CHAPTER 7

Probability Distributions for Continuous Variables

Beekeepers remove honey from beehives, clarify it, and fill jars to sell to consumers. Processing engineers and technicians use probability distributions to ensure that the variation in the amount of honey in each jar falls within acceptable limits.

- What are some other food products sold in jars or bottles that are filled in a processing facility?
- Would you expect most of the jars to contain the amount stated on the label, more than this amount, or less than this amount?

Key Terms

attribute
frequency histogram
frequency polygon
normal distribution
margin of error
confidence interval
confidence level
continuity correction

Literacy Strategy

A frequency table provides a way to organize large amounts of data obtained from sampling. It simplifies analysis. For example, you can use a table to organize the amount of honey in 100 sample jars. Using the information in the table, you can generate a graph and determine probabilities.

Amount of Honey (g)	Number of Jars
145–149.9	1
150–154.9	3
155–159.9	9
160–164.9	17
165–169.9	26
170–174.9	25
175–179.9	17
180–184.9	2

Career Link

Mechanical Engineer

Doris plans to study mechanical engineering in preparation for a career in the food processing industry. She is working at a co-op position on a large farm that harvests and packages honey for sale in stores. The amount of honey in a jar varies, with a known mean and standard deviation. The mean can be adjusted on the filling machine.

- Why must the mean be set higher than the stated amount of honey in the jar?
- Why should the mean not be set too high?

Chapter Problem

Food Service Industry

In the food service industry, machines fill containers to a stated mean value with a known standard deviation. Engineers and technicians ensure that the machines are adjusted properly. The jar shown is labelled 500 g.

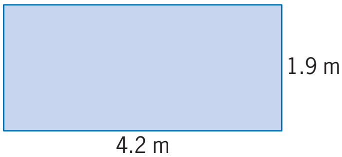
- Would you expect all jars to contain exactly this amount of honey?
- Why is it unwise, from a business point of view, for too many jars to contain less than this amount?
- Suggest a reason why a company would prefer not to have many containers with much more than this amount.
- How can technicians use a probability distribution to properly adjust the filling machine?

Prerequisite Skills

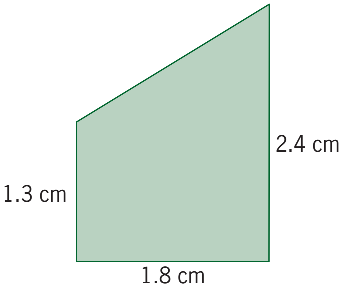
Areas of Rectangles and Trapezoids

1. Determine the area of each figure.

a)



b)



Organizing Data: Histogram, Mean, Standard Deviation, and *z*-Scores

A histogram is a graphical display that uses bars of varying heights to represent the frequency with which data occur.

If there are n data points in a data set, then

sample mean: $\overline{x} = \dfrac{\sum x}{n}$

sample standard deviation: $s = \sqrt{\dfrac{\sum(x - \overline{x})^2}{n - 1}}$

The z-score is the number of standard deviations that the value of a continuous variable is from the mean.

z-score: $z = \dfrac{x - \overline{x}}{s}$

A data value below the mean has a negative z-score. A data value above the mean has a positive z-score.

2. Eighteen students counted the number of coins in their pockets.

Number of Coins					
3	1	4	3	1	2
4	6	3	5	3	6
5	6	3	2	4	7

a) Sketch a histogram for the data.

b) Sketch a relative frequency histogram for the data.

c) Use technology to determine the mean and standard deviation for the number of coins.

d) Determine the z-score for a student with two coins in her pocket.

3. On a final mathematics exam, the students in one class scored a mean of 74 with a standard deviation of 4. The students in another class scored a mean of 72 with a standard deviation of 6.

a) What mark would result in the same z-score for each class?

b) Are there any other marks that would work? Explain why or why not.

4. Thayer collected bonus bills from a hardware store for several months. He found that the 25-cent bill turned up with a relative frequency of 0.4. How many 25-cent bills would be expected in a sample of 80 bills?

Counting Principles: Permutations and Combinations

5. Evaluate.

a) $5!$

b) ${}_3P_3$

c) ${}_{10}P_2$

d) ${}_7C_4$

6. Farouk is visiting an amusement park. He has time to ride 5 of the 17 rides at the park. In how many ways can he choose his 5 rides?

7. Wayne is an author of 12 textbooks. He wants to display one copy of each book on a bookshelf.
   a) In how many ways can he arrange the 12 books from left to right?
   b) He has three books for each grade from 9 to 12. In how many ways can he arrange the books if they are clumped in groups of 3 for each grade in increasing order from left to right?

Probability

8. Sam purchases 5 green gumballs, 7 red gumballs, and 3 white gumballs. The gumballs are in a paper bag. He reaches into the bag and pulls out one gumball. What is the probability that it is green?

9. Six students are lined up to purchase tickets for a movie. What is the probability that they are lined up in alphabetical order by first name?

10. The Drama Club holds a draw at each performance to raise money for props and costumes. They sell 200 tickets at $2 each. There is one prize of each of $100, $75, and $25. What is the expected value of each ticket sold?

Discrete Probability Distributions

11. A board game uses a 12-sided die as shown. It is rolled 7 times. What is the probability that it comes up greater than 9 exactly 4 times?

12. The barriers at a commuter train crossing are down for a total of 12 min every hour. If 100 cars approach the barrier every hour, what is the probability that exactly 20 will find the barriers down?

13. A mathematics class has 4 students with red hair, 6 with black hair, 9 with brown hair, and 7 with blond hair. The teacher randomly chooses 4 students to plan a class pizza party. What is the probability that all 4 have red hair?

14. Of the 1000 students at Eastdale Secondary School, 420 are boys. For a promotion, the cafeteria manager selects 10 students at random to receive a free sample of a new wrap. Use technology to determine the probability that an equal number of boys and girls will receive a free sample.

We often calculate probabilities for a range of values. Keep in mind the differences among terms such as “at most,” “less than,” “greater than,” and “at least.”

15. A department store mails out Saturday-Scratch-n-Save cards to all of the households in a large city. One card in 100 offers a discount of 50%, while the rest offer 5%. Last Saturday, 250 customers used their cards.
   a) What is the probability that exactly 3 customers received a 50% discount?
   b) What is the probability that more than 1 but fewer than 4 customers received a 50% discount?

7.1

Continuous Random Variables

Learning Goals

I am learning to

- distinguish between discrete variables and continuous variables
- work with sample values for situations that can take on continuous values
- represent a probability distribution using a mathematical model
- represent a sample of values of a continuous random variable using a frequency table, a frequency histogram, and a frequency polygon

Minds On...

A beekeeper collects data such as the number of bees in a hive or the amount of honey produced by the bees in a hive.

- Suggest some possible values for the number of bees in a hive and for the amount of honey produced in a hive. You may wish to use the Internet to help you provide reasonable estimates.
- What is different about the types of numbers used for each?

Action!

Investigate Comparing Discrete and Continuous Random Variables

attribute

- a quality or characteristic given to a person, group, or object

Project Prep

The data in your project may take different forms. Make sure you identify your continuous and discrete variables appropriately.

1. Consider attributes of students in your class, such as number of siblings or height.

a) List several attributes that are counted using discrete values.

b) List several attributes that are measured using continuous values.

2. Some students were asked for the number of siblings in their families. The table shows the results.

a) Classify the number of siblings as a discrete or a continuous variable. Explain your reasoning.

b) Represent the data using a histogram.

Number of Siblings	Number of Students
0	3
1	5
2	7
3	3
4	2
5 or more	1

3. Students recorded the time, to the nearest minute, spent on math homework one evening. The table shows the results.

Time (min)	Number of Students
0–10	2
10–20	5
20–30	6
30–40	10
40–50	4
50+	1

If a data value falls on the boundary between two intervals, it is usually placed in the lower interval.

For example, a data value of 10 min is recorded in the 0–10 min interval.

a) Classify time as a discrete or a continuous variable. Explain your reasoning.

b) Why is the time shown in intervals?

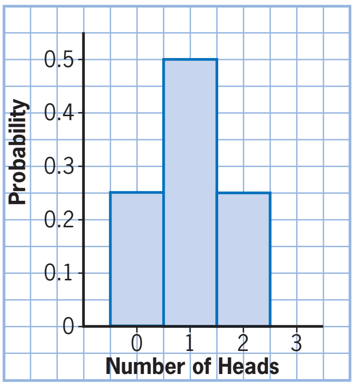
c) Draw a scatter plot of these data. For the time value, use the midpoint of each interval. Sketch a smooth curve through the points on the scatter plot.

d) **Reflect** Does the shape of the curve make sense? Explain.

e) **Extend Your Understanding** Consider the choice of intervals in the table. Why must you be careful not to have too few or too many intervals?

You can represent a probability distribution by a table and a graph that relates each outcome to the probability that it occurs. For discrete data, the variable can take on only certain values, often whole numbers. For example, suppose you flip a coin twice, and count the number of heads.

Number of Heads	Probability
0	$\frac{1}{4}$ or 0.25
1	$\frac{1}{2}$ or 0.50
2	$\frac{1}{4}$ or 0.25



The first column of the table must include all possible outcomes.

- What is the sum of the numbers in the probability column?
- Will this always be the case, assuming that all possible outcomes have been considered?

Each rectangle in the graph has a width of one unit.

- What is the total area of the three rectangles?
- What does this area represent?

Processes

Connecting
Suggest two other examples of a discrete probability distribution that you have already encountered in this course.

For continuous data, you can group the outcomes into intervals. The variable can take on any value in the interval between two numbers, including decimal and fractional values. For example, suppose you measure the distances of various horseshoe throws, and group the distances in intervals.

Distance (m)	Probability
2–4	0.1
4–6	0.2
6–8	0.4
8–10	0.2
10–12	0.1

The first column must include all possible outcomes. What is the sum of the probabilities?

The probability density function defines the continuous probability distribution for a given random variable. The probability that a random variable assumes a value between a and b is the area under the curve between a and b. The total area under the probability distribution graph is equal to 1.

Example 1

Determine a Probability Using a Uniform Distribution

A survey at a doughnut shop shows that the time required for a customer to eat a doughnut varies from 30 sec to 3 min, with all times in between equally likely.

a) What kind of a distribution is this? How do you know?

b) Sketch a graph that illustrates this distribution. Place **Time** on the horizontal axis, and **Probability Density** on the vertical axis. Determine the vertical scale on the graph. Explain your reasoning.

c) What is the probability that a customer will eat a doughnut in 1 to 3 min?

d) How many values are possible for the time required to eat a doughnut? Explain your answer.

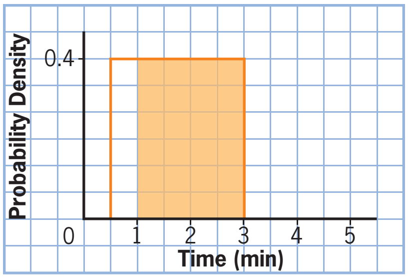
e) Is it possible to determine the probability that a customer will eat a doughnut in exactly 1 min using the area under the graph? Explain your answer.

Solution

a) Since all outcomes are equally likely, this is a uniform distribution.

b) All values from 0.5 min to 3.0 min are equally probable. The graph is a horizontal line from 0.5 min to 3.0 min.

The area under the graph represents the total of all of the probabilities. Therefore, the area must equal 1. The base of the rectangle has a length of 2.5.



The total area of the probability density distribution will always be equal to 1 (or 100%). You can determine probabilities for a range of values by calculating the area under the graph.

$$\begin{aligned} bh &= A \\ 2.5h &= 1 \\ h &= \frac{1}{2.5} \\ &= 0.4 \end{aligned}$$

The top of the graph occurs at 0.4.

c) The probability that a customer will eat a doughnut in 1 to 3 min is equal to the shaded area under the graph from 1.0 min to 3.0 min.

$$\begin{aligned} P(1 \leq X \leq 3) &= \text{area under the graph} \\ &= 0.4(3.0 - 1.0) \\ &= 0.8 \end{aligned}$$

The probability that a customer will eat a doughnut in 1 to 3 min is 0.8.

d) Since this is a continuous distribution, any real number between 0.5 and 3.0 is a possible value. An infinite number of possible values exist for the time required to eat a doughnut.

e) If you pick a single value such as 1 min, the rectangle under the graph will have a width of 0 min. The probability for a single value of a continuous distribution is 0. The area method cannot be used for single values of a continuous variable, only for a range of values.

It is tempting to want to say that the probability that it takes exactly 1 min to eat a doughnut is 0.4. However, this does not work for a continuous distribution. You must determine probabilities by using the area under the graph.

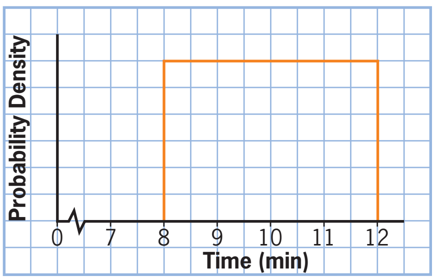
Your Turn

Chris works at a tire store. Chris can change a tire on a rim in 8 to 12 min, with all times in between equally likely.

a) What is the probability that Chris changes a given tire in less than 9 min?

b) What is the probability that it takes between 9 min and 11.5 min?

c) What is the probability that it takes exactly 10 min?



Example 2

Frequency Table, Frequency Histogram, Frequency Polygon

Many businesses use arrays of lights to attract customers. The life of a light bulb follows a continuous distribution. A technician sampled 40 light bulbs in a laboratory. The table shows the lifetime of each bulb, rounded to the nearest day.

Several versions of this sign at Sam the Record Man graced Yonge Street in Toronto from 1937 until 2007.

Life of Light Bulb (days)							
163	152	135	144	161	145	135	151
166	138	153	137	148	145	133	154
141	148	155	150	146	139	165	142
153	160	138	171	148	159	172	148
149	175	149	146	158	154	156	138

a) Can you use the data in the table to determine whether the data seem to follow a uniform distribution? Can you make a reasonable estimate of the mean lifetime of the bulbs?

b) Use a table like the one below to determine the frequency for each interval. The first two intervals have been completed for you.

<table>
<tr><th>Lifetime (days)</th><th>Tally</th><th>Frequency</th></tr>
<tr><td>120–130</td><td></td><td>0</td></tr>
<tr><td>130–140</td><td>𝍸 |||</td><td>8</td></tr>
<tr><td>140–150</td><td></td><td></td></tr>
</table>

c) Inspect the frequency table. Can you now answer part a) more easily?

d) How does a frequency table help you to analyse the raw data from a sampling experiment?

e) Use the frequency table to draw a **frequency histogram**. Then, add a **frequency polygon** to the histogram.

f) How is the shape of the frequency polygon related to the shape of the probability density distribution for the variable? Can you use the area under the frequency polygon to calculate probabilities for any range of values?

frequency histogram

- a graph with intervals on the horizontal axis and frequencies on the vertical axis

frequency polygon

- a segmented line that joins the midpoints of the top of each column in the frequency histogram

Solution

a) The data are difficult to analyse in this form. It is not obvious whether the distribution is uniform or not. Similarly, it is difficult to estimate the value of the mean with any accuracy.

b)

Lifetime (days)	Tally	Frequency
120–130		0
130–140	卌 |||	8
140–150	卌 卌 ||||	14
150–160	卌 卌 |	11
160–170	||||	4
170–180	|||	3
180–190		0

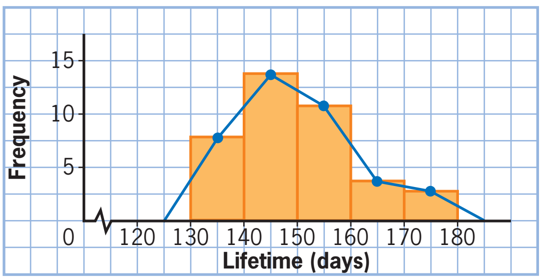
You can check for errors by adding the numbers in the frequency column. What should be the total in this case?

c) From the frequency table, it appears that the frequencies vary from 0 to 14. The distribution is not uniform. The mean lifetime appears to be around 145 days.

d) The frequency table groups the raw data into intervals. The frequency in each interval makes the shape of the distribution more obvious, and gives an indication of the location of the mean.

