frac{1}{20}\right) = 0$

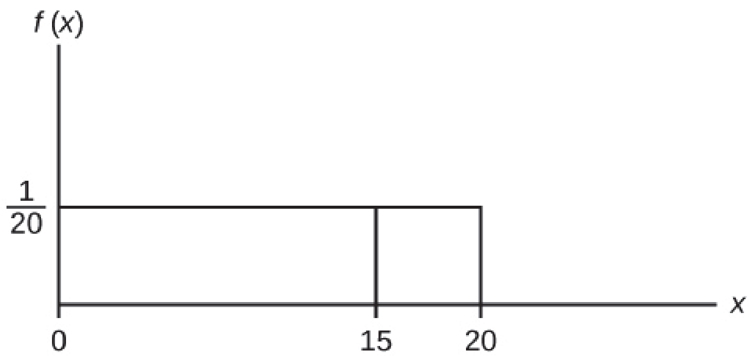

Figure 5.8

$P(X \leq x)$ (can be written as $P(X < x)$ for continuous distributions) is called the cumulative distribution function or CDF. Notice the "less than or equal to" symbol. We can use the CDF to calculate $P(X > x)$. The CDF gives "area to the left" and $P(X > x)$ gives "area to the right." We calculate $P(X > x)$ for continuous distributions as follows: $P(X > x) = 1 – P(X < x)$.

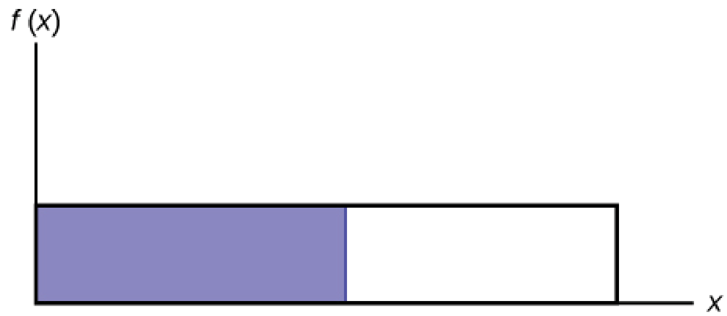

Figure 5.9

Label the graph with $f(x)$ and x. Scale the x and y axes with the maximum x and y values. $f(x) = \frac{1}{20}$, $0 \le x \le 20$.

To calculate the probability that x is between two values, look at the following graph. Shade the region between $x = 2.3$ and $x = 12.7$. Then calculate the shaded area of a rectangle.

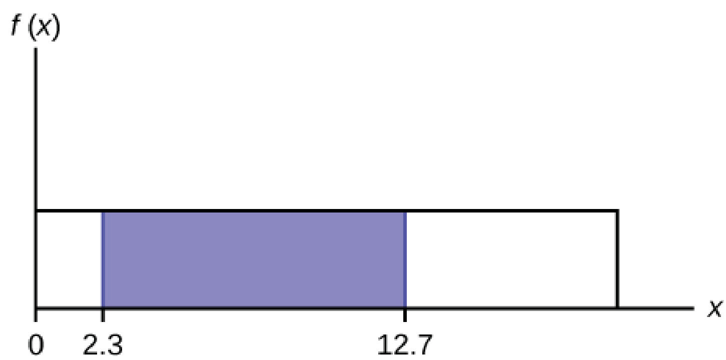

Figure 5.10

$$P(2.3 < x < 12.7) = (\text{base})(\text{height}) = (12.7 - 2.3)\left(\frac{1}{20}\right) = 0.52$$

Try It Σ

5.1 Consider the function $f(x) = \frac{1}{8}$ for $0 \le x \le 8$. Draw the graph of $f(x)$ and find $P(2.5 < x < 7.5)$.

5.2 | The Uniform Distribution

The uniform distribution is a continuous probability distribution and is concerned with events that are equally likely to occur. When working out problems that have a uniform distribution, be careful to note if the data is inclusive or exclusive.

