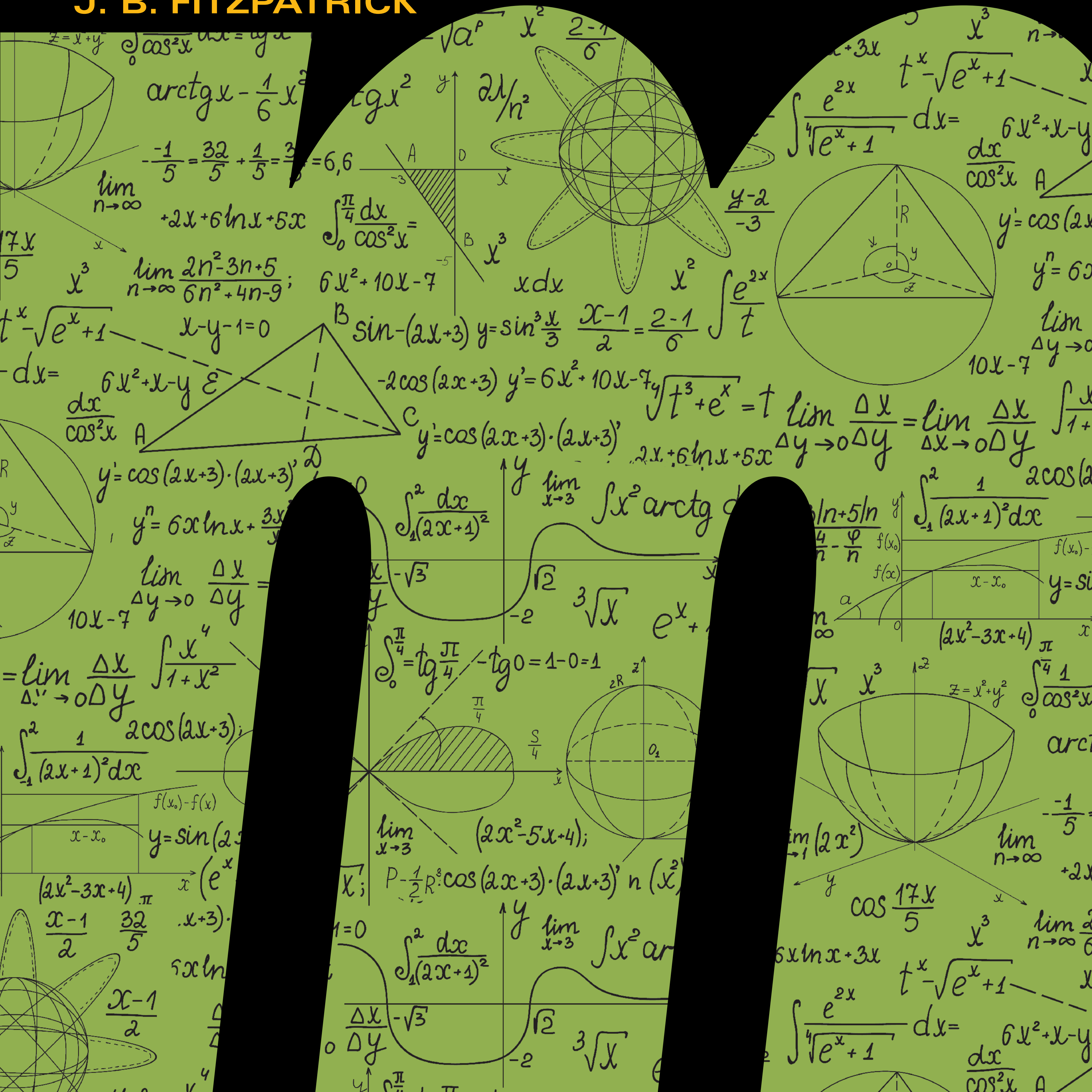
NEW SENIOR

Mathematics

ADVANCED YEAR 12 | 4TH EDITION

BOB AUS
J. B. FITZPATRICK

NSW STAGE 6



CHAPTER 7
Random variables

7.1 DISCRETE RANDOM VARIABLES

When performing a sampling procedure, a number of different outcomes are expected. For example, when rolling a normal die a large number of times, you can expect to observe some of each of the values from {1, 2, 3, 4, 5, 6}. The outcome can vary between rolls. X, the observed outcome, is a **random variable**. In particular, X is a **discrete random variable** because the list of possible outcomes is countable. Discrete random variables are often associated with number or size.