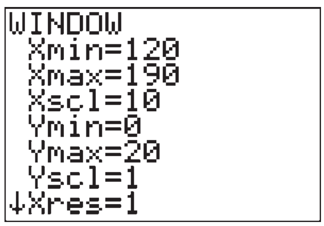
e) **Method 1: Use Paper and Pencil**

- Draw axes on a piece of graph paper. Label the horizontal axis from 120 to 190. Label the vertical axis from 0 to 20.
- Use an interval width of 10 to draw the histogram.
- Mark the midpoint of the top of each bar on the histogram. Join the points with a segmented line to sketch the frequency polygon.



Method 2: Use a Graphing Calculator

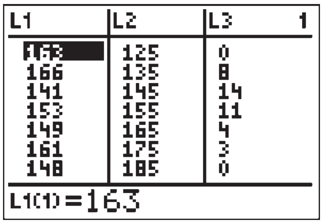
- Turn on the calculator and clear all lists. Adjust the graphing **WINDOW** as shown.
- Enter the raw light bulb data in list **L1**. How many entries are there?
- Enter the midpoint of each interval in list **L2**. How many entries are there?



Project Prep

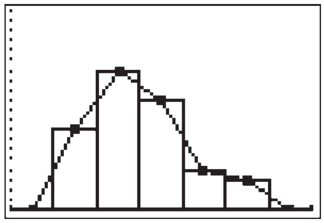
Technology is a good resource when graphing your raw data. Consider using a graphing calculator, spreadsheet, Fathom™, or other data analysis software.

- Enter the frequency data in list **L3**.



Why is the list in L1 not the same length as the lists in L2 and L3?

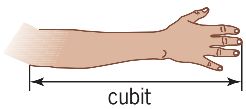
- Display the **STAT PLOT** screen. Turn on **Plot1**. Select the histogram. Ensure that **Xlist** is **L1**, and **Freq** is set to 1.
- Turn on **Plot2**. Select the line graph. Ensure that **Xlist** is **L2** and **Ylist** is **L3**.
- Press GRAPH to display the histogram and frequency polygon.



f) The shape of the frequency polygon gives an indication of the shape of the probability distribution for the variable. The total area under the frequency polygon is not equal to 1. You cannot calculate probabilities using areas under the frequency polygon. You need to use a probability distribution to determine probabilities.

Your Turn

A cubit measures the distance from the elbow to the tip of the outstretched middle finger. Ranjit's class has 30 students. Each student determines the length of a cubit using his or her arm. The frequency table shows the results.



Cubit Length (mm)	Frequency
514–516	1
516–518	2
518–520	3
520–522	4
522–524	8
524–526	5
526–528	4
528–530	2
530–532	1

a) Sketch a frequency distribution histogram.

b) Add a frequency polygon to the histogram in part a).

c) Estimate the mean cubit length in the class.

d) Would this be considered a uniform distribution? Why?

Consolidate and Debrief

Key Concepts

- Some situations result in discrete data. These are often whole numbers.
- Some situations result in continuous data over a range. Continuous data include fractional or decimal values. Continuous distributions are often the result of measurements.
- The probability that a variable falls within a range of values is equal to the area under the probability density graph for that range of values.
- You can represent a sample of values for a continuous random variable using a frequency table, a frequency histogram, or a frequency polygon.
- The frequency polygon approximates the shape of the probability density distribution.

Reflect

R1. The manager of a hotel collects data about the operation of the hotel. Two examples include the number of guests that occupy the hotel each day and the time that a given guest waits for an elevator to arrive. Which of these is discrete and which is continuous? Is it possible to list all values of the discrete distribution? Is it possible to list all values of the continuous distribution? Explain.

R2. Three classmates measured Maya's height. The results are 154 cm, 155 cm, and 157 cm. Suggest reasons why the results are not all the same. Is this a measurement error or an example of bias? Explain.

R3. A classmate cannot accept that the probability that a bottle of juice from the cafeteria contains exactly 280 mL, as shown on the label, is zero. How can you convince him this is correct?

Practise

1. Which of these variables would be expected to result in a discrete distribution?

a) the masses of the cupcakes for sale in the school cafeteria

b) the value of a card drawn at random from a deck of 52 cards

c) the daily barometric pressure measured in your city

2. Which of these variables would be expected to result in a continuous distribution?

a) the number of students with blue eyes in each class in your school

b) the weights of the police officers in the local police force

c) the number of cartons of milk sold in the cafeteria vending machine

d) the number of defective tablet computers in a shipment to an electronics store

3. A men's clothing store developed a customer waist size probability table from a sample of 300 customers, as shown.

Waist Size (in.)	Probability
26–28	0.000
28–30	0.025
30–32	0.175
32–34	0.295
34–36	0.300
36–38	0.160
38–40	0.045
40–42	0.000

a) Does the distribution appear to be uniform? Explain your answer.

b) What is the frequency associated with a waist size from 34 to 36?

c) If a new customer comes in with a waist size of 38, which interval should the data be placed in?

Literacy Link

Canada officially uses SI units for length, such as centimetres. However, in some industries—including the clothing industry—measurements are still recorded using imperial units, such as inches. 1 inch ≈ 2.54 cm.

Apply

4. Donna is monitoring nutrient levels in a local stream. She collects 44 samples of water from various locations. The table shows the volumes of the samples.

Volume of Sample (mL)			
55	63	56	64
55	62	61	65
57	62	59	63
55	65	60	62
58	62	56	61
63	60	64	57
57	64	60	59
55	61	56	56
58	58	65	64
58	57	60	65
63	61	59	59

a) Can you determine from the table if the distribution is uniform? Explain your answer.

b) Devise a plan to determine whether the distribution is uniform. Carry out your plan and draw a conclusion.

Processes

Representing
How did you choose to represent the data in the table to solve this problem? What are the advantages of your representation?

5. **Thinking** Air tanks used by scuba divers are typically filled to a pressure of 3000 psi (pounds per square inch). A sample of 32 tanks at a dive shop was selected, and the table shows the pressures.

Pressure (psi)			
3003	2999	2995	2999
3004	3001	3001	3000
2998	3000	2999	3000
3001	2997	2998	3003
3005	2995	3005	3002
2997	3004	2995	2997
2996	3001	3002	3002
2999	3003	2996	3003

a) Determine whether the distribution is uniform. Explain your method.

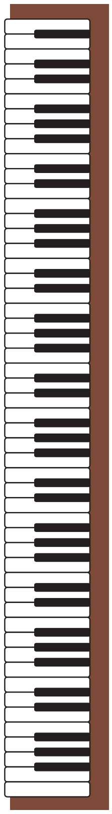
b) Is this distribution discrete or continuous? Give a reason for your answer.

c) Suggest a reason why are there no values in the table recorded with decimal places, such as 2998.347.

Literacy Link

In scuba diving, pressure is measured using the imperial unit of psi, which means pounds per square inch.

6. Consider the music produced by a piano.



a) How many keys are on a standard piano keyboard including both black and white keys?

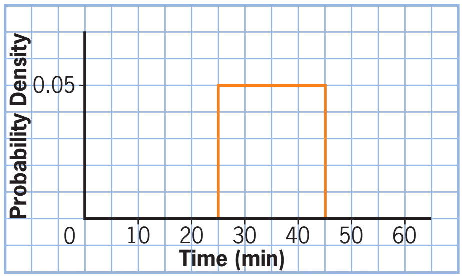
b) The frequency of a piano note varies from a low of about 28 Hz at the left of the keyboard to a high of about 4100 Hz at the right of the keyboard. If you assign a number to each key, and count how many times that key is used in a piece of music, you will get a frequency distribution. Do you expect the distribution to be uniform? Explain your answer.

c) Is the distribution in part b) discrete or continuous? Give a reason for your answer.

Literacy Link

The frequency of a sound is measured in units called *hertz* (Hz). Humans can hear from a low of about 20 Hz to a high of about 20 000 Hz.

7. The Ridgeway High School Paperman Triathlon consists of a 200-m swim, a 5-km bicycle ride, and a 1-km run. The graph shows the probability distribution for the time required to complete the triathlon.



a) What is the probability that a contestant will finish the triathlon in 30 min or less?

b) What is the probability that a contestant will finish the triathlon in 30 to 40 min?

c) Why was this type of distribution used? Explain.

8. A trombone can create musical notes like a piano. Because the musician uses a slider to control the frequency of the notes, the trombone can also play all of the frequencies between the notes that a piano is restricted to. A tenor trombone produces frequencies from a low of about 80 Hz to a high of about 600 Hz.

a) If a trombone player plays many notes at random frequencies, you can collect samples. Is this distribution discrete or continuous? Explain.

b) Suppose a trombone player plays a musical composition. If you measured the frequencies of the notes being played, would you expect the distribution to be discrete or continuous? Explain.

9. Sardines that will be canned are sorted from the main catch. Only fish that are 95 mm to 100 mm long are kept. The table shows the lengths of a sample of 24 sardines.

Length of Sardine (mm)			
99	98	100	95
98	95	95	96
97	100	100	99
100	99	97	97
98	99	97	96
96	99	99	98

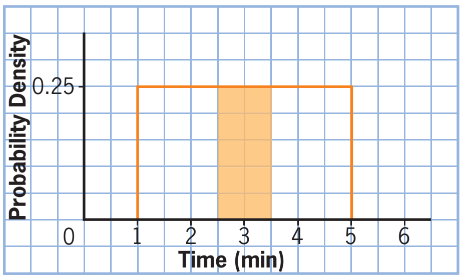
a) Use the table to determine whether the distribution is uniform. Explain your method.

b) The plant manager would like to model these data with a uniform distribution. What should the manager use for the height of the probability density graph?

c) Sketch the probability density graph.

d) What is the probability that a sardine has a length less than or equal to 98 mm?

10. **Communication** Consider the uniform probability distribution shown. Jon says the shaded area represents the probability that the variable has a value of 3. Sunita says the shaded area represents the probability that the variable has a value of 2.5 to 3.5. Ahmed says the shaded area represents the probability that the variable has a value of 2 to 4. Who is correct? Give reasons for your answer.



✓ Achievement Check

11. Carbon resistors are used in all types of electronic circuits including televisions, smartphones, and computers. Carbon resistors act to reduce current flow. They usually have a fixed resistance value measured in kilo-ohms (kΩ). When they are manufactured, those within 5% of a stated resistance value are set aside and form an approximately uniform distribution. Consider a 100 kΩ resistor.

a) What are the low and high cutoff values for this distribution?

b) Sketch a probability distribution for these resistors, assuming a uniform distribution.

c) What is the height of the probability distribution?

d) What is the probability that a given resistor has a resistance within 0.25% of the stated resistance value?

Literacy Link

Electrical resistance is measured in units called *ohms*, named after German high school teacher Georg Ohm (1789–1854). He developed the electrical theory of resistance that underlies most of the electrical and electronic devices that we are familiar with.

12. Doppler radar in a device similar to a police radar gun can be used to measure speeds of birds in flight. The table shows the measured speeds of 30 Canada geese.

Speed (km/h)				
78.5	68.0	64.7	75.1	68.3
76.6	76.6	71.2	73.7	73.5
76.7	64.6	66.7	68.8	73.6
80.2	72.4	69.6	77.3	74.8
72.5	66.1	73.5	75.0	68.9
71.7	65.7	67.4	70.9	73.0

a) Are these data discrete or continuous? Explain your answer.

b) Choose an appropriate interval width. Give a reason for your choice.

c) Construct a frequency table.

d) Use your frequency table to construct a frequency histogram.

e) Add a frequency polygon to the frequency histogram.

f) Estimate the mean of the speed data.

g) Estimate the probability that a goose is exceeding 80 km/h. Explain your thinking.

13. **Thinking** Should hair colour be considered discrete or continuous data? Give reasons for your answer.

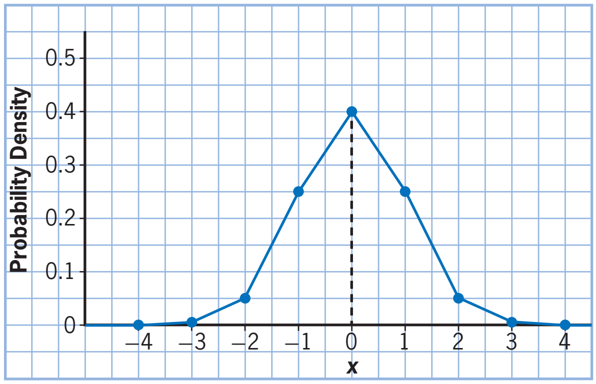
Extend

14. Some musical instruments can play only certain frequencies, while others can play any frequency over a range.

a) Suggest three instruments that can play only discrete frequencies.

b) Suggest three instruments that can play any frequency over a range.

15. **Thinking** Most distributions are not uniform, but the probability that a value occurs between two boundary values is still equal to the area under the probability density curve. Study the probability density graph shown.



a) Develop a method for determining the area under the plot between any two points.

b) Use your method to determine the probability that the value of x lies between 0 and 3.

7.2

The Normal Distribution and *z*-Scores

Learning Goals

I am learning to

- determine the theoretical probability for a continuous random variable over a range of values
- determine the mean and standard deviation of a sample of values
- calculate and explain the meaning of a *z*-score
- solve real-world probability problems involving normal distributions

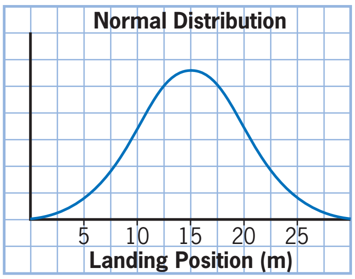
Minds On...

In a spot landing contest, airplane pilots take turns trying to land as close as possible to a line painted on the runway. Most will land close to the line, with fewer and fewer as the distance from the line increases. This kind of probability distribution is referred to as a **normal distribution**.

normal distribution

- a probability distribution around a central value, dropping off symmetrically to the right and left, forming a bell-like shape

Suppose that the landing zone is 30 m long, with the target line at 15 m. The graph of a normal distribution for the spot landing contest would look something like this:



Suggest two or three other situations that might reasonably be expected to form a normal distribution.

Study the shape of the normal distribution. What features can you see?

Action!

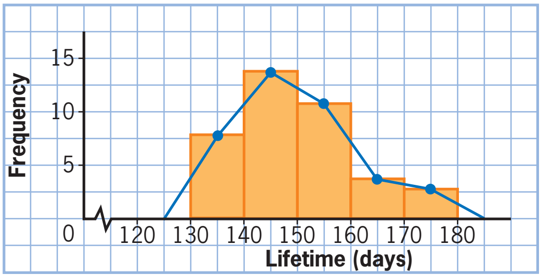
Mathematically, the two "tails" of the normal distribution continue forever. In a real situation, the probability of finding a datum far from the central peak is essentially zero.

