Continuous Random Variables

Lesson objectives

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

1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 327 #s 1 - 5 & 7

Definitions

Attribute

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

Frequency Histogram

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

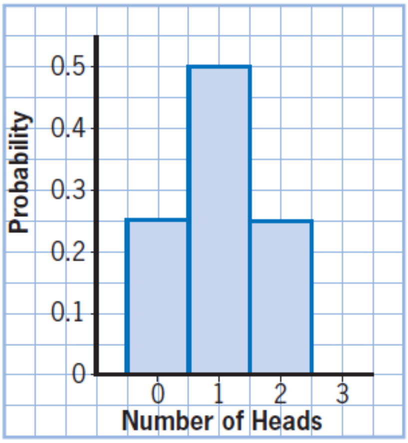
Frequency Polygon

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

Investigate on Page 320

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?

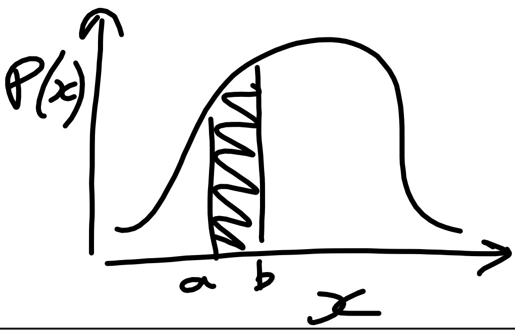
For continuous data, you can group the outcomes into intervals. The variable can take on any value in the interval between two numbers, including decimal and fractional values. For example, suppose you measure the distances of various horseshoe throws, and group the distances in intervals.

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.

4m ?



$2 < x \leq 4$ ← ✓ 4m

$4 < x \leq 6$ ← ✗

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.

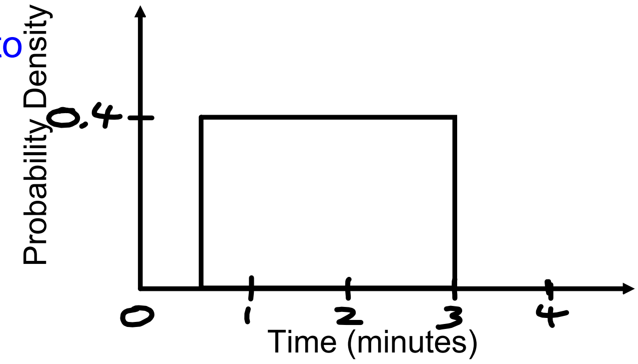
c) What is the probability that a customer will eat a doughnut in 1 to 3 min?

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

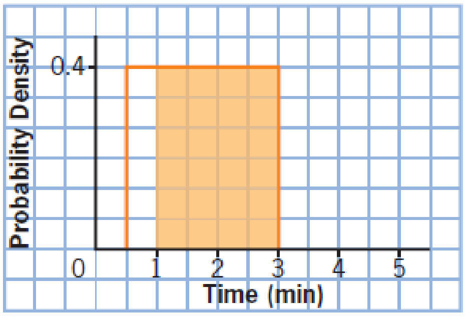
e) Is it possible to determine the probability that a customer will eat a doughnut in exactly 1 min using the area under the graph? Explain your answer.

a) This will be a uniform distribution since all of the times (outcomes) are equally likely.

b) Horizontal axis must include 0.5 to 3 minutes. Because the area of the graph must equal 1 (represents all probabilities), the y-axis (probability density) will go up to 1 ÷ 2.5 = 0.4



c) The probability that a customer will eat a donut in 1 to 3 minutes is equal to area under the graph for the region.



$P(1 \leq x \leq 3)$ = area under graph

= 0.4(3.0 - 1.0)

= 0.8

The probability of a donut being eaten in 1 to 3 minutes is 0.8

d) Since time is a continuous variable, there is an infinite number of values that could exist from 0.5 to 3 minutes.

e) If choosing a single value, such as 1 minute, the area would equal zero since the width of the rectangle is zero. The area method cannot be used to find the probability of an individual value for a continuous variable, only for a range.

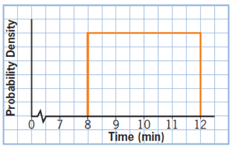
Your Turn

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?



