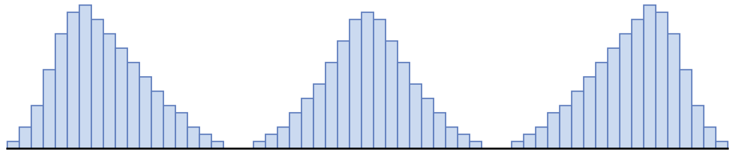
3.4 Normal Distribution

In Section 3.1, you learned about histograms that were skewed left or right and histograms that were symmetrical. While no situation in the real world is perfect, many natural relationships, when displayed as a histogram, will form a bell-shaped distribution like the centre image below.



In this section, you will learn about **normal distribution**, a symmetrical, bell-shaped histogram with a number of significant statistical properties.

normal distribution—a symmetrical, bell-shaped histogram used in statistical analysis

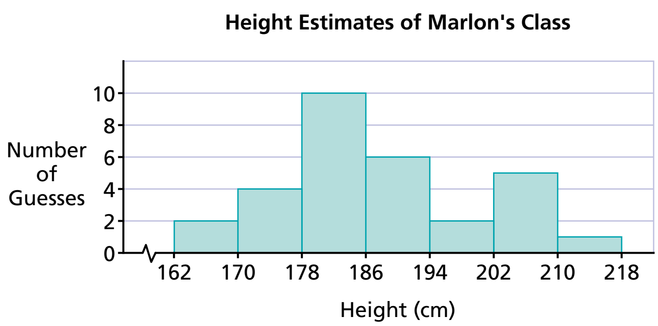
Example 1 Exploring Height

Marlon's class has been challenged to guess their teacher's height in centimetres. Listed below are the estimates, submitted anonymously. Calculate the mean and standard deviation, and then create a histogram.

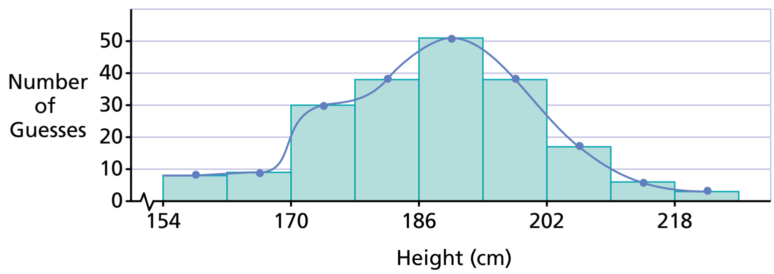
183	183	174	212	178	207	186	178
204	172	189	183	184	190	184	168
190	180	183	190	185	162	200	206
196	187	204	185	206	175		

Solution

Using a calculator, Marlon determines that the mean, $\bar{x}$, is 187, and the standard deviation, σ, is 12.1. Marlon would like seven intervals in his histogram, so he rounds up the range to 56 and gets a bin width of $\frac{56}{7} = 8$.



Marlon decides to collect 120 more samples and the histogram becomes the following:



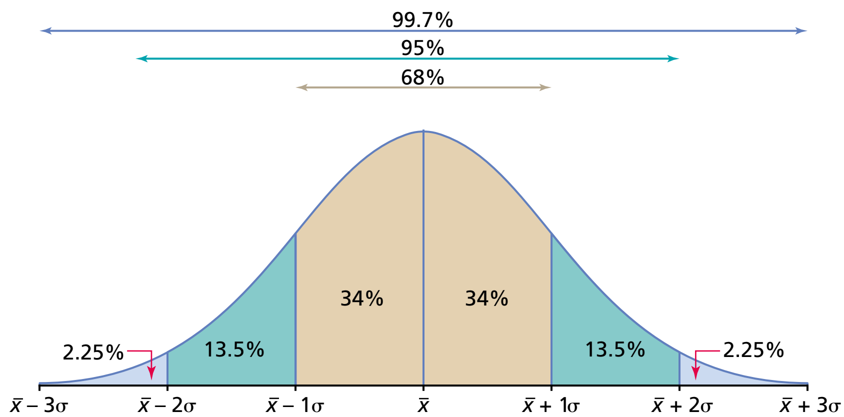
If you draw a smooth curve close to or through the tops of the rectangles in the histogram, you get a curve that looks like a normal curve, like the one shown below.