If you collect sample values of a variable that is expected to follow a normal distribution, the sample data will cluster around a central peak to form a bell-like shape, or "bell curve."

For a discrete distribution, you can calculate probabilities using counting techniques. For a continuous distribution, you can calculate probabilities by determining the area under a graph for a range of values. In this section, you will learn how to determine probabilities if you know that the distribution is a normal distribution.

Investigate The Effect of Interval Width on a Frequency Polygon

1. Review the frequency histogram and the frequency polygon from the light bulb data used in section 7.1. To estimate the mean, is one more useful than the other? Give a reason for your answer.



Life of Light Bulb (days)			
163	152	135	144
166	138	153	137
141	148	155	150
153	160	138	171
149	175	149	146
161	145	135	151
148	145	133	154
146	139	165	142
148	159	172	148
158	154	156	138

2. Work with a partner to construct a frequency table with intervals of 5 days, and another with intervals of 20 days.

3. The frequency polygon displays a segmented picture of the actual frequency distribution. Use the tables to produce frequency histograms and frequency polygons.

4. Compare the three frequency histograms. Which appears to display the data in the most useful way? Give a reason for your answer.

5. Compare the three frequency polygons. Which appears to follow the smoothest curve?

6. **Reflect** Would an interval width of 1 day be suitable for a frequency table in this case? Explain why or why not.

7. **Extend Your Understanding** Suppose the testing laboratory wanted to use an interval of 1 day to produce a very smooth frequency polygon. What changes would need to be made to the process of gathering data?

If you gather a large amount of data and use a small interval width, the frequency polygon looks like a smooth function.

Example 1

Use a Frequency Distribution to Estimate Probabilities

Fruityfizz Soft Drinks bottles its products in containers marked 500 mL. The table shows the frequency distribution for a sample of 200 bottles.

Volume (mL)	Frequency, *f*
490–492	0
492–494	0
494–496	2
496–498	11
498–500	43
500–502	81
502–504	48
504–506	14
506–508	1
508–510	0

a) Add a relative frequency column to the table.

b) Use the table to determine the probability that a given bottle will contain less than 500 mL of soft drink.

c) Use the table to determine the probability that a given bottle will contain between 498 mL and 502 mL of soft drink.

d) Is it possible to determine the probability that a given bottle will contain exactly 500 mL of soft drink using the table? Explain your answer.

e) Sketch a scatter plot of the frequency versus the volume. For the horizontal axis, use the midpoint of each interval.

f) Sketch a scatter plot of the relative frequency versus the volume. For the horizontal axis, use the midpoint of each interval. Sketch a smooth curve through the points. How does the shape of the graph of the frequency data compare to the shape of the graph of the relative frequency data?

g) Could you use the area under the relative frequency graph to answer parts b) and c)? Explain your answer.

h) Interpret the answers to parts b) and c) in relation to areas under a probability density graph.

Solution

a) To determine the relative frequency, divide each frequency by the total number of bottles tested.

Volume (mL)	Frequency, f	Relative Frequency, $rf = \frac{f}{200}$
490–492	0	0.000
492–494	0	0.000
494–496	2	0.010
496–498	11	0.055
498–500	43	0.215
500–502	81	0.405
502–504	48	0.240
504–506	14	0.070
506–508	1	0.005
508–510	0	0.000

b) To determine the probability, add the relative frequencies from 490 mL to 500 mL.

$0 + 0 + 0.010 + 0.055 + 0.215 = 0.280$

The probability that a given bottle will contain less than 500 mL of soft drink is 0.280.

c) To determine the probability, add the relative frequencies from 498 mL to 502 mL.

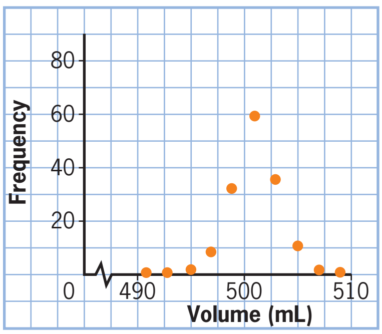
$0.215 + 0.405 = 0.620$

The probability that a given bottle will contain between 498 mL and 502 mL of soft drink is 0.620.

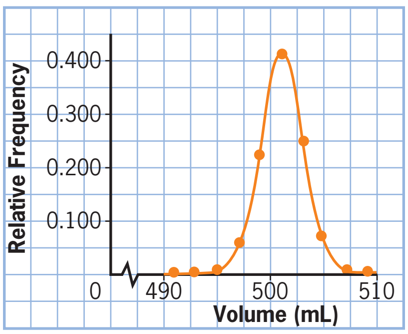
d) Probabilities for a continuous distribution cover a range of values. The probability that a given bottle will contain exactly 500 mL of soft drink would require an interval width of 0 mL, which is not available on the table. In fact, the probability that a given bottle will contain exactly 500 mL of soft drink is zero.

Consider the number of possible values that the volume of soft drink in the bottle could take. Can you use the definition of probability to explain why the answer to part d) makes sense?

e)



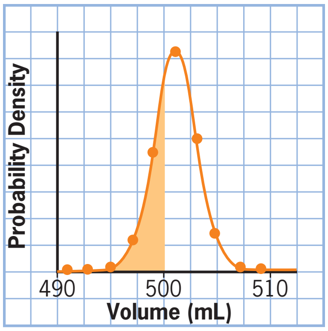
f)



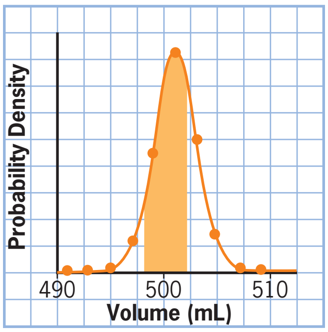
The graph of the relative frequency data has the same shape as the graph of the frequency data.

g) In order to use the area under a graph to calculate probabilities, the total area must be equal to 1. This is called a probability density graph. The relative frequency graph in part f) has the same general shape as the probability density graph, but the area under the graph is not equal to 1. You cannot use it to calculate probabilities by finding areas.

h) The probability that a bottle contains less than 500 mL is equal to the area under the probability density graph from the far left up to $x = 500$.



The probability that a bottle contains between 498 mL and 502 mL is equal to the area under the probability density graph from $x = 498$ to $x = 502$.



The vertical scale on the probability density graph must be chosen such that the total area under the graph is equal to 1. This is not easy to do. However, you can determine the area under a curve fairly easily if the distribution is normal.

Your Turn

A men's clothing store kept records of the waist sizes of a sample of their customers, measured to the nearest quarter of an inch. The table shows the relative frequencies.

Waist Size (in.)	Relative Frequency
26–28	0.000
28–30	0.025
30–32	0.175
32–34	0.295
34–36	0.300
36–38	0.160
38–40	0.045
40–42	0.000

Although Canada officially uses SI units for length, such as centimetres, clothing measurements are often recorded using imperial units, such as inches. 1 inch ≈ 2.54 cm.

What is the probability that a customer has a waist size

a) between 30 and 32 in.?

b) of more than 36 in.?

c) between 30 and 36 in.?

d) of exactly 38 in.?

If a variable is expected to follow a normal distribution, you can take a representative sample. You can use the mean and standard deviation of the sample to approximate the mean and standard deviation of the underlying normal distribution. The approximation becomes more accurate as more data are collected.

Example 2

Spot Landing Contest

Forty aircraft participated in a spot landing contest at the Wainfleet Ring Aerodrome. The landing zone was 30 m long, with the target line at the 15 m mark. The table shows the touchdown position of each aircraft along the landing zone. The data in the table are expected to follow a normal distribution.

Project Prep

Many continuous data sets are distributed normally. Consider including that in your data analysis.

Landing Zone Position (m)			
10.6	18.9	17.7	22.9
12.2	11.9	12.2	10.6
17.0	13.4	14.0	14.6
14.0	15.5	18.9	13.4
11.6	12.5	18.0	9.8
11.3	16.2	10.6	14.6
18.0	13.1	11.9	12.5
14.6	17.4	15.2	11.9
19.8	22.3	14.6	15.5
10.6	17.7	12.2	15.5

a) Determine the mean and the standard deviation of the spot landing data.

b) What is the z-score for a pilot who lands her plane at a position of 18.3 m?

c) What is the probability that a pilot lands at a position of 18.3 m or less?

d) What is the probability that a pilot lands at a position of more than 18.3 m?

e) What is the probability that a pilot lands at a position between 12.2 m and 18.3 m?

Solution

a) Method 1: Use Paper and Pencil

Add the landing distances, and divide by 40.

$$\overline{x} = \frac{\sum x}{n}$$
$$= \frac{585.2}{40}$$
$$= 14.63$$

The mean of the landing distances is 14.63 m.

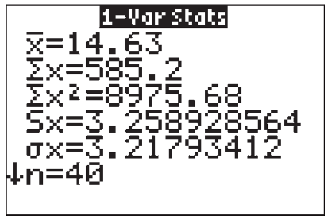
Use the equation for the standard deviation.

$$s = \sqrt{\frac{\sum x^2 - n(\overline{x})^2}{n-1}}$$
$$= \sqrt{\frac{8975.68 - 40(14.63)^2}{40-1}}$$
$$\approx 3.259$$

The standard deviation is 3.259 m.

Method 2: Use a Graphing Calculator

- Turn on the calculator and clear all lists.
- Enter the landing distance data in list **L1**.
- Press STAT and scroll to **CALC**.
  Select **1:1-Var Stats**.
- Press 2ND, then **1** to select **L1**.
  Press ENTER.



The statistical summary will display as shown. The mean, $\overline{x}$, is 14.63 and the standard deviation, Sx, is 3.259.

b) Use the equation for a z-score:

$$z = \frac{x - \overline{x}}{s}$$
$$= \frac{18.3 - 14.63}{3.259}$$
$$\approx 1.12$$

The z-score for a landing at 18.3 m is 1.12.

The z-scores for a normal distribution follow a normal distribution themselves, with a mean of 0 and a standard deviation of 1. The area under the distribution of z-scores is equal to 1. It is possible to construct a probability table for these z-scores. You can find this table on pages 480–481.

c), d), e)

Method 1: Use a Table

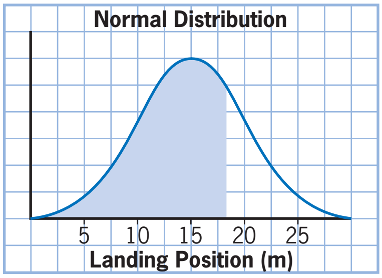
To determine $P(X \le 18.3)$, start by moving down the z-column until you reach a z-score of 1.1. Then, move to the right until you reach a z-score of 1.12. Read the probability as 0.8686.

Z	0.00	0.01	0.02	0.03	
0.0	0.5000	0.5040	0.5080	0.5120	0.
0.1	0.5398	0.5438	0.5478	0.5517	0.
0.2	0.5793	0.5832	0.5871	0.5910	0.
0.3	0.6179	0.6217	0.6255	0.6293	0.
0.4	0.6554	0.6591	0.6628	0.6664	0.
0.5	0.6915	0.6950	0.6985	0.7019	0.
0.6	0.7257	0.7291	0.7324	0.7357	0.
0.7	0.7580	0.7611	0.7642	0.7673	0.
0.8	0.7881	0.7910	0.7939	0.7967	0.
0.9	0.8159	0.8186	0.8212	0.8238	0.
1.0	0.8413	0.8438	0.8461	0.8485	0.
1.1	0.8643	0.8665	0.8686	0.8708	0.
1.2	0.8849	0.8869	0.8888	0.8907	0.

The entry in the table represents the probability that the variable has a value less than or equal to the value represented by the z-score.

The probability that a pilot lands at a position of 18.3 m or less is 0.8686.

Note that this is equal to the area under the normal distribution curve to the left of 18.3 m.

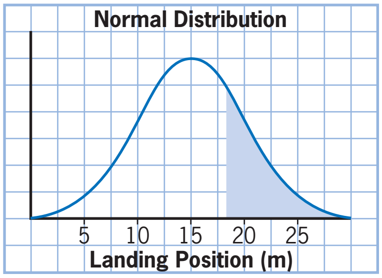


Recall that the probability that a variable takes on a specific value is zero. Therefore, you get the same probability for $P(X \leq 18.3)$ as you do for $P(X < 18.3)$. The "extra" area for $P(X = 18.3)$ is equal to zero.

The probability that a pilot lands at a distance of more than 18.3 is equal to 1 minus the probability that the pilot lands at 18.3 m or less. These two probabilities are complements of one another.

$$\begin{aligned} P(X > 18.3) &= 1 - P(X \leq 18.3) \\ &= 1 - 0.8686 \\ &= 0.1314 \end{aligned}$$

The probability that a pilot lands at a distance of more than 18.3 m is equal to 0.1314. Note that this is equal to the area under the normal distribution curve to the right of 18.3 m.



The probability that a pilot lands at a position between 12.2 m and 18.3 m is equal to the probability that the pilot lands at 18.3 m or less minus the probability that the pilot lands at 12.2 m or less.

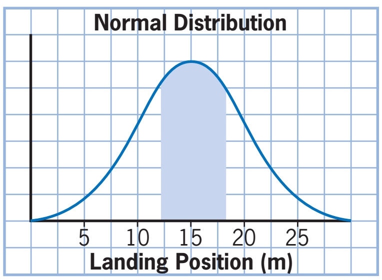
The z-score for a pilot landing at 12.2 m is

$$\begin{aligned} z &= \frac{x - \overline{x}}{s} \\ &= \frac{12.2 - 14.63}{3.259} \\ &\approx -0.75 \end{aligned}$$

Use the table on pages 480–481 to determine the associated probability for a z-score of -0.75 as 0.2266.

$$\begin{aligned} P(12.2 \leq X \leq 18.3) &= P(X \leq 18.3) - P(X \leq 12.2) \\ &= 0.8686 - 0.2266 \\ &= 0.6420 \end{aligned}$$

The probability that a pilot lands at a position between 12.2 m and 18.3 m is equal to 0.6420. This is equal to the area under the normal distribution curve to the right of 12.2 m and to the left of 18.3 m.

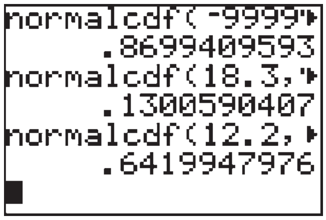


c), d), e)

Method 2: Use a Graphing Calculator

- Press 2ND, then VARS to access **DISTR**.
- Select **2:normalcdf(**.

The normal cumulative distribution function calculates the area under the normal curve from a lower to an upper bound. Choose −99999 as the lower bound and an upper bound of 18.3. Use the values for the mean and standard deviation that were calculated in parts a) and b). For example: **normalcdf(−99999, 18.3, 14.63, 3.259)**.



The probability that a pilot lands at 18.3 m or less is 0.8699.

You can use a subtraction process to determine the probability that a pilot lands at more than 18.3 m, as in the table method. Alternatively, use the normal distribution function with a lower bound of 18.3 and an upper bound of 99999. The probability is 0.1301.

To determine the probability that a pilot lands at 12.2 m to 18.3 m, use a lower bound of 12.2 and an upper bound of 18.3. The probability is 0.6420.

Note that the graphing calculator solution gives slightly different results from those using the table, since the z-score values in the table are rounded to two decimal places.

The normal distribution curve continues to infinity in both directions. When determining a probability, sometimes you need to continue the area to infinity. The graphing calculator has no symbol for infinity, so a large number like 99 999 is used instead.