First find the height of the rectangle (value of the probability density)

= 1 ÷ (12.0 - 8.0)

= 0.25

a) $P(8 \leq x < 9)$ = area under graph

= 0.25(9.0 - 8.0)

= 0.25

The probability of a tire being changed in under 9 minutes is 0.25

b) $P(9 < x < 11.5)$ = area under graph

= 0.25(11.5 - 9.0)

= 0.625

The probability of a tire being changed between 9 and 11.5 minutes 0.625

c) The area method cannot be used to find the probability of an individual value for a continuous variable, only for a range. Therefore the probability of Chris changing a tire in exactly 10 minutes is zero.

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.

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.

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

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.

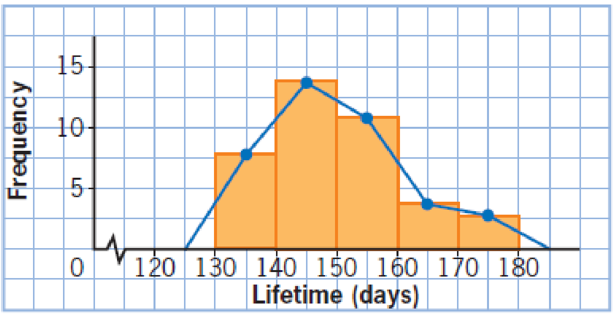
f) How is the shape of the frequency polygon related to the shape of the probability density distribution for the variable? Can you use the area under the frequency polygon to calculate probabilities for any range of values?

a) Analysing microdata is difficult for continuous data. This is a reason why it is often grouped and then analysed. In its current form it is difficult to see if the distribution is uniform or not. Likewise, it is difficult to make an estimate for the mean, aside from totaling all the values.

c) As the frequencies are different we can say that the distribution is not uniform. An estimate for the mean value would be about 150. You would find the answer to $\Sigma m_i x_i \div n$

d) The frequency table groups this raw data into class intervals. The frequency of each interval makes the distribution more obvious and gives an indication of the location of the central measures of tendency.

e)

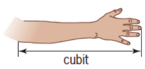


A frequency polygon is created by joining the midpoints of each bar together with a straight line.

f) The frequency polygon with give an indication of the shape of the probability distribution. The area under it will not equal one though. You cannot calculate probabilities from a frequency polygon, you need to use a probability distribution.

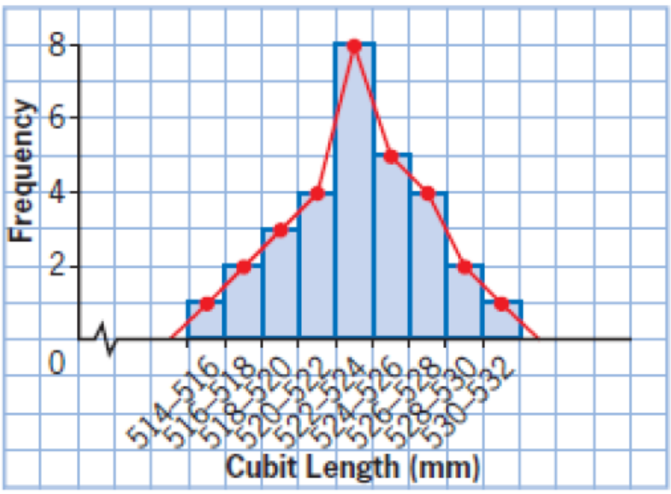
Your Turn

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?



c) Mean length $= \dfrac{\sum m_i f_i}{\sum f_i}$

$= [515(1) + 517(2) + 519(3) + 521(4) + 523(8) + 525(5) + 527(4) + 529(3) + 531(1)] \div 30$

$= 15696 \div 30$

$= 523.2$ mm

The estimated mean cubit length is 523.2 mm

d) Distribution is NOT uniform because the frequencies are different

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.

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.

The number of guests that occupy the hotel each day is discrete data, while the time a guest waits for an elevator is continuous data. It is possible to list all values of the discrete distribution, since these values from 0 up to the maximum capacity of the hotel. It is not possible to list all values of the continuous distribution, since time can be recorded in fractions of a second.

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.

The various height measurements could be caused by several things. For example, Maya standing differently each time, bad measuring technique, and misreading the measuring device. This is most likely a measurement error, not bias.

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?

The probability that a variable falls within a range of values is equal to the area under the probability density graph for all that range of values. The area method cannot be used for single values of a continuous variable, only for a range of values. A continuous random variable can take on an infinite number of values. The probability that it will equal a specific value is always zero.