Given enough data and small enough intervals, Marlon would eventually get a perfectly symmetrical bell-shaped curve. A distribution with a histogram that follows a normal curve is called a normal distribution.

CHARACTERISTICS OF NORMAL DISTRIBUTIONS

A normal distribution has the following properties:

- It is symmetrical; the mean, median, and mode are equal and fall at the line of symmetry for the curve.
- It is shaped like a bell, peaking in the middle and sloping down toward the sides. It approaches zero at the extremes.
- Approximately 68% of the data is within one standard deviation of the mean.
- Approximately 95% of the data is within two standard deviations of the mean.
- Approximately 99.7% of the data is within three standard deviations of the mean.
- The notation used to describe a normal distribution, of the variable X, is $X \sim N(\bar{x}, \sigma^2)$, where $\bar{x}$ is the mean and σ^2 is the variance (the square of the standard deviation).



The graph of the normal distribution $X \sim N(\bar{x}, \sigma^2)$

Example 2 Using Normal Distribution

Julie is an engineer who designs roller coasters. She wants to develop a ride that 95% of the population can ride. The average adult in North America has a mass of 71.8 kg, with a standard deviation of 13.6 kg.

(a) What range of masses should she be prepared to anticipate?

(b) If she wanted to provide for 99.7% of the general population, what range of masses should she be prepared to anticipate?

(c) What assumption is Julie making in this example that could cause problems?

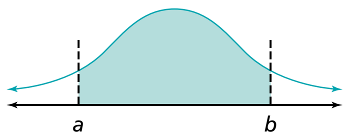
Solution

(a) In a normal distribution, 95% of the data is within two standard deviations of the mean. With a mean of 71.8 and a standard deviation of 13.6, that means that 95% of the data will likely be between $71.8 - 2(13.6)$, or 44.6 kg and $71.8 + 2(13.6)$, or 99 kg.

(b) To get an interval into which 99.7% of the data fits, you need to widen it to three standard deviations from the mean (the interval 31 kg to 112.6 kg).

(c) Julie has assumed that adult masses are normally distributed about the mean, which is unlikely due to the large difference between men and women. To use the properties of normal distribution, Julie needs to separate male and female data, and then calculate the new mean and standard deviation for each population. She could then determine the appropriate intervals.

The area under every normal curve equals 1. In any normal distribution, the percent of the data that lies between two specific values, *a* and *b*, is the area under the normal curve between endpoints *a* and *b*.



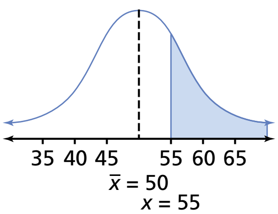
Example 3 Area Under a Normal Curve

If X~(50, 5^2), draw a diagram that represents the percent of data that have these values for *X*:

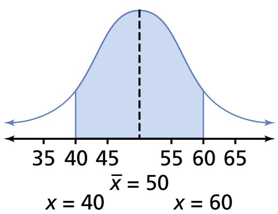
(a) $x > 55$ **(b)** $40 < x < 60$ **(c)** $x < 38$

Solution

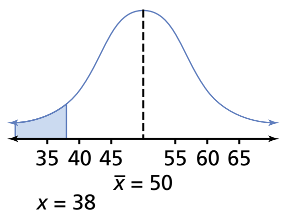
(a)



(b)



(c)



KEY IDEAS

Data that are normally distributed have the following properties:

- Normal distributions are symmetrical and approach zero at the extremes.
- Of the data, 68% is within one standard deviation of the mean; 95% is within two standard deviations of the mean; and 99.7% is within three standard deviations of the mean.
- The mean, median, and mode are equal and fall on the line of symmetry of the distribution.
- The notation used to describe a normal distribution, of the variable X, is $X \sim N(\bar{x}, \sigma^2)$, where $\bar{x}$ is the mean and σ^2 is the variance (the square of the standard deviation).
- The area under any normal curve is 1. The percent of the data that lies between two values in a normal distribution is equivalent to the area under the normal curve between these values.

3.4 Exercises

A

1. **Knowledge and Understanding** Which of the following statements are properties of a normal distribution? Explain.