Processes

Selecting Tools and Computational Strategies

Did you have a choice of methods to use for this problem? If so, which did you choose? What were your reasons for the choice?

Your Turn

Kunal is trying out for the school football team. He wants to know how far he can kick the ball for a field goal. The table shows the data from 20 trials.

Field Goal Distance (yd)			
17	27	31	25
25	44	35	24
31	48	42	48
45	34	41	38
40	43	45	21

a) Determine the mean and standard deviation of the data.

b) What is the probability that Kunal kicks a distance of less than 30 yd?

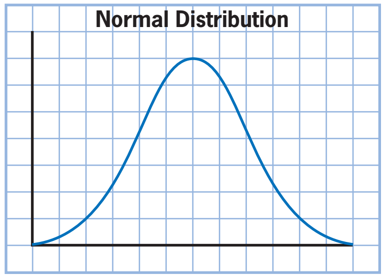
c) What is the probability that Kunal kicks a distance of 20 yd to 40 yd?

In some sports, measurements are recorded using imperial units, such as yards. One yard is slightly shorter than one metre. 1 yd ≈ 0.9 m.

Consolidate and Debrief

Key Concepts

- The frequency polygon approximates the shape of the frequency distribution.
- You must use a range of values to determine the theoretical probability for a continuous random variable.
- The probability that a continuous random variable takes on any single value is zero.
- A normal distribution is a probability distribution around a central value, dropping off symmetrically to the right and left, forming a bell-like shape.
- You can determine the probability that a variable will lie within a range of values by finding an area under the normal distribution.
- You can use z-scores to determine probabilities, either from a table or by using technology.



Reflect

R1. Consider the field goal distances in the Your Turn following Example 2. As Kunal practises and gains more skill, would you expect the mean and standard deviation to remain the same? Explain your answer.

R2. Relate to the graph of a normal distribution to explain why $P(z < -5)$ is almost zero.

Practise

Use this information to answer #1 to #4.

Roberta rides her bicycle to school. She records her times for one week, and determines a mean of 10.2 min and a standard deviation of 2.4 min. She believes the time follows a normal distribution.

Choose the best answer for #1 and #2.

1. What is the z-score for a trip that takes 13.8 min?

A -3.6

B -1.5

C 1.5

D 3.6

2. A trip has a z-score of -0.5. How long did the trip take?

A 9.0 min

B 9.7 min

C 10.7 min

D 11.4 min

3. What is the probability that a trip will take twice as long as the mean?

4. **Communication** Roberta makes similar measurements the following week, and calculates a different mean and standard deviation. Why did this happen? How can she obtain more reliable values for the mean and standard deviation?

Apply

5. The corner of Bloor and Bathurst streets in Toronto is the home of Honest Ed's, a huge discount store famous all over the world. The sign contains about 23 000 light bulbs.

The filament of an incandescent light bulb slowly evaporates, limiting the life of the bulb. Turning the bulb on and off also shortens its lifetime. Technicians tested a sample of 500 bulbs. The table shows the frequencies for the lifetime of the bulbs.

Lifetime (days)	Frequency
300–325	2
325–350	15
350–375	38
375–400	55
400–425	91
425–450	94
450–475	73
475–500	68
500–525	40
525–550	14
550–575	9
575–600	1

a) Sketch a frequency histogram and a frequency polygon for these data.

b) Estimate the mean life of these light bulbs.

c) Add a relative frequency column to the table.

d) What is the probability that a given light bulb will fail in 400 days or fewer?

e) If you want to be reasonably sure that there would never be a burned out bulb on the sign, how often should you replace all of the bulbs? Explain.

6. Mieke is using a tennis ball machine to launch tennis balls over the net at various speeds. The table shows the speeds from a sample of 40 launches.

Speed of Ball (km/h)			
63	69	62	63
60	62	58	63
59	65	67	64
51	57	58	57
63	60	62	61
58	68	61	53
60	56	50	59
60	66	62	69
61	55	57	65
61	63	63	71

a) Use an interval width of 5 km/h, starting at 49 km/h. Construct a frequency table.

b) Use your frequency table to construct a frequency histogram.

c) Add a frequency polygon to the frequency histogram.

d) Do the data appear to follow a normal distribution? Explain your answer.

e) Estimate the mean.

f) Add a relative frequency column to the table.

g) What is the probability that a given ball will launch at a speed of 59 to 69 km/h?

Processes

Selecting Tools and Computational Strategies
When dealing with large amounts of data, some methods are easier than others. What tools did you select to solve this problem? What were your reasons for the selection?

7. Refer to the tennis ball speed data in #6.

   a) Use an interval width of 2 km/h, starting at 49 km/h. Construct a frequency table.

   b) Use your frequency table to construct a frequency histogram.

   c) Add a frequency polygon to the frequency histogram.

   d) Estimate the mean.

   e) Which interval width made it easier to estimate the mean speed? Give a reason for your answer.

   f) Inspect the two frequency polygons. Which gives a clearer picture of the actual frequency distribution? Why is this so?

8. **Application** In the sport of archery, an archer controls the horizontal deviation of the arrow from the centre of the target by aiming the bow properly from left to right. Marian shot a total of 36 arrows toward a target and measured the horizontal deviation of each impact from the centre line of the target. A negative number indicates an error to the left, while a positive number indicates an error to the right.

Horizontal Error (cm)					
–2	5	–3	0	8	0
–5	–4	–1	–5	–3	–5
–1	0	0	–2	3	3
–6	2	–4	–1	6	3
–6	–1	–7	–4	2	–3
–8	1	–2	4	6	5

   a) Use an interval width of 2 cm, starting at −10 cm. Construct a frequency table.

   b) Add a relative frequency column to your frequency table.

   c) What is the probability that the horizontal error is less than 4 cm to either side of the centre of the target?

   d) Some archers use a sight to help them aim. Suppose an archer misadjusted the sight so that it had a bias to the left. How would this show up in the data?

9. **Application** Terry's Tree Farm planted Christmas tree seedlings last year. The table shows the current heights of a sample of 20 trees.

Tree Height (cm)			
37.4	35.0	34.8	35.7
32.4	35.2	38.0	36.1
32.8	38.0	38.5	37.6
32.0	36.3	31.1	35.5
30.4	34.1	36.2	35.8

   a) Determine the mean height.

   b) Determine the standard deviation of the heights.

   c) Are there enough data to predict whether the distribution of the heights follows a normal distribution? Explain. Include a graph or table as part of your answer.

10. A student from region A scored 400 on a standardized math test with a mean of 350 and a standard deviation of 35. A student from region B scored 67 on a standardized math test with a mean of 62 and a standard deviation of 5. Both are being considered for a scholarship at the same university.

a) How can the university decide which candidate is the better student?

b) What assumptions must the university make?

c) Apply the method in part a), and determine which candidate is the better student. Give a reason for your answer.

11. **Application** In many cases, variables follow a distribution that falls off symmetrically to either side of a central maximum. Examples include heights of girls of the same age, lengths of parts produced by a machine, or blood pressures of a sample of patients. The British statistician Karl E. Pearson (1857–1936) suggested calling this curve the "normal distribution."

a) Suggest a practical example of a variable that you think should follow a normal distribution.

b) Search the Internet for measured values of your choice.

c) Copy the values into an appropriate technology.

d) Determine the mean and standard deviation of the data.

e) Select a value and determine the z-score for that value.

f) Determine the probability that the variable has a value greater than or equal to the value chosen in part e).

✓ Achievement Check

12. The Tuwheeler motorcycle factory produces pistons for motorcycle engines. The table shows the piston diameters for a production run of 20 pistons. The factory rejects pistons that have a diameter of less than 8.38 cm or more than 8.42 cm.

Diameter (cm)			
8.42	8.40	8.37	8.39
8.38	8.39	8.44	8.41
8.42	8.43	8.43	8.42
8.42	8.41	8.40	8.39
8.41	8.38	8.41	8.38

a) Determine the mean of these data.

b) Determine the standard deviation of these data.

c) Determine the z-scores that limit the acceptable range of diameters.

d) What is the probability that a given manufactured piston falls outside the acceptable range?

e) In a production run of 500 pistons, how many would be expected to be unacceptable?

13. Application A coal-fired power plant releases some radioactive substances into the air. For example, a 1000-MW coal plant releases a mean of 5200 kg of uranium per year, with a standard deviation of 1300 kg. The release of uranium follows a normal distribution.

a) What is the probability that the plant will release less than 4000 kg of uranium in a given year?

b) What is the probability that the plant will release more than 6000 kg of uranium in a given year?

c) What is the probability that the release will be within 10% of the mean?

Extend

14. Application A teacher has taught Grade 12 Mathematics for 25 years. He has kept records indicating that the long-term mean mark in his class is 72, with a standard deviation of 5. In his current class, the mean is 65, with a standard deviation of 7. He suspects that his tests have been harder this year, and decides to adjust all of the marks so that the mean becomes 72 with a standard deviation of 5. This process is known as "belling the marks," or "marking on the curve."

a) Suppose that you received a mark of 70 in this year's class. Devise a method to determine your mark after adjustment. Explain your method.

b) What is your adjusted mark?

c) Can you suggest any other reasons why the marks this year might be lower than expected?

15. Thinking Recall the soft drink volume data from Example 1. These data are already grouped, and the volumes of individual bottles are not known.

Volume (mL)	Frequency, *f*
490–492	0
492–494	0
494–496	2
496–498	11
498–500	43
500–502	81
502–504	48
504–506	14
506–508	1
508–510	0

a) How can you determine a more accurate estimate for the mean of these data? Devise an approach, and explain why you think it will work.

b) Carry out your plan. What is the mean?

c) Review the frequency polygon for these data. Does your mean make sense? Give a reason for your answer.

7.3

Applications of the Normal Distribution

Learning Goals

I am learning to

- recognize the general characteristics of a normal distribution
- use technology to simulate a normal distribution in order to investigate its properties
- determine probabilities for a normal distribution

Minds On...

In the fall, apple picking is a popular activity in many orchards across Ontario.

- Would you expect that all of the apples on a tree have the same mass?
- Are the masses of the apples distributed normally? How can you tell?
- Are all distributions with a central maximum and a bell-like shape normal distributions?

Action!

Investigate a Normal Distribution Using a Simulation

Materials

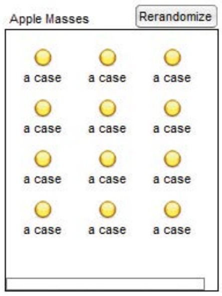
- computer with Fathom™

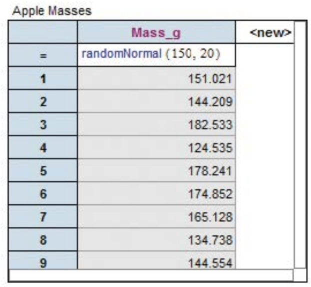
You can use technology such as Fathom™ or a graphing calculator to generate random values from a normal distribution. You can then increase the number of values to see what effect this has on the characteristics of the distribution.

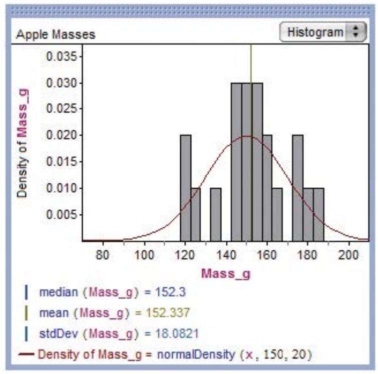
A variety of apple has a mean mass of 150 g, with a standard deviation of 20 g. This is the underlying normal distribution.

Suppose you pick 20 of these apples from a tree, and determine the mean and standard deviation of your sample, $\bar{x}$ and s. How will these compare to the mean and standard deviation of the underlying normal distribution, μ and σ?

1. Open the Fathom™ file **AppleMasses.ftm** provided by your teacher. The data in this fil generated randomly from a distribution with a mean of 150 and a standard deviation of 20. You will see a collection, a case table with 20 random values from a normal distribution with a mean of 150 and a standard deviation of 20. You will also see a histogram of the data with the mean and median values displayed.







   a) Do the data appear to lie around a mean of 150 g?

   b) Do most of the values fall within two standard deviations?

   c) How do the mean and standard deviation of the sample compare with those from the underlying normal distribution?

   d) Does the histogram approximate a normal distribution, with values symmetric about a central maximum? How do the mean and median compare?

2. Press the **Rerandomize** button. This will select 20 different values for the masses of the apples. Press the **Rerandomize** button several times. Watch the effect on the graph and the measures.

   a) What changes do you see in the graph and in the measures? How do the mean and median compare?

   b) With a small number of values like this, how well does the sample approximate the underlying normal distribution?

3. Right click on the collection box, and select **New Cases…**. Add 80 more values for a total of 100. If necessary, adjust the scales of the axes on the graph.

   a) What differences can you see in the histogram with 100 cases?

   b) Press the **Rerandomize** button several times. Observe what happens to the measures and the histogram. How do the mean and median compare?

4. Repeat #3 for a total of 1000 cases, then a total of 10 000 cases. You can add only 5000 cases at a time, so you will need to perform this step twice to obtain a total of 10 000 cases.

5. **Reflect** Summarize how the number of data points in the sample affects the shape of the frequency histogram and the measures of sample mean and sample standard deviation.

6. **Extend Your Understanding** Suppose a sample of apples is taken from a different orchard, and more values appear to the left of the central peak than the right—that is, the distribution is skewed to the right. How would the values of the mean and median compare? How would they compare if the distribution were skewed to the left?

Example

Values Within One Standard Deviation of the Mean

Suppose the birth mass of a breed of guinea pig follows a normal distribution with a mean of 100 g and a standard deviation of 10 g.

a) What is the probability that a birth mass from a large sample lies within one standard deviation of the mean? two standard deviations?

b) Does the answer to part a) depend on the mean and standard deviation of the distribution?

Solution

a) Method 1: Use a *z*-Score Table

One standard deviation below the mean results in a z-score of -1, while one above the mean results in a z-score of $+1$. Two standard deviations below the mean results in a z-score of -2, while two above the mean results in a z-score of $+2$. Use the table on pages 480–481 to determine the probability that a value lies in each range.

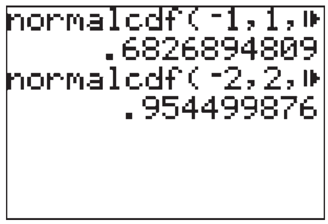
$$\begin{aligned} &P(-1 \le z \le 1) \\ &= P(z \le 1) - P(z \le -1) \\ &= 0.8413 - 0.1587 \\ &= 0.6826 \end{aligned}$$

$$\begin{aligned} &P(-2 \le z \le 2) \\ &= P(z \le 2) - P(z \le -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \end{aligned}$$

The probability that a value lies within one standard deviation of the mean is about 68%, and within two standard deviations is about 95%.

Method 2: Use a Graphing Calculator

Use the **normalcdf(** function from the **DISTR** menu by inputting **normalcdf(−1,1,0,1)** and **normalcdf(−2,2,0,1)**. The probability that a value lies within one standard deviation of the mean is about 68% The probability that a value lies within two standard deviations of the mean is about 95%.



b) Since one standard deviation from the mean will always result in z-scores of -1 and $+1$, the answer to part a) does not depend on the mean and standard deviation.

Your Turn

A sample of male patients at a hospital showed a mean systolic blood pressure of 124.7 mmHg with a standard deviation of 14.5 mmHg.