On the other hand, if the list of possible outcomes is not countable, the variable is a **continuous random variable**. For example, when measuring the heights of a sample of people, although you might expect the measurements to fall within a range (say, 140 cm to 190 cm), each individual value is dependent only on the degree of accuracy of the measuring instrument.

Continuous random variables are often associated with height, mass and time. Further work will be done with continuous random variables later in the chapter.

Discrete variables and whole numbers

A discrete random variable is not restricted to taking on whole number values; the important criterion is that the number of outcomes must be countable. For example, shoe sizes using the British measuring system increase in half sizes such as 7, $7\frac{1}{2}$, 8, . . ., but since the total number of different sizes can be counted, the variable is discrete.

Statistical models

Statistical modelling is a process used to predict real-world events. A **statistical model** consists of equations that are based on assumptions and simplifications that produce a mathematical result. The predictions of the model can then be tested against some real-world data. Since it is unlikely that a model will be completely accurate, it is often modified to make it better fit the data. As an example, one might suggest that adults have the same arm span as their height. However, after collecting a large quantity of data, arm span might be found to be closer to 97% of a person's height.

Notation

Capital letters (e.g. X and Y) are used for random variables, and their corresponding lower-case letters (e.g. x and y) are used for the values that the random variable takes. Subscripts distinguish the various possible values of X. For example, the probability that the variable X takes the value x_2 is denoted as $P(X = x_2)$.

By varying the number of times a coin is tossed in the activity below, you will see how the probability of obtaining a fixed number of heads changes.

MAKING CONNECTIONS

Tossing three coins

Use a spreadsheet to simulate the results of tossing three coins n times.

In some cases, the variable is not the actual outcome but rather a value assigned to an outcome. Consider an experiment where three coins are tossed. Let X stand for the number of heads obtained. There are 8 possible outcomes.

Outcome observed	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
Value of X	3	2	2	2	1	1	1	0

So, X can be any of the values from $\{0, 1, 2, 3\}$. Assuming that each of the 8 observed outcomes are equally likely, which is a reasonable assumption using fair coins, a probability table can be created for the variable X.

x	0	1	2	3
$P\ (X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Example 1

Four coins are tossed. What is the probability of obtaining exactly two heads?

Solution

Write out the sample space for the experiment.

(There are two options for each coin, so the sample space will consist of $2 \times 2 \times 2 \times 2 = 16$ options.)

$$\left\{\begin{matrix} \text{HHHH,} & \text{HHHT,} & \text{HHTH,} & \text{HHTT} \\ \text{HTHH,} & \text{HTHT,} & \text{HTTH,} & \text{HTTT} \\ \text{THHH,} & \text{THHT,} & \text{THTH,} & \text{THTT} \\ \text{TTHH,} & \text{TTHT,} & \text{TTTH,} & \text{TTTT} \end{matrix}\right\}$$

Define the variable to be used: Let $X =$ the number of heads obtained, so $X = 2$.

From the sample space write out the events which correspond to the situation.

(Here there are six events.)

$$\left\{\begin{matrix} \text{HHTT,} & \text{HTHT,} & \text{HTTH} \\ \text{THHT,} & \text{THTH,} & \text{TTHH} \end{matrix}\right\}$$

State the probability for each successful event: $P(\text{HHTT}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

and

$P(\text{HHTT}) = P(\text{HTHT}) = P(\text{HTTH}) = P(\text{THHT}) = P(\text{THTH}) = P(\text{TTHH}) = \frac{1}{16}$

Calculate the required probability by adding together the individual probabilities for each outcome.

$$P(X = 2) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{6}{16} = \frac{3}{8}$$

When the outcomes are equally likely you can use $P(\text{Event})$:

$$P(\text{Event}) = \frac{\text{number of successful outcomes}}{\text{total number of outcomes}}$$

Finding sample size

A more interesting question would be to find the sample size necessary to meet a particular condition. In many of these questions it is useful to use the complementary event. Recall that $P(\overline{A}) = 1 - P(A)$.