Example 5.2

The data in **Table 5.1** are 55 smiling times, in seconds, of an eight-week-old baby.

6 | THE NORMAL DISTRIBUTION

Figure 6.1 If you ask enough people about their shoe size, you will find that your graphed data is shaped like a bell curve and can be described as normally distributed. (credit: Ömer Ünlü)

Introduction

Chapter Objectives

By the end of this chapter, the student should be able to:

- Recognize the normal probability distribution and apply it appropriately.
- Recognize the standard normal probability distribution and apply it appropriately.
- Compare normal probabilities by converting to the standard normal distribution.

The normal, a continuous distribution, is the most important of all the distributions. It is widely used and even more widely abused. Its graph is bell-shaped. You see the bell curve in almost all disciplines. Some of these include psychology, business, economics, the sciences, nursing, and, of course, mathematics. Some of your instructors may use the normal distribution to help determine your grade. Most IQ scores are normally distributed. Often real-estate prices fit a normal distribution. The normal distribution is extremely important, but it cannot be applied to everything in the real world.

In this chapter, you will study the normal distribution, the standard normal distribution, and applications associated with them.

The normal distribution has two parameters (two numerical descriptive measures), the mean (μ) and the standard deviation (σ). If X is a quantity to be measured that has a normal distribution with mean (μ) and standard deviation (σ), we designate this by writing

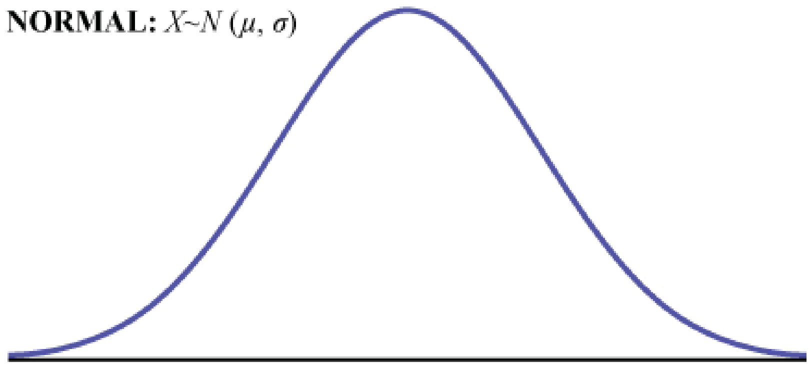

Figure 6.2

The probability density function is a rather complicated function. **Do not memorize it**. It is not necessary.

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

The cumulative distribution function is $P(X < x)$. It is calculated either by a calculator or a computer, or it is looked up in a table. Technology has made the tables virtually obsolete. For that reason, as well as the fact that there are various table formats, we are not including table instructions.

The curve is symmetrical about a vertical line drawn through the mean, μ. In theory, the mean is the same as the median, because the graph is symmetric about μ. As the notation indicates, the normal distribution depends only on the mean and the standard deviation. Since the area under the curve must equal one, a change in the standard deviation, σ, causes a change in the shape of the curve; the curve becomes fatter or skinnier depending on σ. A change in μ causes the graph to shift to the left or right. This means there are an infinite number of normal probability distributions. One of special interest is called the **standard normal distribution**.

Collaborative Exercise

Your instructor will record the heights of both men and women in your class, separately. Draw histograms of your data. Then draw a smooth curve through each histogram. Is each curve somewhat bell-shaped? Do you think that if you had recorded 200 data values for men and 200 for women that the curves would look bell-shaped? Calculate the mean for each data set. Write the means on the x-axis of the appropriate graph below the peak. Shade the approximate area that represents the probability that one randomly chosen male is taller than 72 inches. Shade the approximate area that represents the probability that one randomly chosen female is shorter than 60 inches. If the total area under each curve is one, does either probability appear to be more than 0.5?