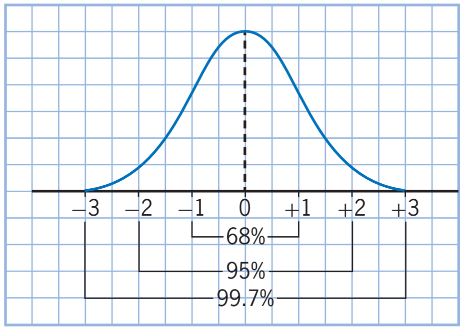
a) Use the z-score table on pages 480–481 to determine the probability that a measurement from a large sample lies within three standard deviations of the mean.

b) Use technology to determine the probability that a measurement from a large sample would lie within three standard deviations of the mean.

Consolidate and Debrief

Key Concepts

- You can use the normal distribution to model the frequency and probability density distributions of continuous random variables.
- The normal distribution has a central peak, and is symmetric about the mean.
- The mean and median are equal.
- About 68% of the data values are within one standard deviation of the mean, about 95% of the data values are within two standard deviations of the mean, and about 99.7% are within three standard deviations of the mean.



Reflect

R1. A factory produces bolts with a mean length of 5.0 cm and a standard deviation of 0.1 cm. You select a random sample from a batch and find that it has a length of 4.7 cm. Is this a surprising value? Explain your answer.

R2. Machine A fills 1000 one-kilogram honey jars using 1200 kg of honey. Machine B fills 1000 jars using 1050 kg of honey. When the jars are tested, 0.1% of the jars from each machine contain less than 1 kg of honey. Are these results possible? Explain your answer, using diagrams to help you.

Practise

Choose the best answer for #1 and #2.

1. Which of these situations might be reasonably expected to follow a normal distribution?

A the heights of students in grade 3 at an elementary school

B the mass of peanut butter in a sample of jars marked "1 kg"

C the distance that a candidate for the football team can throw a football

D all of these

2. Which statement is true concerning a normal distribution?

A The curve is symmetrical about a central peak.

B The median is always less than the mean.

C All of the data values will occur within two standard deviations of the mean.

D The mean of a sample always matches the mean of the underlying normal distribution.

3. A breed of adult cat has a mean mass of 4.2 kg with a standard deviation of 0.5 kg. An article in a pet magazine claims that 1 out of 40 such cats will have a mass of more than 5.2 kg. Does this make sense? Explain.

4. **Communication** A factory produces steel washers that will fit onto standard bolts. Why would a quality control engineer be interested in the mean and standard deviation of these washers?

Apply

5. **Application** A tire company produces bicycle tires with a mean outside diameter of 700 mm and a standard deviation of 13.2 mm.

a) Use technology to model a random sample of 10 tires from this company.

b) Determine the mean and standard deviation of your sample.

c) How closely do your mean and standard deviation match those of the underlying normal distribution?

d) Model random samples of 100 tires and 1000 tires.

e) Determine the sample mean and standard deviation for each sample in part d).

f) Summarize the values of $\bar{x}$, s, μ, and σ in a table for all three samples.

g) How do the values of the sample measures compare to the underlying normal distribution measures as the size of the sample increases?

6. To play an octave on a piano, your hand must span a distance of 16.4 cm. A sample of music students showed a mean hand span of 21.8 cm, with a standard deviation of 2.4 cm.

a) What is the probability that a student could not play an octave?

b) What is the probability that a student could play one and one-half octaves?

Literacy Link

If you start on any white key on a piano, such as C, and count up or down seven white keys, you will reach another C. This interval is known as an *octave*, because it contains eight notes.

7. Lithium cells used to power digital watches have a mean lifetime of 8 years with a standard deviation of 1.5 years. The Tempus Watch Company sells a sealed model of watch with power cells that are not replaceable. The watch costs $9.95, and the company offers a 5-year replacement warranty if the cell fails.

a) What percent of its watches would the company expect to replace under a 5-year warranty plan?

b) Assume that the company sells 100 000 watches this year, and it costs $5.00 to replace a defective watch, including shipping. How much should the company budget to replace watches under warranty?

c) An advertising executive suggests boosting sales by offering a 10-year warranty. Is this a reasonable idea? Use calculations similar to parts a) and b) to provide support for your answer.

8. The mean mark on a mathematics exam was 70%, with a standard deviation of 10%.

a) If 200 students wrote the exam, how many would be expected to score between 60% and 80%?

b) How many students would be expected to score above 90%?

c) How many students would be expected to score below 50%?

9. Thinking Precision racing pistons for a high-performance automobile engine need to have a diameter of 5 in. accurate to within 0.0001 in. A parts manufacturing company would like 95% of its pistons to fall within this range. What standard deviation is needed to meet this requirement?

✓ Achievement Check

10. Frank's Footlongs sells hot dogs at the beach. Analysis of his sales shows that the number of hot dogs sold on a given day follows a normal distribution with a mean of 120 and a standard deviation of 11.

a) Is it reasonable to treat these data as if they formed a continuous distribution? Give a reason for your answer.

b) What is a reasonable minimum number of hot dogs and buns for Frank to have on hand at the beginning of the day? Give reasons for your answer.

c) What is the probability that the stand will sell fewer than 25 hot dogs on a given day?

d) Frank's season lasts from May 1 until September 30. On how many days should he expect to sell between 100 and 140 hot dogs?

e) Should he expect to sell 200 or more hot dogs on at least one day during the season? Support your answer with numbers.

11. Application Honey jars from the farm where Doris works say they contain 500 g of honey. The table shows the actual amounts from a sample of 30 jars.

Mass of Honey (g)					
503	505	504	500	502	505
506	502	501	501	503	501
502	506	504	505	505	503
504	501	499	501	502	504
502	500	501	503	506	504

a) Determine the mean and standard deviation of the sample of honey jars.

b) What percent of the data in the table actually fall within one standard deviation of the mean?

c) How does the answer in part b) compare to the expected percent of the data within one standard deviation of the mean?

Extend

12. Rudy's Sandwich Shoppe sells corned beef sandwiches advertised to contain 200 g of corned beef. Rudy has set his slicing machine to a mean of 220 g. The actual amount follows a normal distribution. He ran a quality control check, and found that 5% of his sandwiches contained less than 200 g of corned beef.

a) What was the standard deviation of the amount of corned beef in the sandwiches?

b) Rudy would like to ensure that no more than 0.5% of the sandwiches contain less than 200 g of corned beef. Suggest two different actions that he can take.

c) From a business point of view, which action in part b) would be more desirable?

13. Communication Pafnuty Chebyshev was a Russian mathematician who worked in the field of statistics during the middle of the 19th century. Chebyshev's Theorem states that no more than $\frac{1}{k^2}$ of the values of a distribution lie more than k standard deviations from the mean. The value of k must be greater than 1. This theorem applies to all distributions, not just the normal distribution.

a) What is the maximum proportion of values that may lie more than two standard deviations from the mean?

b) Does this theorem agree with what you know about the normal distribution? Explain your answer, with examples.

c) Write an expression for the proportion of values that must lie within k standard deviations of the mean.

7.4

Confidence Intervals

Learning Goals

I am learning to

- distinguish among the meanings of common confidence levels such as 90%, 95%, and 99%
- determine the margin of error for a population mean estimated using a sample
- determine the upper and lower limits of the confidence interval

Minds On...

Fishing is a popular activity in parks such as Algonquin Provincial Park. Many of the lakes in the park contain brook trout as well as lake trout. It is important to collect data on the fish populations and watch for any significant or sudden changes that may indicate problems with the health of the marine ecosystem.

- How can wildlife researchers determine reasonable estimates of the number and species distributions in a large lake?
- Why is it not possible to obtain exact numbers?
- How confident can researchers be that their estimates are close to the correct values?

Action!

margin of error

- the range of values that a particular measurement is said to be within
- the smaller the margin of error, the greater the accuracy of the measurement

confidence interval

- the range of possible values of the measured statistic at a particular confidence level

A researcher might summarize the findings as "Lake trout make up 20.4% of the fish population in Lake Lavieille. This estimate is considered correct within ±3.0%, 19 times out of 20."

What does this statement mean?

Suppose the researcher samples the fish in the lake 20 times. The researcher is confident that the percent of lake trout found in the samples will be within 3% of 20.4% in 19 out of the 20 samples. The **margin of error** is 3% above or below the mean.

The **confidence interval** range is

$20.4\% - 3\% = 17.4\%$ to $20.4\% + 3\% = 23.4\%$

The **confidence level** is 19 out of 20, or 95%. How does this percent connect to the standard deviation of a normal distribution?

Investigate Simulating Fish Populations Using Game Chips

Complete this investigation as a class.

1. Use a bag containing 200 game chips. The bag contains a number of blue chips preselected by the teacher but unknown to the students, as well as other chips to make up a total of 200 in each bag.

2. Ensure that the chips are well mixed. Reach into the bag and pull out 30 chips at random, replacing the chosen chip each time. This is the sample size, n. Record the total number of blue chips chosen in the 30 trials. Determine the experimental probability, p, of getting a blue chip. Express p as a decimal.

Confidence Level	*z*-Score
90%	1.645
95%	1.960
99%	2.576

3. Calculate the margin of error, as a decimal, from the formula
$$E = z\sqrt{\frac{p(1-p)}{n}}$$
Use the z-score for a confidence level of 95%.

4. Use the margin of error to determine the confidence interval. Express the probability of getting a blue chip and the confidence interval as percents.

5. Have several class members take turns pulling 30 chips from the bag with replacement. Record the number of blue chips chosen in every 30 trials. Pool the class results for the total number of blue chips drawn. Determine the margin of error and the confidence interval for the class results using a confidence level of 95%.

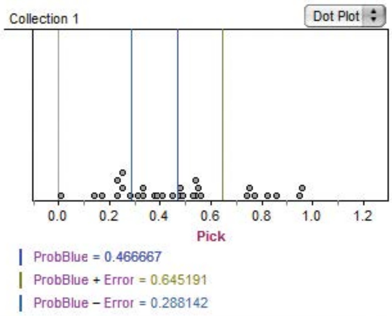
6. **Reflect** Compare the individual confidence interval from #4 with that for the class results in #5. Why does the confidence interval decrease as the sample size n increases?

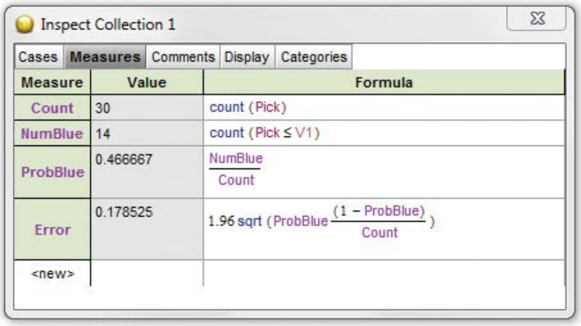
7. **Extend Your Understanding** You could use technology such as Fathom™ to simulate a large number of experiments.
   a) Open the Fathom™ file **FishSample.ftm** provided by your teacher. This data file simulates drawing 30 chips from a population with a proportion of blue chips equal to the value of p on the slider. The values plotted on the graph represent the proportion of blue chips in your sample, plus or minus the margin of error for a confidence level of 95%.
   b) Rerandomize the sample. How do the upper and lower bounds on the confidence interval change? Does the true proportion fall inside of your confidence interval?
   c) Right click on the sample and add another 30 cases. How do the upper and lower bounds on the confidence interval change?
   d) Add another 140 cases. How does the dot plot change? How do the upper and lower bounds on the confidence interval change?

Materials
- bag
- 200 game chips (blue and other colours)

confidence level
- the probability that a particular statistic is within the range indicated by the margin of error
- commonly used confidence levels are 90%, 95%, and 99%

e) Summarize the effect of increasing the number in the sample on the margin of error and the confidence interval. A sample Fathom™ screen for a 40% probability of getting a blue chip is shown.





Measure	Value	Formula
Count	30	count (Pick)
NumBlue	14	count (Pick ≤ V1)
ProbBlue	0.466667	$\frac{\text{NumBlue}}{\text{Count}}$
Error	0.178525	1.96 sqrt $\left(\text{ProbBlue}\frac{(1-\text{ProbBlue})}{\text{Count}}\right)$
<new>		

The minimum sample size required to obtain reliable results for the confidence interval is usually at least 30.

Example 1

Lake Trout in Lake Lavieille

Lake Lavieille is one of the largest lakes in Algonquin Park. In 2009, 234 lake trout were caught out of a total catch of 911. In 2012, 141 lake trout were caught out of a catch of 689.

a) Determine the percent of lake trout caught for each year.

b) Determine the margin of error for each year. Use a 95% confidence level.

c) Determine the confidence interval for each year.

d) Do the two confidence intervals overlap? Give a specific example.

e) Is it reasonable to conclude that the percent of lake trout in Lake Lavieille decreased from 2009 to 2012? Give a reason for your answer.

Solution

a) For 2009:

$$p = \frac{234}{911} \approx 0.257$$

For 2012:

$$p = \frac{141}{689} \approx 0.205$$

The percent of lake trout caught is about 25.7% for 2009 and 20.1% for 2012.

b)

Confidence Level	*z*-Score
90%	1.645
95%	1.960
99%	2.576

For 2009:

$$E = z\sqrt{\frac{p(1-p)}{n}}$$
$$\approx 1.96\sqrt{\frac{0.257(1-0.257)}{911}}$$
$$\approx 0.028$$

For 2012:

$$E = z\sqrt{\frac{p(1-p)}{n}}$$
$$\approx 1.96\sqrt{\frac{0.205(1-0.205)}{689}}$$
$$\approx 0.030$$

The margin of error is about 2.8% for 2009 and 3.0% for 2012.

c) For 2009:

Lower limit = 25.7% − 2.8%
= 22.9%

Upper limit = 25.7% + 2.8%
= 28.5%

For 2012:

Lower limit = 20.5% − 3.0%
= 17.1%

Upper limit = 20.5% + 3.0%
= 23.1%

The confidence interval is 22.9% to 28.5% for 2009 and 17.1% to 23.1% for 2012.

d) The confidence intervals overlap; for example, both include 23.0%.

e) A 95% confidence level is used. The two confidence intervals overlap, meaning that the 2012 data could fall within the 2009 confidence interval. You cannot definitely conclude that the percent of lake trout has decreased.

How would the resulting intervals change if you used a different confidence interval?

Your Turn

An opinion poll surveyed 100 households who were watching television at a particular time. Of these, 75% were watching *Hockey Night in Canada*.

a) Determine the margin of error at a 99% confidence level.

b) Determine the confidence interval for this situation.

c) How would a news source state the results?

Example 2

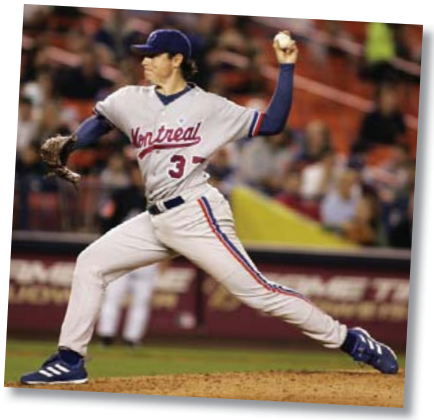
Interpreting Survey Results

The first Major League Baseball franchise outside of the United States was the Montréal Expos, who played from 1969 to 2004. In 2005, the Expos moved to Washington, DC, and are now known as the Washington Nationals. A survey was conducted to determine whether a major league baseball team should come back to Montréal. Of the 1589 people surveyed, 69% were in favour of baseball coming back.

a) Determine the margin of error for this survey at a confidence level of 95%.

b) For what range of percents can you be 95% confident that people would be in favour of baseball returning to Montréal?

c) A second survey at a confidence level of 95% found that 56% were in favour, with a margin of error of 5.2%. Approximately how many people were surveyed?