Example 2

A netball player scores a goal from 75% of her shots. How many shots at goal would she need to have at least a 95% chance of scoring at least one goal?

Solution

Use complementary events: P(scoring at least one goal) = 1 − P(scoring no goals)

You require: P(scoring no goals) < 0.05

Systematically vary the number of misses to find the required value: let M be the event that she misses the goal.

$$\begin{aligned} P(M) &= 0.25 \\ P(MM) &= 0.25 \times 0.25 \\ &= 0.0625 \\ P(MMM) &= 0.25 \times 0.25 \times 0.25 \\ &= 0.015\,625 \\ &< 0.05 \end{aligned}$$

The netball player would need to take at least three shots at goal to be at least 95% certain of scoring at least one goal.

Discrete probability distributions

Consider an experiment where three coins are tossed. The table of probabilities for each outcome forms what is known as a discrete probability distribution.

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

It lists each of the x values possible and the associated probabilities. Strictly speaking you should indicate that the probability for any other value of x is zero, but in cases such as this it is obvious so you do not need to state it explicitly.

Some conditions need to be met before you can say you are dealing with a discrete probability distribution.

For a discrete probability distribution: $0 \le P(X) \le 1$ for all values of x,
and $\sum P(X = x) = 1$ (where $\sum$ stands for 'the sum of').

If either of these conditions is not met, you do not have a discrete probability distribution.

Example 3

Does the following table represent a discrete probability distribution?

x	1	3	5	7
$P(X = x)$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{5}$

Solution

Are all the probabilities between 0 and 1, inclusive; i.e. is $0 \le P(X) \le 1$? Yes.

Do the probabilities add up to 1?

$$\begin{aligned} \sum P(X = x) &= \frac{1}{5} + \frac{1}{10} + \frac{3}{10} + \frac{2}{5} = \frac{2}{10} + \frac{1}{10} + \frac{3}{10} + \frac{4}{10} \\ &= \frac{10}{10} \\ &= 1 \end{aligned}$$

Both conditions have been met, so the table of data represents a discrete probability distribution.

The probabilities can be represented as fractions, decimals or percentages.

Example 4

The table of data below represents a discrete probability distribution.

x	3	4	5	6	7
$P(X = x)$	0.14	k	0.36	0.21	0.13

Find the value of k.

Solution

Add up the given probabilities: $0.14 + k + 0.36 + 0.21 + 0.13 = 0.84 + k$

As the sum of probabilities = 1, then: $0.84 + k = 1$

Solve for k: $k = 1 - 0.84$

$k = 0.16$

Example 5

The following discrete probability distribution represents a five-sector spinner where the areas of the sectors are not equal.

x	1	2	3	4	5
$P(X = x)$	k	$\frac{k}{2}$	$\frac{k}{3}$	$\frac{k}{4}$	0.25

Which of the values for k makes this a discrete probability distribution?

A 0.48 **B** 0.1875 **C** 0.25 **D** 0.36

Solution

Add up the terms involving the unknown: $k + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} = \frac{25k}{12}$

As the sum of probabilities = 1, then: $\frac{25k}{12} + 0.25 = 1$

$$\frac{25k}{12} = \frac{3}{4}$$

$$k = \frac{9}{25} = 0.36$$

The answer is D.

Note that the other options came from common mistakes that students make.

A ignores the probability of 0.25 and just solves $\frac{25k}{12} = 1$.

B assumes that there were no fractions and the equation to be solved is $4k = 0.75$.