6.1 | The Standard Normal Distribution

The **standard normal distribution** is a normal distribution of **standardized values called *z*-scores**. **A *z*-score is measured in units of the standard deviation.** For example, if the mean of a normal distribution is five and the standard deviation is two, the value 11 is three standard deviations above (or to the right of) the mean. The calculation is as follows:

$x = \mu + (z)(\sigma) = 5 + (3)(2) = 11$

The z-score is three.

The mean for the standard normal distribution is zero, and the standard deviation is one. The transformation $z = \frac{x - \mu}{\sigma}$ produces the distribution $Z \sim N(0, 1)$. The value x comes from a normal distribution with mean μ and standard deviation σ.

Z-Scores

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the *z*-score is:

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

The *z*-score tells you how many standard deviations the value *x* is above (to the right of) or below (to the left of) the mean, μ. Values of x that are larger than the mean have positive *z*-scores, and values of x that are smaller than the mean have negative *z*-scores. If x equals the mean, then x has a *z*-score of zero.

Example 6.1

Suppose $X \sim N(5, 6)$. This says that x is a normally distributed random variable with mean $\mu = 5$ and standard deviation $\sigma = 6$. Suppose $x = 17$. Then:

$$z = \frac{x - \mu}{\sigma} = \frac{17 - 5}{6} = 2$$

This means that $x = 17$ is **two standard deviations** (2σ) above or to the right of the mean $\mu = 5$. The standard deviation is $\sigma = 6$.

Notice that: $5 + (2)(6) = 17$ (The pattern is $\mu + z\sigma = x$)

Now suppose $x = 1$. Then: $z = \frac{x - \mu}{\sigma} = \frac{1 - 5}{6} = -0.67$ (rounded to two decimal places)

This means that $x = 1$ is 0.67 standard deviations (-0.67σ) below or to the left of the mean $\mu = 5$. Notice that: $5 + (-0.67)(6)$ is approximately equal to one (This has the pattern $\mu + (-0.67)\sigma = 1$)

Summarizing, when z is positive, x is above or to the right of μ and when z is negative, x is to the left of or below μ. Or, when z is positive, x is greater than μ, and when z is negative x is less than μ.

Try It Σ

6.1 What is the *z*-score of x, when $x = 1$ and $X \sim N(12,3)$?

Example 6.2

Some doctors believe that a person can lose five pounds, on the average, in a month by reducing his or her fat intake and by exercising consistently. Suppose weight loss has a normal distribution. Let X = the amount of weight lost(in pounds) by a person in a month. Use a standard deviation of two pounds. $X \sim N(5, 2)$. Fill in the blanks.

a. Suppose a person **lost** ten pounds in a month. The *z*-score when $x = 10$ pounds is $z = 2.5$ (verify). This *z*-score tells you that $x = 10$ is ________ standard deviations to the ________ (right or left) of the mean _____ (What is the mean?).

Solution 6.2

a. This *z*-score tells you that $x = 10$ is **2.5** standard deviations to the **right** of the mean **five**.

b. Suppose a person **gained** three pounds (a negative weight loss). Then z = __________. This *z*-score tells you that $x = -3$ is ________ standard deviations to the __________ (right or left) of the mean.

Solution 6.2

b. $z = \mathbf{-4}$. This *z*-score tells you that $x = -3$ is **four** standard deviations to the **left** of the mean.

Suppose the random variables X and Y have the following normal distributions: $X \sim N(5, 6)$ and $Y \sim N(2, 1)$. If $x = 17$, then $z = 2$. (This was previously shown.) If $y = 4$, what is z?

$$z = \frac{y - \mu}{\sigma} = \frac{4 - 2}{1} = 2 \text{ where } \mu = 2 \text{ and } \sigma = 1.$$

The z-score for $y = 4$ is $z = 2$. This means that four is $z = 2$ standard deviations to the right of the mean. Therefore, $x = 17$ and $y = 4$ are both two (of **their own**) standard deviations to the right of **their** respective means.