Solution

a) In the first survey, $p = 69\%$.

$$E = z\sqrt{\frac{p(1-p)}{n}}$$

$$E = 1.96\sqrt{\frac{0.69(1-0.69)}{1589}}$$

$$E \approx 0.023$$

The margin of error is $\pm 2.3\%$.

b) The lower end of the range is $69\% - 2.3\% = 66.7\%$. The upper end of the range is $69\% + 2.3\% = 71.3\%$. There is a 95% chance that the percent of people in favour of another major league baseball team in Montréal is between 66.7% and 71.3%.

c) In the second survey, $p = 56\%$.

The margin of error is $E = 5.2\%$.

$$E = z\sqrt{\frac{p(1-p)}{n}}$$ Rearrange to solve for n.

$$E^2 = z^2\left(\frac{p(1-p)}{n}\right)$$ Square both sides to eliminate the square root.

$$n = \frac{z^2(p(1-p))}{E^2}$$ Multiply both sides by n. Divide both sides by E^2.

$$n = \frac{(1.96)^2(0.56(1-0.56))}{0.052^2}$$ Substitute and evaluate.

$$n \approx 350$$

The second sample surveyed about 350 people.

Your Turn

A pharmaceutical manufacturer makes more than 500 000 pills of a certain drug each day. The company randomly samples 400 pills daily to check that they meet the proper weight and size standards. On a given day, 52 pills were found to be substandard.

a) What is the margin of error for this sample at a confidence level of 90%?

b) If the company would like to cut the margin of error in half, how would the sample size have to change?

Repeated Sampling

Suppose that samples of the same size are repeatedly taken from a population that follows a normal distribution with a mean of μ and a standard deviation of σ. The means of the samples will be normally distributed with a mean of μ and a standard deviation of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Consider the role of the sample size n. If $n = 1$, the standard deviation of the sample means is the same as the standard deviation of the normal distribution. However, as the value of n increases, the standard deviation of the sample means decreases. The usual rule is that a value of $n = 30$ produces a reasonable estimate of the population mean.

You can calculate the margin of error for a sample mean from the formula $E = z\frac{\sigma}{\sqrt{n}}$ or $E = z\sigma_{\bar{x}}$. This is also known as the standard error.

Example 3

Determining Confidence Levels for a Sample Mean

At an agricultural fair, the masses of 8 giant pumpkins entered in a contest were 11 kg, 13 kg, 15 kg, 18 kg, 12 kg, 14 kg, 10 kg, and 16 kg. Results from past fairs suggest that the masses are normally distributed with a mean of 14.2 kg and a standard deviation of 2.5 kg. Determine a 90% confidence interval for the sample mean.

Project Prep

Does your project include continuous variables? Can you calculate and apply confidence intervals?

Solution

Calculate the sample mean and standard deviation of the sample means.

$$\begin{aligned}\bar{x} &= \frac{11 + 13 + 15 + 18 + 12 + 14 + 10 + 16}{8} \\ &\approx 13.6\end{aligned}$$

$$\begin{aligned}\sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{2.5}{\sqrt{8}} \\ &\approx 0.88\end{aligned}$$

For a 90% confidence level, $z = 1.645$.

$$\begin{aligned} \text{Lower limit} &= \overline{x} - E \\ &= \overline{x} - z\sigma_{\overline{x}} \\ &\approx 13.6 - (1.645)(0.88) \\ &\approx 12.2 \end{aligned}$$

$$\begin{aligned} \text{Upper limit} &= \overline{x} + E \\ &= \overline{x} + z\sigma_{\overline{x}} \\ &\approx 13.6 + (1.645)(0.88) \\ &\approx 15.0 \end{aligned}$$

The 90% confidence interval for the mean mass is 12.2 kg to 15.0 kg. This means that with 90% confidence, you can predict that the mean mass of a pumpkin lies within the interval from 12.2 kg to 15 kg.

Your Turn

A consumers' group tested batches of light bulbs to see how long they lasted. The results, in hours, from one batch were 998, 1234, 1523, 1760, 937, 1193, 996, 1002, 986, 1285, 1163, and 1716. The manufacturer claims that the life of the light bulbs is normally distributed with a mean of 1200 h and a standard deviation of 420 h.

a) Calculate the mean of the sample and the standard deviation for the sample means.

b) Determine the 99% confidence interval for the sample mean.

Consolidate and Debrief

Key Concepts

- The confidence level is the probability that a particular statistic is within the range indicated by the margin of error.
- Commonly used confidence levels are 90%, 95%, and 99%. These are related to the z-scores of the distribution.
- A margin of error is the range of values that a particular statistic is said to be within. For a statistic with probability p, the margin of error is $E = z\sqrt{\dfrac{p(1-p)}{n}}$.
- The greater the sample size, the smaller the margin of error. The smaller the margin of error, the greater the accuracy of the measurement.
- The confidence interval is the range of possible values of the measured statistic.
- For repeated samples of the same size taken from the same population with a normal distribution, the standard deviation of the sample means is $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}$ and the margin of error is $E = z\dfrac{\sigma}{\sqrt{n}}$.

Reflect

R1. Can you use the terms "confidence level" and "margin of error" interchangeably? Give reasons for your answer.

R2. A computer manufacturer found that a mean of 5.6% of a model of tablet computers were returned as defective within one year. The service department considered this number accurate within 1.4%, 9 times out of 10. What information does the confidence interval provide? What is the benefit of using the confidence interval?

R3. A 95% confidence interval can be stated as 19 times out of 20. Restate a 90% and a 99% confidence interval in a similar manner.

Practise

Choose the best answer for #2 and #3.

1. Common confidence levels are 90%, 95%, and 99%. A sample is taken from a population with a given mean and standard deviation. Within approximately how many standard deviations of the mean will the values in the confidence interval lie for a

a) 90% confidence level?

b) 95% confidence level?

c) 99% confidence level?

2. A poll conducted by the student newspaper found that 78% of the students who ate lunch at the school cafeteria ordered a salad at least twice per week. The poll is considered accurate within ±5%, 17 times out of 20. What is the confidence level for the poll?

A 75%

B 85%

C 95%

D 98%

3. During a municipal election the local newspaper polled 251 people. The paper reported that 57% said they were in favour of candidate A for mayor. The result was considered accurate within 6.1%, 19 times out of 20. Which of the following statements is false?

A The margin of error is 6.1%.

B In a similar poll, 95% of the time between 50.9% and 63.1% of the people would be found in favour.

C The confidence interval is 57% ± 6.1%.

D In a similar poll, 95% of the time 57% or more of the people would be found in favour.

4. An automobile dealer offers a new line of tires. The tires last a mean of 75 000 km with a standard deviation of 5000 km, following a normal distribution. The tire life experienced by 100 customers is recorded. What is the expected standard deviation of the sample means?

Apply

5. The Canadian commercial pilot written exam consists of 100 multiple choice questions. Last year, the students enrolled in a community college aviation course recorded a mean mark of 82% among 25 candidates.

a) Determine the margin of error at a 99% confidence level.

b) Determine the confidence interval for the exam marks.

c) State the results in the usual format for the course newsletter.

6. **Communication** A political party received an average of 34% support in recent polls plus or minus 3.4%, 19 times out of 20. Two subsequent polls showed 38% support and 27% support. How would you report on the meaning of these polls to the party membership?

7. **Application** A Single Crème cookie is made using a cream filling between two wafers. The amount of cream follows a normal distribution with a mean of 25 g and a standard deviation of 2.0 g. The company claims its new Double Crème line contains twice the amount of filling. A random sample of 20 such cookies were found to contain cream content as shown.

Mass of Cream (g)				
48.9	47.3	47.3	45.5	52.9
50.1	46.0	47.9	48.5	48.2
47.5	51.9	49.7	47.8	50.1
46.9	51.0	45.9	45.4	47.1

a) Calculate the mean of the sample and the standard deviation for the sample means. What assumption must you make?

b) Determine the 95% confidence interval for the sample mean.

c) Is the company justified in claiming that the Double Crème line contains twice the filling of the Single Crème line? Give reasons for your answer.

8. A concrete manufacturer knows from experience that setting times for concrete follow a normal distribution with a standard deviation of $\sigma = 8.5$ min. The manufacturer wants to use the slogan "Our concrete quick-sets in t minutes" in its advertising campaign. A technician pours 25 test squares of equal size and finds the mean setting time to be 72.2 min.

a) Determine a 95% confidence interval for the actual mean setting time of the concrete.

b) Advise the manufacturer on a reasonable value for t. Give a reason for your answer.

9. A honey farm rates its honey on a colour scale from 1 to 20, ranging from very light orange to deep orange. The colour of honey follows a normal distribution with a standard deviation of 2.5. A technician tests a sample of 50 jars of honey, resulting in a sample mean of 12. Determine a 95% confidence interval for the colour of the honey.

10. A survey of businesses showed that a mean of 17% of gross income was spent on office overhead, with a standard deviation of 5%, following a normal distribution. At a 99% confidence level, the margin of error was 10.5%. How many businesses were surveyed?

11. As part of Earth Day celebrations, an environmental scientist participated in a program to measure water clarity in 70 locations in Lake Ontario using a clarity measuring disk. The scientist reported that the lake water was clear to a mean depth of 5 m with a standard deviation of 1 m. The margin of error was 0.20 m. What confidence level was used?

12. A study of patients with low-back pain reported that the sample mean duration of the pain was 18.3 months. The duration follows a normal distribution in the population with a standard deviation of 5.9 months. The margin of error for the population mean was 1.5 months at a 95% confidence level. How many patients were in the study?

13. A manufacturer of computer hardware knows that the life of its hard drives follows a normal distribution with a standard deviation of 2400 h. Over the past three years, the mean life, based on a sample of 900 hard drives, was 16 000 h of use. What is the 95% confidence interval for the mean life of a hard drive?

✓ Achievement Check

14. When tomato farmers harvest their crop, they use an automated tomato picker that separates the tomato from the vine and eliminates any non-ripe tomatoes. The process of eliminating the non-ripe tomatoes is not perfect due to the speed with which the tomatoes are actually picked. Some green tomatoes get into the load. The buyer takes a single scoop of tomatoes from a random spot in the load. A sample of 300 tomatoes contains 92% of acceptable standard.

a) What is the confidence interval of the load if the confidence level is 95%?

b) The current load measures 41.992 tonnes. What is the interval for the mass of acceptable tomatoes?

c) Tomatoes sell for $94.40 per tonne. What is the range of values that this load is worth?

d) A load is accepted as long as the sample contains at least 67% acceptable tomatoes. How does the range of values change for a load at this level?

Literacy Link

The term *tonne* refers to the SI or metric unit equal to 1000 kg. The term *ton* refers to the imperial unit equal to 2000 lb.

15. A friend shows you a caption from the school newspaper:

"Two recent polls show that 57% of students would vote for Adam and 51% would vote for Meghan in the next student council election. These results are accurate to within ±3%, 19 times in 20."

Explain what the caption means.

Extend

16. **Thinking** Suppose you select 30 men from the entire population of 20-year-old men, and measure their weights.

a) Are you more likely to end up with men close to the population mean or far away from the population mean? Give a reason for your answer.

b) What effect does this have on the standard deviation of the sample mean compared to the population mean?

c) How does the formula for the standard deviation of the sample means, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, fit with the effect in part b)?

17. Consider the formula for the margin of error: $E = z\sqrt{\frac{p(1-p)}{n}}$. A poll is being conducted to determine the percent of voters who will vote for candidate Alpha with a confidence level of 95%. What is the minimum number of voters who must be interviewed to guarantee a margin of error of no more than 2%, regardless of the value of p? Explain your reasoning.

7.5

Connections to Discrete Random Variables

Learning Goals

I am learning to

- make connections between a normal distribution and a binomial distribution
- make connections between a normal distribution and a hypergeometric distribution
- recognize the role of the number of trials in these connections

Minds On...

Recall from Chapter 4 that tossing a coin several times and recording the number of heads obtained is an example of a binomial distribution.

- How does the shape of the distribution depend on the number of times the experiment is tried?
- Predict the form of a graph of the number of heads possible when a coin is flipped five times. Make a sketch of your prediction.

Action!

Investigate 1 Compare the Binomial Distribution to the Normal Distribution

Materials

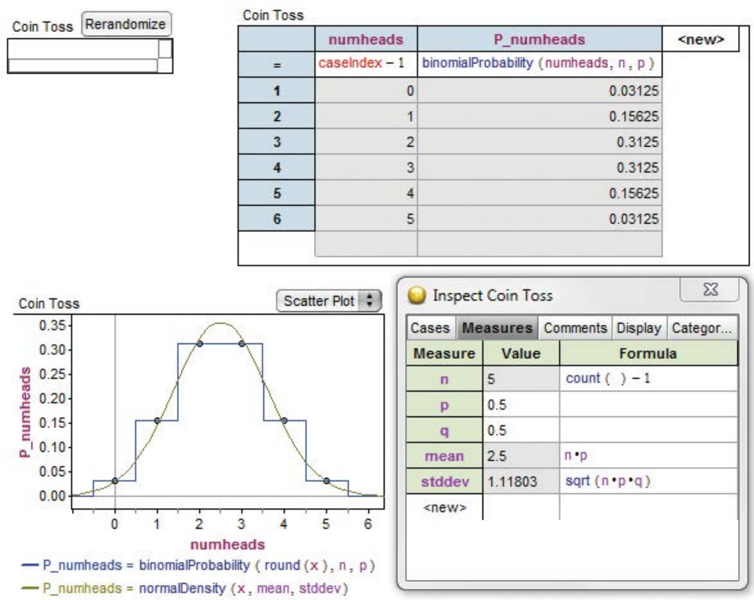
- computer with Fathom™

A coin is tossed five times, and the number of heads is recorded. What does the probability distribution look like? How does its appearance change as the number of tosses is increased?

1. Open the Fathom™ file **CoinToss.ftm** provided by your teacher. This file shows the probability distribution for the number of heads flipped on five trials. The binomial probability for each possible number of heads is shown in blue. The approximate normal distribution is shown in green superimposed over the binomial distribution.

To approximate the binomial distribution with a normal distribution, you can calculate the mean from the formula $\mu = np$, and the standard deviation from the formula $\sigma = \sqrt{npq}$. The derivation of these formulas is beyond the scope of this course.

2. How well does the normal distribution match the binomial distribution?



3. Right click on the collection box, and add 5 new cases. Adjust the scales on the axes of the graph if necessary. How is the fit with 10 tosses of the coin?

4. Right click on the collection box, and add 10 new cases. Adjust the scales on the axes of the graph if necessary. How is the fit with 20 tosses of the coin?

5. **Reflect** How does the fit of the normal distribution to the binomial distribution depend on the number of trials? Use your Fathom™ simulation to try a larger number of trials, such as 100 and then 1000.

6. **Extend Your Understanding** Suppose that a weighted coin is used in the coin toss. Now the probability of getting a head is only 0.3.

   a) Predict the effect on the binomial distribution for a large number of trials. How would it compare to the normal distribution?

   b) Adjust the probability values for p and q in the Fathom™ simulation, and check your prediction. Were you right? Explain your reasoning.

You can "step back" the Fathom™ simulation by choosing **Undo** from the **Edit** menu. Alternatively, you can use the keyboard shortcut **CTRL-z**.