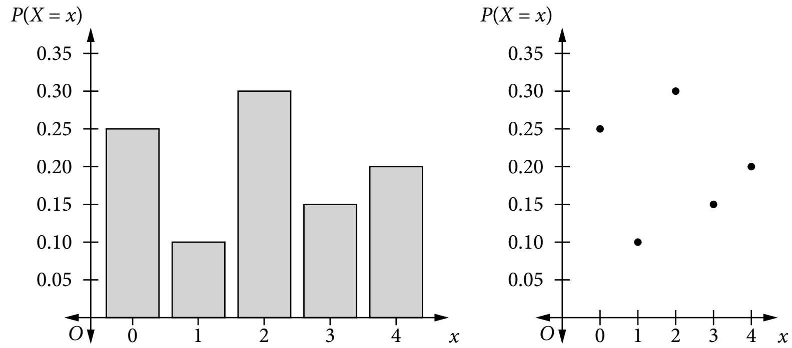
C assumes that there were no fractions and the equation to be solved is $4k = 1$.

As with many areas of mathematics, you may find it useful to draw a graph as a pictorial representation of the distribution.

Consider the following table of values:

x	0	1	2	3	4
$P(X = x)$	0.25	0.1	0.3	0.15	0.2

The graph of the distribution is shown in two different ways: a bar graph on the left and a dot graph on the right.



There is another notation that can be used for describing a discrete probability distribution that is particularly useful if the possible values all have the same probability. To describe the distribution for rolling a normal six-sided die, you can write: $P\,(X = x) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{for all other values of } x \end{cases}$

Here the probability attached to every value is explicitly stated, not just values with non-zero probability.

EXERCISE 7.1 DISCRETE RANDOM VARIABLES

1 Find the probability of each of the following events. Express your answers as fractions in simplest form.

(a) The probability of obtaining exactly one head if three coins are tossed.
(b) The probability of obtaining at least one head if three coins are tossed.
(c) The probability of obtaining exactly two even numbers if two dice are rolled.
(d) The probability of obtaining exactly one odd number if two dice are rolled.
(e) The probability of obtaining a pair of numbers that are the same if two dice are rolled.

2 Wayne is a soccer player who takes the penalty kicks awarded to his team. He has a 60% chance of scoring from the penalty spot. For Wayne to be 95% certain of scoring at least one goal, how many penalty kicks would he need to take?

3 Do the following tables represent discrete probability distributions? Give a reason for your answer.

(a)

x	0	1	2	3	4
$P(X=x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

(b)

x	2	3	4	5	6
$P(X=x)$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$

(c)

x	−1	0	1
$P(X=x)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

(d)

x	1	2	3	4	5
$P(X=x)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$-\frac{1}{5}$

(e)

x	1	1.5	2	2.5	3
$P(X=x)$	20%	15%	30%	18%	17%

(f)

x	$1\frac{1}{2}$	$1\frac{3}{4}$	2	$2\frac{1}{4}$	$2\frac{3}{4}$
$P(X=x)$	0.24	0.16	0	0.38	0.22

4 For each of the following examples find the value of k that makes each table represent a discrete probability distribution.

(a)

x	0	1	2	3	4	5
$P(X=x)$	$\frac{1}{18}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{5}{18}$	k	$\frac{3}{18}$

(b)

x	0	1	2	3	4	5
$P(X=x)$	$\frac{1}{8}$	$\frac{5}{24}$	k	$\frac{7}{24}$	$\frac{1}{24}$	$\frac{1}{6}$

(c)

x	5	6	7	8	9
$P(X=x)$	$\frac{1}{9}$	k	$\frac{k}{2}$	$\frac{4}{9}$	$\frac{1}{9}$

(d)

x	8	9	10	11	12
$P(X=x)$	$\frac{1}{6}$	$3k$	$\frac{1}{3}$	k	$\frac{1}{6}$

5 Are the following random variables discrete or continuous data?

(a) The number of students in each of your classes at school.
(b) The height of teachers at your school.
(c) The sizes of the shirts worn by each of the students in your mathematics class.
(d) The neck circumference of each of the members of the Australian netball team.
(e) The number of tosses of a coin before a head is observed.
(f) The time it takes for a process worker to complete 50 items.
(g) The distance jumped in the long jump by the competitors at the sports carnival.
(h) The number of red lights you stop at, per day, on the way to school over the course of a month.
(i) The number of whole lessons missed, per student, this year by members of your English class.
(j) The retail price in cents per litre charged for petrol over the course of a year.
(k) The number of people in the queue when you enter the bank each Friday for a year.
(l) The number of passionfruit collected from each vine in your orchard this season.