The *z*-score allows us to compare data that are scaled differently. To understand the concept, suppose $X \sim N(5, 6)$ represents weight gains for one group of people who are trying to gain weight in a six week period and $Y \sim N(2, 1)$ measures the same weight gain for a second group of people. A negative weight gain would be a weight loss. Since $x = 17$ and $y = 4$ are each two standard deviations to the right of their means, they represent the same, standardized weight gain **relative to their means**.

Try It Σ

6.2 Fill in the blanks.

Jerome averages 16 points a game with a standard deviation of four points. $X \sim N(16,4)$. Suppose Jerome scores ten points in a game. The z–score when $x = 10$ is –1.5. This score tells you that $x = 10$ is _____ standard deviations to the ______(right or left) of the mean______(What is the mean?).

The Empirical Rule

If X is a random variable and has a normal distribution with mean μ and standard deviation σ, then the **Empirical Rule** says the following:

- About 68% of the x values lie between $–1\sigma$ and $+1\sigma$ of the mean μ (within one standard deviation of the mean).
- About 95% of the x values lie between $–2\sigma$ and $+2\sigma$ of the mean μ (within two standard deviations of the mean).
- About 99.7% of the x values lie between $–3\sigma$ and $+3\sigma$ of the mean μ (within three standard deviations of the mean). Notice that almost all the x values lie within three standard deviations of the mean.
- The z-scores for $+1\sigma$ and $–1\sigma$ are +1 and –1, respectively.
- The z-scores for $+2\sigma$ and $–2\sigma$ are +2 and –2, respectively.
- The z-scores for $+3\sigma$ and $–3\sigma$ are +3 and –3 respectively.

The empirical rule is also known as the 68-95-99.7 rule.

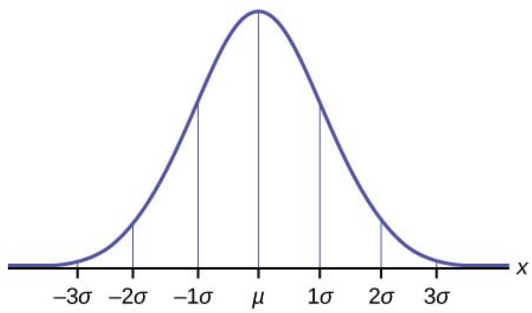

Figure 6.3

Example 6.3

The mean height of 15 to 18-year-old males from Chile from 2009 to 2010 was 170 cm with a standard deviation of 6.28 cm. Male heights are known to follow a normal distribution. Let X = the height of a 15 to 18-year-old male from Chile in 2009 to 2010. Then $X \sim N(170, 6.28)$.

a. Suppose a 15 to 18-year-old male from Chile was 168 cm tall from 2009 to 2010. The *z*-score when x = 168 cm is z = _______. This *z*-score tells you that x = 168 is ________ standard deviations to the ________ (right or left) of the mean _____ (What is the mean?).

Solution 6.3

a. –0.32, 0.32, left, 170

b. Suppose that the height of a 15 to 18-year-old male from Chile from 2009 to 2010 has a *z*-score of z = 1.27. What is the male's height? The *z*-score (z = 1.27) tells you that the male's height is ________ standard deviations to the __________ (right or left) of the mean.

Solution 6.3

b. 177.98, 1.27, right

Try It Σ

6.3 Use the information in **Example 6.3** to answer the following questions.

a. Suppose a 15 to 18-year-old male from Chile was 176 cm tall from 2009 to 2010. The *z*-score when x = 176 cm is z = _______. This *z*-score tells you that x = 176 cm is ________ standard deviations to the ________ (right or left) of the mean _____ (What is the mean?).

b. Suppose that the height of a 15 to 18-year-old male from Chile from 2009 to 2010 has a *z*-score of z = –2. What is the male's height? The *z*-score (z = –2) tells you that the male's height is ________ standard deviations to the __________ (right or left) of the mean.