(a) It is symmetrical about the mean.
(b) Exactly half of the values fall within one standard deviation of the mean.
(c) As you get farther from the mean, the frequency approaches zero.
(d) Almost all the data is within three standard deviations of the mean.
(e) The most frequent value is 1.
(f) The mean is exactly equal to the median and the mode.

2. Why do mathematicians use standard deviation to measure the spread of a body of data instead of interquartile range?

3. **Application** For each of the following data sets,

(i) calculate the mean, median, and standard deviation;
(ii) create a histogram or bar graph; and
(iii) explain why you think the distribution is or is not approximately normal.

(a) {41.5, 42.4, 42.6, 42.7, 42.9, 43.0, 43.6, 44.0, 44.5, 44.6, 44.6, 44.8, 45.0, 45.3, 45.5, 45.5, 45.6, 45.7, 45.8, 46.1, 46.3, 46.4, 46.5, 46.6, 46.8, 47.0, 47.2, 47.6, 47.6, 47.9}
(b) {2, 4, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10, 11, 11, 11, 12, 12, 13, 13, 15}

4. Which of the following tables of data resembles a normal distribution? Explain.

(a)

10–19	20–29	30–39	40–49	50–59	60–69
3	5	17	20	11	4

(b)

2–5	6–9	10–13	14–17	18–21	22–25
2	8	8	3	4	5

(c)

10–24	25–39	40–54	55–69	70–84	85–99
2	7	16	10	4	1

5. Using the properties of normal distribution, estimate the standard deviation of the normal data samples in Question 4.

B

6. Out of 100 packages of jawbreakers, 68 packages contain between 120 and 150. Use your knowledge of normal distribution to estimate the average number of jawbreakers and the standard deviation of the sample.

7. **Communication** As a general rule, statistical analysis states that the standard deviation can be estimated using the formula $\sigma = \frac{range}{6}$.

(a) Do you think this estimate of σ is reasonable? Explain.

(b) Using the data in Question 3, what estimations of σ do you get?

8. An upcoming renovation to your school will leave exposed beams in the hallways. It is estimated that the tallest 15% of students might hit their head on these new beams. If your school has a mean height of 173 cm with a standard deviation of 7.7 cm, estimate what range of students would be affected? (**Hint**: Consider the tall half of a normal distribution.)

9. The amount of coffee an automatic machine dispenses (in ounces) can be represented by the normal distribution $X \sim N(10.2, 0.6^2)$.

(a) What range does 68% of the quantity of coffee dispensed lie between?

(b) Draw a diagram that represents each of the following.

(i) the percent of cups dispensed that contains greater than 10.8 oz

(ii) the percent of cups dispensed that contains between 9.6 oz and 10.2 oz

(iii) the percent of cups dispensed that contains less than 9 oz

(c) Is there a significant risk of a 12-oz cup overflowing? Explain.

10. Burns Appliance Co. offers a replacement warranty on their toaster ovens, which have a mean lifespan of 8.5 years, with a standard deviation of 0.8 years. How long a warranty would they establish if they could only afford to repair no more than 2.5% of the toaster ovens they make? (**Hint**: Consider the lower half of a normal distribution.)

Technolink For more information on using the **randNorm** command, see Appendix C.7 on page 406.

11. Using a TI-83 Plus calculator, generate a random sample of 15 numbers with $\bar{x} = 10$, $\sigma = 2$ that is normally distributed by following these steps.

(i) Press [MATH] and use the arrow keys to select the **PRB** menu.
(ii) Press [6] to select the **randNorm(** command.
(iii) Key in 10, 2, 15) and press [STO▸] [2nd] [ENTER].
(iv) Press [2nd] [Y=] to access the STAT PLOTS menu. Press [1] and turn it on using the arrows and [ENTER]. Make sure only one Stat Plot is on.
(v) Select the histogram using the arrow keys and press [ENTER].
(vi) Press [ZOOM] [9] to set up a standard display.

(a) Describe the distribution. Is it symmetrical? Bell-shaped?
(b) How many data are within one standard deviation of the mean?
(c) Change the number in Step (iii) from 15 to 100. What changes do you see in the distribution? How about 500?