Investigate 2 Compare the Hypergeometric Distribution to the Normal Distribution

Materials

- computer with Fathom™

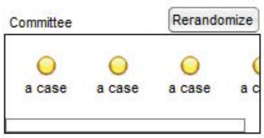
A committee of 4 is chosen from a group of 200 people. How many males are on the committee? How does the distribution change as the size of the committee is increased? Does the population size have any effect?

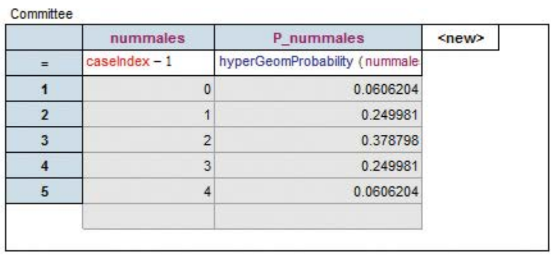
1. Open the Fathom™ file **Committee.ftm** provided by your teacher. The file shows the probability distribution for the number of males selected in a committee of 4 people from a sample of 100 males and 100 females. The hypergeometric probability for each possible number of males is shown in blue. The approximate normal distribution is shown in green over the hypergeometric distribution.

   If the sample size is small compared to the population size, the probability of selecting a male is approximately equal to the number of males divided by the population size.

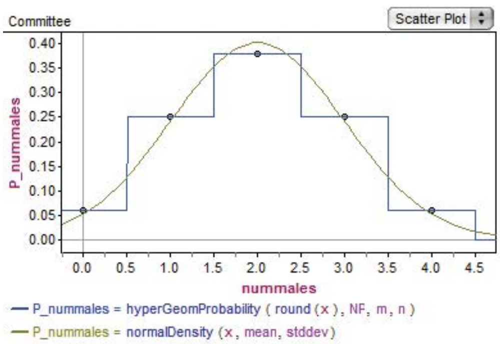
   To approximate the hypergeometric distribution with a normal distribution, you can calculate the mean from the formula $\mu = np$ and the standard deviation from the formula $\sigma = \sqrt{npq\left(\frac{NP - n}{NP - 1}\right)}$. The derivation of these formulas is beyond the scope of this course.

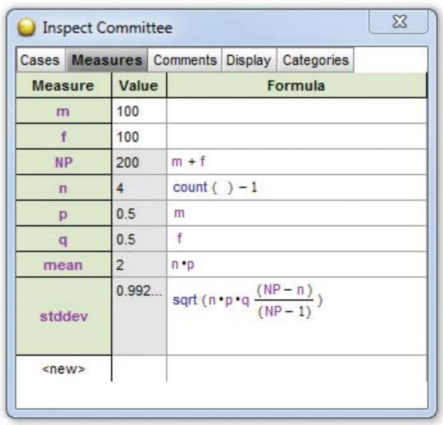
2. How well does the normal distribution match the hypergeometric distribution?




Committee

	nummales	P_nummales	<new>
=	caseIndex − 1	hyperGeomProbability (nummale	
1	0	0.0606204	
2	1	0.249981	
3	2	0.378798	
4	3	0.249981	
5	4	0.0606204	




Inspect Committee

Cases | Measures | Comments | Display | Categories

Measure	Value	Formula
m	100	
f	100	
NP	200	m + f
n	4	count ( ) − 1
p	0.5	m
q	0.5	f
mean	2	n • p
stddev	0.992...	sqrt (n • p • q $\frac{(NP - n)}{(NP - 1)}$)
<new>		

3. Change the committee membership from 4 to 10. Right click on the collection box, and add 6 new cases. Adjust the scales on the axes of the graph if necessary. How is the fit with 10 members on the committee?

4. Change the committee membership to 20. Right click on the collection box, and add 10 new cases. Adjust the scales on the axes of the graph if necessary. How is the fit with 20 members on the committee?

5. **Reflect** The committee membership must remain a small fraction of the population size, typically less than one-tenth. Why is this necessary? Consider the values of p and q in the above investigation in your response.

6. **Extend Your Understanding** Suppose that a committee of 4 were chosen in a random selection from a very small population, say a club with 6 male members and 2 female members. Can you see any problems with calculating the number of males on the committee? Consider some extreme cases. Use the Fathom™ simulation from the investigation to explore different scenarios.

You can calculate a binomial probability using the formula from chapter 4. However, this formula requires the use of factorials. If the number of successes being considered is large, this can lead to a large number of computations. You can avoid this by using the approximation.

The normal approximation for a binomial distribution is usually considered reasonable if the values of np and nq are both greater than 5: $np > 5$ and $nq > 5$.

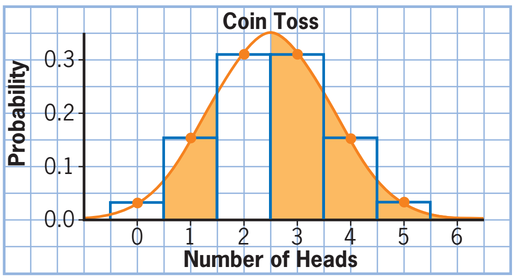
The normal approximation for a hypergeometric distribution is usually considered reasonable if the sample size is small compared to the size of the population, typically less than one-tenth: $n < \frac{1}{10}NP$, where n is the sample size and NP is the size of the population.

Since the binomial and hypergeometric distributions are discrete but the normal distribution is continuous, you must apply a **continuity correction** when using the approximation.

continuity correction

- a correction applied when using the normal approximation to correct for the difference between a discrete and continuous distribution

For example, suppose you want to determine the probability that, in 5 tosses of a coin, there is exactly 1 head. Refer to the graph shown.



The area under the normal distribution that represents this probability runs from 0.5 to 1.5. Therefore, you must calculate $P(0.5 \le X \le 1.5)$.

Similarly, if you want the probability of getting 3 or more heads, you need the area under the graph from 2.5 to infinity on the right, or $P(X \ge 2.5)$.

On the other hand, if you want to determine the probability of getting more than 3 heads, you need to calculate $P(X \ge 3.5)$.

Example 1

Normal Approximation for a Binomial Distribution

Suzette rolls a die 36 times. She records the number of times the die shows a 6.

a) Is it reasonable to approximate this distribution with a normal distribution? Give a reason for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) Determine the probability that the die will show a 6 at least 10 times.

d) What is the probability that the die will show a 6 more than 10 times, fewer than 10 times, and at most 10 times?

Solution

a)
$$np = 36 \times \frac{1}{6} = 6$$
$$nq = 36 \times \frac{5}{6} = 30$$

Recall that in a binomial distribution there are only two outcomes: success and failure. The probability of a success is p, and the probability of a failure is q.

Since both of these are greater than 5, the normal approximation is reasonable in this case.

b)
$$\mu = np = 36 \times \frac{1}{6} \approx 6$$
$$\sigma = \sqrt{npq} = \sqrt{36 \times \frac{1}{6} \times \frac{5}{6}} \approx 2.236$$

The mean is 6, and the standard deviation is 2.236.

Processes

Reflecting
Why do the two methods in part c) not give exactly the same answer? Which method is more accurate? Why? Can you get a better answer from the table by interpolating between two probabilities? Try this, and see how the two methods compare.

c) Since you want the die to show 6 at least 10 times, you must apply a continuity correction and calculate $P(X \geq 9.5)$.

Method 1:
Use a Table

Calculate the z-score for 9.5.

$$z = \frac{X - \mu}{\sigma} \approx \frac{9.5 - 6}{2.236} \approx 1.565$$

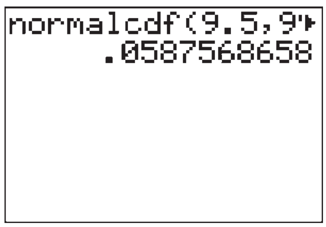
Refer to the table on pages 480–481.

$$\begin{aligned} P(X \geq 9.5) &= 1 - P(x < 9.5) \\ &= 1 - P(z < 1.57) \\ &= 1 - 0.9418 \\ &= 0.0582 \end{aligned}$$

The probability of 6 occurring 10 or more times is about 0.0582.

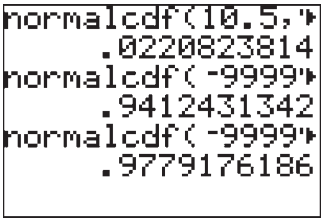
Method 2:
Use a Graphing Calculator

Use the **normalcdf(** function from the **DISTR** menu to calculate $P(X > 9.5)$ by inputting **normalcdf(9.5,999999, 6,2.236)**.



The probability of 6 occurring 10 or more times is about 0.0588.

d) If you want more than 10 successes, calculate $P(X \geq 10.5)$ by inputting normalcdf(10.5,999999,6,2.236). The probability is about 0.022.



If you want fewer than 10 successes, calculate $P(X \leq 9.5)$ by inputting **normalcdf(−999999,9.5,6,2.236)**. The probability is about 0.941.

If you want at most 10 successes, calculate $P(X \leq 10.5)$ by inputting **normalcdf(−999999,10.5,6,2.236)**. The probability is about 0.978.

Your Turn

The probability of rolling a 6 on a weighted die is 0.25. The die is rolled 25 times.

a) Is it reasonable to approximate this distribution with a normal distribution? Give reasons for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) Determine the probability that the die will show a 6 fewer than 8 times.

Example 2

Normal Approximation for a Hypergeometric Distribution

Chris works at a local daycare on a co-op work term. Chris plays a game with the children that involves pulling marbles from a bag. The bag contains 24 black marbles and 36 red marbles, well mixed. One of the children reaches in and takes out 5 marbles without looking. Chris records the number of black marbles.

a) Is it reasonable to approximate this distribution with a normal distribution? Give a reason for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) What is the probability that exactly 3 of the marbles are black?

d) What is the probability that a child pulls out at least 3 black marbles? more than 3 black marbles? fewer than 3 black marbles? at most 3 black marbles?

e) How does the answer to part c) compare with the probability calculated from the hypergeometric distribution?

Solution

a) There are 60 marbles in the bag, of which 5 are chosen. The number of trials is less than 10% of the population. The normal approximation is reasonable for this hypergeometric distribution.

b)
$$\begin{aligned}\mu &= np\\ &= 5 \times \frac{24}{60}\\ &= 2\end{aligned}$$

$$\begin{aligned}\sigma &= \sqrt{npq\left(\frac{NP-n}{NP-1}\right)}\\ &= \sqrt{5 \times \frac{24}{60} \times \frac{36}{60}\left(\frac{60-5}{60-1}\right)}\\ &\approx 1.058\end{aligned}$$

The mean is 2, and the standard deviation is 1.058.

c) To determine the probability of 3 black marbles, you must calculate $P(2.5 \le X \le 3.5)$. Use the **normalcdf** function from the **DISTR** menu by inputting **normalcdf(2.5,3.5,2,1.058)**.

```
normalcdf(2.5,3▸
     .2401238396
```

If you prefer to use the table method for parts c) and d), follow steps similar to those used in Example 1, part c), Method 1. What *z*-score values will you use?

The probability of getting exactly 3 black marbles is about 0.240.

d) If you want at least 3 black marbles, calculate $P(X \ge 2.5)$ by inputting **normalcdf(2.5,999999,2,1.058)**. The probability is about 0.318.

If you want more than 3 black marbles, calculate $P(X \ge 3.5)$ by inputting **normalcdf(3.5,999999,2,1.058)**. The probability is about 0.078.

```
normalcdf(2.5,9▸
     .3182529417
normalcdf(3.5,9▸
     .0781291021
```

If you want fewer than 3 black marbles, calculate $P(X < 2.5)$ by inputting **normalcdf(−999999,2.5,2,1.058)**. The probability is about 0.682.

If you want at most 3 black marbles, calculate $P(X < 3.5)$ by inputting **normalcdf(−999999,3.5,2,1.058)**. The probability is about 0.922.

```
normalcdf(-9999▸
     .6817470583
normalcdf(-9999▸
     .9218708979
```

e) There is no hypergeometric probability function on a graphing calculator. You can use the combinatoric method from section 4.5.

To locate the nPr and nCr functions on some graphing calculators, select the MATH PRB functions. Explore your own calculator and record how to access these functions.

$$\begin{aligned}P(X=3) &= \frac{{}_{24}C_3 \times {}_{36}C_2}{{}_{60}C_5}\\ &\approx 0.233\end{aligned}$$

```
24 nCr 3*36 nCr▸
     .233473807З
```

The probability of getting exactly 3 black marbles is about 0.23. The two answers are about the same.

Your Turn

Allison has a drawer full of unmatched socks. The drawer contains 30 blue socks, 30 green socks, and 30 yellow socks. She pulls seven socks from the drawer and records the number of blue socks in the sample.

a) Is it reasonable to approximate this distribution with a normal distribution? Give a reason for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) What is the probability that 3, 4, or 5 of the socks are blue?

Consolidate and Debrief

Key Concepts

- As the number of trials increases, a binomial distribution takes on the characteristics of a normal distribution.
- If the values of np and nq are both greater than 5, you can approximate the binomial distribution using a normal distribution.
- If the sample size is small compared to the population size, a hypergeometric distribution takes on the characteristics of a normal distribution.
- If the sample size is less than one-tenth of the population size, $n < \frac{1}{10}NP$, you can approximate the hypergeometric distribution using a normal distribution.
- You must use a continuity correction when approximating a discrete distribution with a normal distribution. For example, if you want the probability of rolling a 6 exactly 3 times, you must calculate $P(2.5 \leq X \leq 3.5)$. If you want the probability of rolling at least 3 sixes, you must calculate $P(X \geq 2.5)$. However, if you want the probability of rolling more than 3 sixes, you must calculate $P(X \geq 3.5)$.

Reflect

R1. Suggest another situation that is usually modelled with a binomial distribution but could reasonably be approximated with a normal distribution.

R2. Suggest another situation that is usually modelled with a hypergeometric distribution but could reasonably be approximated with a normal distribution.

R3. Suggest possible reasons why you might prefer to use a normal approximation rather than a binomial or hypergeometric distribution.

R4. Describe a situation where it is not possible to use a normal approximation for the binomial or hypergeometric distribution.

Practise

Choose the best answer for #1 and #2.

1. A card is randomly drawn from a deck of 52. Drawing a diamond is a success and anything else is a failure. The card is replaced and the deck is shuffled. The experiment is repeated 30 times. To model this experiment using a normal distribution, what mean and standard deviation should you use?

A 7.5, 2.372

B 7.5, 5.625

C 2.739, 2.372

D 2.739, 5.625

2. A bag of jellybeans contains 200 beans, of which 30 are red. Susan reaches into the bag and pulls out 15 beans at random. To model this experiment using a normal distribution, what standard deviation should you use?

A 1.913

B 1.778

C 1.383

D 1.333

3. Two dice are rolled. A double is considered a win, and anything else is a loss. What is the minimum number of rolls that should be made to model this situation using a normal distribution?

4. A barrel at the Pro Shop contains 30 white golf balls, 20 yellow golf balls, and 10 orange golf balls. A contest requires a contestant to blindly select several balls without replacement. The prize depends on the number of orange golf balls obtained. What is the maximum number of balls that could be selected to model the contest using a normal approximation?

5. Five cards are dealt from a deck of 52. The number of hearts is counted.

a) Is it reasonable to model this distribution with a normal distribution? Explain.

b) What mean should you use?

c) What standard deviation should you use?