Example 6.4

From 1984 to 1985, the mean height of 15 to 18-year-old males from Chile was 172.36 cm, and the standard deviation was 6.34 cm. Let Y = the height of 15 to 18-year-old males from 1984 to 1985. Then $Y \sim N(172.36, 6.34)$.

The mean height of 15 to 18-year-old males from Chile from 2009 to 2010 was 170 cm with a standard deviation of 6.28 cm. Male heights are known to follow a normal distribution. Let X = the height of a 15 to 18-year-old male from Chile in 2009 to 2010. Then $X \sim N(170, 6.28)$.

Find the *z*-scores for x = 160.58 cm and y = 162.85 cm. Interpret each *z*-score. What can you say about x = 160.58 cm and y = 162.85 cm?

Solution 6.4

The *z*-score for x = 160.58 is z = –1.5.
The *z*-score for y = 162.85 is z = –1.5.
Both x = 160.58 and y = 162.85 deviate the same number of standard deviations from their respective means and in the same direction.

Try It Σ

6.4 In 2012, 1,664,479 students took the SAT exam. The distribution of scores in the verbal section of the SAT had a mean μ = 496 and a standard deviation σ = 114. Let X = a SAT exam verbal section score in 2012. Then $X \sim N(496, 114)$.

Find the *z*-scores for x_1 = 325 and x_2 = 366.21. Interpret each *z*-score. What can you say about x_1 = 325 and x_2 = 366.21?

Example 6.5

Suppose x has a normal distribution with mean 50 and standard deviation 6.

- About 68% of the x values lie between $-1\sigma = (-1)(6) = -6$ and $1\sigma = (1)(6) = 6$ of the mean 50. The values $50 - 6 = 44$ and $50 + 6 = 56$ are within one standard deviation of the mean 50. The z-scores are –1 and +1 for 44 and 56, respectively.
- About 95% of the x values lie between $-2\sigma = (-2)(6) = -12$ and $2\sigma = (2)(6) = 12$. The values $50 - 12 = 38$ and $50 + 12 = 62$ are within two standard deviations of the mean 50. The z-scores are –2 and +2 for 38 and 62, respectively.
- About 99.7% of the x values lie between $-3\sigma = (-3)(6) = -18$ and $3\sigma = (3)(6) = 18$ of the mean 50. The values $50 - 18 = 32$ and $50 + 18 = 68$ are within three standard deviations of the mean 50. The z-scores are –3 and +3 for 32 and 68, respectively.

Try It Σ

6.5 Suppose X has a normal distribution with mean 25 and standard deviation five. Between what values of x do 68% of the values lie?

Example 6.6

From 1984 to 1985, the mean height of 15 to 18-year-old males from Chile was 172.36 cm, and the standard deviation was 6.34 cm. Let Y = the height of 15 to 18-year-old males in 1984 to 1985. Then $Y \sim N(172.36, 6.34)$.

a. About 68% of the y values lie between what two values? These values are ________________. The z-scores are ________________, respectively.

b. About 95% of the y values lie between what two values? These values are ________________. The z-scores are ________________ respectively.

c. About 99.7% of the y values lie between what two values? These values are ________________. The z-scores are ________________, respectively.

Solution 6.6

a. About 68% of the values lie between 166.02 and 178.7. The z-scores are –1 and 1.

b. About 95% of the values lie between 159.68 and 185.04. The z-scores are –2 and 2.

c. About 99.7% of the values lie between 153.34 and 191.38. The z-scores are –3 and 3.

Try It Σ

6.6 The scores on a college entrance exam have an approximate normal distribution with mean, $\mu = 52$ points and a standard deviation, $\sigma = 11$ points.

a. About 68% of the y values lie between what two values? These values are ________________. The z-scores are ________________, respectively.

b. About 95% of the y values lie between what two values? These values are ________________. The z-scores are ________________, respectively.

c. About 99.7% of the y values lie between what two values? These values are ________________. The z-scores are ________________, respectively.