12. The bowling scores of two players are being compared.

Kate: 89, 99, 120, 100, 91, 110, 125, 91, 95, 133, 124, 78, 92, 128, 139, 88, 100, 125, 144, 76, 84, 92, 110, 104, 103, 128, 72, 86, 102, 73

Bernie: 71, 82, 88, 89, 90, 90, 94, 95, 97, 98, 99, 100, 100, 101, 102, 102, 102, 104, 105, 106, 109, 110, 112, 114, 115, 118, 119, 133, 137, 144

(a) Calculate the mean and standard deviation for each player.
(b) What percent of each player's scores fall within one standard deviation of the mean?
(c) What bowling score would be three standard deviations above the mean for Kate? For Bernie?

Technolink For more information on using Fathom™ to simulate normal distributions, see Appendix D.9 on page 423.

13. Using Fathom™, generate a random sample of 20 numbers with $\bar{x} = 25$, $\sigma = 5$ that is normally distributed by following these steps.

(i) Drop a **New Case Table** into the Fathom™ workspace.
(ii) Change the attribute label **<new>** to **X** and press **Enter**.
(iii) Select **New Cases…** from the **Data** menu, key in 20, and press **Enter**.
(iv) Double-click on the collection icon and double-click in the **Formula** column across from **X**.
(v) Key in **randomNormal(25,5)** in the text box and click on **OK**.
(vi) Drag a **New Graph** off the toolshelf and drop it into the Fathom™ document. Drop the **X** attribute from the Case Table on the horizontal axis and change the graph type to histogram.
(vii) Set the bin width to 5 by double-clicking within the graph and changing the value displayed in the text box that appears.

(a) Describe the distribution. Is it symmetrical? Bell-shaped?
(b) How many data are within one standard deviation of the mean ?
(c) Add 100 new cases by clicking on the **Case Table** and selecting **New Cases…** from the **Data** menu. What changes do you see in the distribution? (You will have to adjust your scale by clicking on the graph and selecting **Rescale Graph Axis** from the **Graph** menu.)
(d) Change the number in part (c) from 100 to 500. What changes do you see in the distribution?

C **14.** **Thinking, Inquiry, Problem Solving** Rajinder's Data Management class wrote a unit test and the teacher reported that the results were normally distributed with a mean of 61 (out of 85) and a standard deviation of 7.2.

(a) Angelique claims that she scored 38 on the test. Using your knowledge of normal distribution, is this result likely?

(b) What would Rajinder expect the highest mark to be, given the statistical data that he knows about the test?

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

15. **Knowledge and Understanding** If automobile windshields have a thickness of 14.6 mm and a standard deviation of 0.02 mm, use your knowledge of normal distribution to predict what percent of windshields have a thickness

(a) between 14.58 mm and 14.6 mm

(b) less than 14.58 mm

(c) more than 14.64 mm

(d) between 14.56 mm and 14.62 mm

16. **Application** If the mass of 35 dogs in your neighbourhood were normally distributed, with a mean of 11.2 kg and a standard deviation of 2.8, how many dogs would you expect to have a mass

(a) between 8.4 kg and 14 kg? **(b)** between 5.6 kg and 16.8 kg?

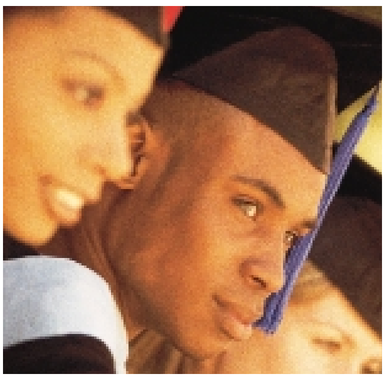
(c) between 2.8 kg and 19.6 kg? **(d)** 11.2 kg or less?

17. **Thinking, Inquiry, Problem Solving** Flooring materials being produced at a lumber mill have an average thickness of 7.9 mm, with a standard deviation of 0.4 mm.

(a) What range do about 68% of the flooring materials fall within?

(b) If you wanted 95% of the materials to fall within the range 7.7 mm and 8.15 mm, what mean and standard deviation would be required?

18. **Communication** Explain why a selection of 10 students from your class can have marks that aren't normally distributed when the marks of the whole class are normally distributed?



Chapter Problem

Comparing Marks

Recall the mean you calculated on page 162 and the standard deviation you calculated on page 172. Refer to the data on page 140.

CP11. How many Grade 12 students in Justin's school have an average mark within one standard deviation of the mean? Two? Three?

CP12. Is this distribution a normal distribution?