Apply

6. A special HOV (high-occupancy vehicle) lane along a highway is reserved for cars carrying two or more people. Police records indicate that 8% of the cars in the HOV lane are occupied by fewer than two people. A random police check observed 100 cars.

a) Use the binomial distribution to determine the probability that exactly 10 of the cars contained one person.

b) Use the normal approximation to determine the probability that exactly 10 of the cars contained one person.

c) Compare the answers to parts a) and b).

7. **Thinking** A coin is tossed 12 times.

a) Use technology to help you create a probability table for the number of heads using the binomial distribution.

b) Construct a probability distribution.

c) What does the height of each bar on the graph represent?

d) Determine the total area under the probability distribution.

e) Is it reasonable to model this experiment using a normal distribution? Explain.

f) Construct the normal approximation.

g) Devise a method to measure the area under the normal distribution as accurately as possible.

h) How does the area under the normal distribution compare to the area under the probability histogram?

8. A card is drawn randomly from a deck and then replaced. The deck is shuffled. Ten trials are carried out.

a) Use the binomial distribution to determine the probability that there are exactly 5 diamonds in 10 trials.

b) Could you reasonably model this distribution using a normal approximation? Explain.

c) Determine the mean and standard deviation of the normal approximation.

d) Use the normal approximation to determine the probability of getting exactly 5 diamonds.

e) How does the answer to part d) compare with the answer to part a)?

✓ Achievement Check

9. Microwave ovens made in China are packaged into containers and shipped to Canada. About 1% are expected to be dented during transport. A sample of five ovens is removed from a container that holds 200. If one is found dented, the container is rejected.

a) Use the hypergeometric distribution to determine the probability that no ovens in the sample are dented.

b) Determine the mean and standard deviation of a normal approximation to this distribution.

c) Use the normal approximation to determine the probability that no ovens in the sample are dented. How does the answer compare with the answer from part a)?

10. An insurance company knows that 12% of the homeowners in a town of 900 are customers. The marketing department calls 50 homes at random.

a) What is the probability that 10 or more of these are already customers?

b) What method did you choose to solve this problem? Give reasons for your choice.

11. **Application** Honey jars from the farm where Doris works say they contain 500 g of honey. A technician measures a sample of 30 jars. The mean content is 502.83 g, with a standard deviation of 1.95 g. The technician can adjust the machine that fills the jars to change the mean. Assume that the standard deviation remains unchanged.

a) Determine the probability that a honey jar contains less than 500 g of honey.

b) Do you need to use a continuity correction factor? Explain.

c) The owners of the company would like to ensure that the probability that a jar contains less than 500 g is at most 0.005. What setting for the mean is required?

Extend

12. A multiple choice test consists of 50 questions with 5 possible answers each. Students need 60% or more to pass.

a) Andre has not studied for the test and guesses randomly at the answers. What is the probability that he will pass the test?

b) Maria has done some studying, and can narrow down the choices to 2 of the 5 provided for each question. What is the probability that she will pass the test?

13. **Thinking** A police survey shows that 5% of drivers passing an intersection are distracted. The police initiate a public education program to inform drivers of the danger. After a month, the police observe 120 cars at the same intersection and find 4 drivers are distracted.

a) If the education program had no effect, how many drivers out of 120 would you expect to be distracted?

b) What is the probability that 4 or fewer drivers would be found distracted out of 120 by pure chance?

c) Would you conclude that the program was effective? Explain your answer.

Chapter 7 Review

Learning Goals

Section	After this section, I can
7.1	• distinguish between discrete variables and continuous variables • work with sample values for situations that can take on continuous values • represent a probability distribution using a mathematical model • represent a sample of values of a continuous random variable using a frequency table, a frequency histogram, and a frequency polygon
7.2	• determine the theoretical probability for a continuous random variable over a range of values • determine the mean and standard deviation of a sample of values • calculate and explain the meaning of a *z*-score • solve real-world probability problems involving normal distributions
7.3	• recognize the general characteristics of a normal distribution • use technology to simulate a normal distribution in order to investigate its properties • determine probabilities for a normal distribution
7.4	• distinguish among the meanings of common confidence levels such as 90%, 95%, and 99% • determine the margin of error for a population mean estimated using a sample • determine the upper and lower limits of the confidence interval
7.5	• make connections between a normal distribution and a binomial distribution • make connections between a normal distribution and a hypergeometric distribution • recognize the role of the number of trials in these connections

7.1 Continuous Random Variables, pages 320–331

1. Advanced scuba divers sometimes breathe enriched air called nitrox. Nitrox contains 32% or 36% oxygen rather than the usual 21% found in ordinary air. Nitrox is toxic to humans if breathed for too long. Before using a nitrox-filled tank, the diver must verify and record the actual percent of oxygen in the tank. The table shows the oxygen content of a sample of 15 tanks.

Percent Oxygen				
32.1	32.2	31.8	32.2	32.1
32.0	31.8	32.1	32.0	31.9
32.0	31.9	31.9	32.2	31.8

a) Can you determine from the table whether the distribution is uniform? Explain your answer.

b) Devise a plan to determine whether the distribution is uniform. Carry out your plan, and draw a conclusion.

2. A sporting goods company produces custom wooden arrows. Any arrows with a length less than 69.2 cm or greater than 73.0 cm are rejected. The remaining arrows follow an approximately uniform distribution.

a) What is the height of the probability distribution?

b) What is the probability that an arrow has a length less than 71.1 cm?

c) What is the probability that an arrow has a length between 70.6 cm and 71.6 cm?

3. Ryan is raising fruit flies as a science project. The table shows the frequencies of the lifetimes of the flies.

Lifetime (days)	Frequency
5–7	1
7–9	3
9–11	13
11–13	24
13–15	27
15–17	20
17–19	9
19–21	3

a) Sketch a frequency histogram for these data.

b) Sketch a frequency polygon for these data.

c) Estimate the mean life of the fruit flies.

7.2 The Normal Distribution and *z*-Scores, pages 332–345

4. Refer to the fruit fly data in #3.

a) Add a column of relative frequencies to the table.

b) What is the probability that a given fruit fly will die before the end of the first week?

c) What is the probability that a fruit fly will live from 11 to 17 days?

5. Triple Q Farms grows soybeans. A farmer is testing a new strain of plant. After 3 months, 28 seeds produced plants with heights as shown.

Soybean Height (cm)			
25.1	48.8	41.0	47.4
39.7	40.2	41.1	41.6
44.2	49.5	36.9	30.0
32.0	36.8	36.1	37.1
37.6	34.4	44.6	45.8
38.7	38.8	32.0	44.0
42.9	34.9	43.3	33.8

a) Determine the mean height.

b) Determine the standard deviation of the heights.

c) Sketch a frequency histogram for these data.

d) Do the heights appear to be normally distributed? Explain.

6. Current engineering graduates earn a mean starting salary of \$62 000 in Canada, with a standard deviation of \$2500. Assuming that the salaries are normally distributed, what is the probability that a graduate will find a job with a starting salary of more than \$65 000?

7.3 Applications of the Normal Distribution, pages 346–351

7. A police radar unit is set up to monitor vehicles on a stretch of highway with a speed limit of 100 km/h. Long-term records for this location show that speeds vary normally with a mean of 105 km/h and a standard deviation of 7 km/h. Drivers who exceed the speed limit by 20 km/h accumulate demerit points as well as pay a fine.

a) What percent of the drivers will accumulate demerit points?

b) What is the probability that a given vehicle has a speed between 99 km/h and 101 km/h?

8. An electric car requires a mean recharge time of 4.5 h if the batteries are fully discharged. The manufacturer guarantees that the car will fully recharge in no more than 5 h. If the probability that a recharge takes more than 5 h is to be kept to 0.0001%, what is the standard deviation of the battery recharge time?

7.4 Confidence Intervals, pages 352–361

9. Marketing analysts for a soft drink manufacturer conducted a survey in a large town. They determined that 42% of the 150 people surveyed regularly bought the soft drink.
   a) Determine the margin of error at a 95% confidence limit.
   b) Determine the confidence interval for the market share for the soft drink.

10. The mean lifetime expected from a model of automobile follows a normal distribution with a standard deviation of 9500 km. A sample of 100 cars showed a mean lifetime of 190 000 km.
   a) Determine the margin of error at a 90% confidence limit.
   b) Determine the confidence interval for the mean.

7.5 Connections to Discrete Random Variables, pages 362–371

11. A bus company has records showing that its buses arrive on time 95% of the time. Suppose the company operates 65 bus trips each day. The CEO has asked for the probability that 60 or more of these buses arrive on time.

   a) Use the binomial distribution to determine the probability that 60 or more buses arrive on time.
   b) Could this distribution be reasonably modelled using a normal approximation? Give reasons for your answer.
   c) Determine the mean and standard deviation of the normal approximation.
   d) Use the normal approximation to determine the probability that 60 or more buses arrive on time.
   e) How does the answer to part d) compare with the answer to part a)?

12. A newspaper reports that 350 of the 3500 people living in a small town have the flu. The data management class at the local high school has 25 students.
   a) Do you expect this to be a representative sample? Explain.
   b) Could you reasonably model this distribution using a normal approximation? Give reasons for your answer.
   c) Determine the mean and standard deviation of the normal approximation.
   d) Use the normal approximation to determine the probability that at least 5 of the students have the flu.
   e) Use technology to compare the answer in part c) to that calculated from the hypergeometric distribution.

Chapter 7 Test Yourself

✓ Achievement Chart

Category	Knowledge/ Understanding	Thinking	Communication	Application
Questions	1, 2, 5	6, 10, 13	10, 11, 15	3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15

Multiple Choice

Choose the best answer for #1 to #6.

1. Which of these variables would be expected to produce a discrete distribution?

A the distance of a long jump during a track meet

B the mass of a bolt produced by a factory

C the number of students with the flu in a given class at your school

D the total monthly rainfall at a weather station

2. Which of these variables would be expected to produce a continuous distribution?

A the number of customers at a restaurant at a given time

B the mass of a hawk recorded during a migration

C the number of hamburgers sold each day in the school cafeteria

D the number of defective watches in a shipment to a department store

3. The average speeds of five contestants in a bicycle race were 24.2 km/h, 28.1 km/h, 21.6 km/h, 22.0 km/h, and 31.2 km/h. What is the mean of these speeds?

A 22.0 km/h **B** 24.2 km/h

C 25.4 km/h **D** 31.2 km/h

4. What is the standard deviation of the speeds in #3?

A 3.70 km/h **B** 4.13 km/h

C 4.4 km/h **D** 5.04 km/h

5. One hundred twenty students qualified for the high jump event at a track meet. The table shows the probability distribution for first jumps.

Height of Jump (cm)	Probability
120–130	0.025
130–140	0.040
140–150	0.081
150–160	0.150
160–170	0.175
170–180	0.218
180–190	0.117
190–200	0.092
200–210	0.073
210–220	0.017
220–230	0.012

What is the frequency associated with a jump between 180 cm and 190 cm?

A 0

B 0.117

C 14

D 120

6. Which statement is true concerning a normal distribution?

A The curve may be skewed to the left or right.

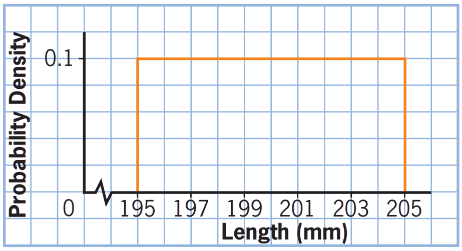
B The median is equal to the mean.

C 95% of the data values will occur within one standard deviation of the mean.

D The mean of a sample is always less than the mean of the underlying normal distribution.

Short Answer

7. An assembly line produces flip-flops of stated length 200 mm. The graph shows the probability distribution of the actual lengths. What is the probability that a given flip-flop will have a length greater than 203 mm?



8. Karen's Zumba class has 50 students. She conducts a Zumba endurance contest. The table shows the probability distribution of the participants lasting increasing lengths of time.

Endurance (min)	Probability
30–35	0.04
35–40	0.04
40–45	0.10
45–50	0.20
50–55	0.24
55–60	0.16
60–65	0.10
65–70	0.08
70–75	0.04

   a) How many students lasted between 45 min and 50 min?

   b) What is the probability that a student lasted less than an hour?

9. When the ketchup dispenser at a fast-food restaurant is completely depressed, it dispenses a mean of 15 mL of ketchup with a standard deviation of 0.75 mL. Assuming that the distribution is normal, what is the probability that a hamburger will receive less than 14 mL of ketchup in one complete press?

10. Legs for a wooden dining room table are produced by a computer numerically controlled (CNC) lathe. The cutting blade lasts a mean time of 550 h with a standard deviation of 20 h. To avoid cutting errors, the shop manager would like to keep the probability of a failure to less than 0.002. How many hours should the blade be used before replacement? Explain.

11. A pharmaceutical company has determined the probability that a new antacid will relieve stomach distress is 75%. The antacid is given to 1000 patients. The number of patients who reported relief is recorded.

   a) Could you reasonably model this distribution using a normal approximation? Explain.

   b) Determine the mean and standard deviation of the normal approximation.

Extended Response

12. Students arrive at school at various times before the bell rings for the first class. The table shows data for 200 students.

Time (min)	Frequency
0–3	2
3–6	6
6–9	18
9–12	32
12–15	49
15–18	42
18–21	27
21–24	19
24–27	5
27–30	0

   a) Sketch a frequency distribution histogram.

   b) Add a frequency polygon to the histogram in part a).

   c) Do the data appear to follow a normal distribution? Explain your reasoning.

13. Airliners taking off from City Central Airport produce a mean noise level of 108 dB (decibels), with a standard deviation of 6.7 dB. To encourage airlines to refit their aircraft with quieter engines, any airliners with a noise level above 120 dB must pay a "nuisance fee."

   a) What percent of the airliners will be billed a nuisance fee?

   b) After two years, a sample of 1000 airliners showed that 4 were billed the nuisance fee. Assuming that the program has been effective, and that the standard deviation has not changed, what is the new mean noise level?

14. A mail order company is planning to deliver small parcels using remote-controlled drones direct to households within 10 km of its warehouses, each located in a large city. As a test, drones delivered 500 parcels. A total of 420 parcels were delivered within the advertised time limit of 30 min. Determine a 99% confidence interval for the proportion of parcels delivered within 30 min.

15. A high school operates 450 computers. On a given day, an average of 15 of these computers are out of service. A computer lab contains 30 computers. A teacher would like to use the lab with her class of 27 students.

a) Is this a binomial or a hypergeometric distribution? Explain.

b) Could you reasonably model this distribution using a normal approximation? Explain.

c) Determine the mean and standard deviation of the normal approximation.

d) Use the normal approximation to determine the probability that there will be enough working computers.

Chapter Problem

Food Service Industry

The honey farm has installed a filling machine for honey jars that hold 500 g of honey. Doris is calibrating the new machine. She sets the machine to a mean of 500 g, and performs a test run of 48 jars, or two cases. The table shows the results.

Honey (g)					
499	501	498	497	499	500
501	499	501	501	498	502
498	498	502	500	499	499
496	503	500	501	499	498
498	501	499	502	500	500
500	502	501	500	500	500
500	501	500	500	500	503
499	499	497	502	499	502

a) Determine the mean and standard deviation of these data.

b) Assume that the distribution is normal. What is the probability that a jar contains less than 500 g of honey?

c) The probability that a jar contains less than 500 g is allowed to be at most 0.005. What setting for the mean should Doris try? What assumptions did you make?

d) How can Doris check that she has set the machine correctly?

e) Using the setting in part c), what is the probability that a jar might contain more than 510 g of honey?

f) As a manager of the company, what policy would you put in place to reassure any customers who receive less honey than expected?