6 You have 10 cards. Five of the cards are hearts, three are diamonds and two are spades. You draw two cards, with replacement, from the pack. Let X be the number of diamonds drawn.

(a) What are the only values that X can take?
A 0 only **B** 0 and 1 only **C** 0, 1 and 2 only **D** 1 and 2 only

(b) Complete the table to show the probability distribution of X. Express the probabilities in decimal form.

x	0	1	2
$P(X=x)$			

(c) What is the probability of drawing exactly one red card from the pack?
(d) What is the probability of drawing at least one red card from the pack?

7 The following table represents a discrete probability distribution.

x	3	5	7	9
$P(X=x)$	$\frac{1}{5}$	$\frac{2}{15}$	$\frac{1}{3}$	k

The value of k is:

A $\frac{2}{3}$ **B** 0 **C** 1 **D** $\frac{1}{3}$

8 Which value of k makes the following table a discrete probability distribution?

x	1	2	3	4
$P(X=x)$	$k+0.1$	$k-0.1$	$k+0.55$	$k-0.15$

A 0.6 **B** 0.15 **C** 0.9 **D** 0.1

9 A board game uses a spinner divided equally into eight sections, each of which has a different number from 1 to 8 written on it. Before a player can put a piece on the board they must spin a 6, 7 or 8.

(a) What is the probability of spinning a 6, 7 or 8?
(b) What is the probability of not rolling a 6, 7 or 8 on the first 3 attempts? Give your answer in decimal form, correct to 2 decimal places.
(c) What is the probability that a player has not been able to start after the sixth attempt? Give your answer in decimal form, correct to 2 decimal places.
(d) Part of the game design process is to ensure it is not too difficult for a player to start. How many attempts does a player need to be at least 97% certain that they can start to play the game?

10 An apartment complex is built with the units having two, three or four bedrooms in the ratio 3 : 6 : 1. One of the units is chosen at random.

(a) Complete the table to show the probability distribution of the number of bedrooms the unit contains. Write the probabilities in decimal form.

x	2	3	4
$P(X = x)$			

(b) What type of graph would best represent the data?

11 Two six-sided dice are rolled. Let X be the total of the two dice.

(a) What are the lowest and highest totals possible?

(b) Complete the table to show the probability distribution of X. Express the probabilities as fractions in simplest form.

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$											

(c) Find $P(X \geq 6)$. **(d)** Find $P(X < 10)$. **(e)** Find $P(4 \leq X \leq 10)$.

12 Two coins are tossed. T stands for the number of tails obtained.

(a) Complete the table to show the probability distribution of T. Express the probabilities in decimal form.

t	0	1	2
$P(T = t)$			

(b) What type of graph would best represent this distribution? Draw the graph.

13 A six-sided die is rolled, and X is the cube of the number showing.

(a) Complete the table to show the probability distribution of X. Express the probabilities as fractions in simplest form.

x	1	8	27	64	125	216
$P(X = x)$						

(b) Find $P(X < 100)$.

14 In a game a coin is tossed and a six-sided die is rolled. If the coin shows tails then X is the score on the die. If the coin shows heads then X is the square of the score on the die.

(a) Draw a table to show the probability distribution of X. Express the probabilities as fractions in simplest form.

(b) Find $P(X \leq 12)$. **(c)** Find $P(X \geq 16)$. **(d)** Find $P(X \leq 9 \mid \text{coin showed heads})$.

15 On the way to work Enzo must pass through three sets of traffic lights. The probability that he will stop at any particular set of lights is $\frac{3}{5}$.

(a) Assuming the traffic lights are independent of each other, construct a table to show the probability distribution of X, the number of traffic lights at which Enzo stops.

(b) What is the probability that, on a particular day, Enzo stops at:

(i) exactly two sets of traffic lights

(ii) no more than two sets of traffic lights

(iii) at least two sets of traffic lights

(iv) fewer than two sets of traffic lights?

(c) What do you notice about the answers to **(b)(iii)** and **(iv)**? Why does this occur?

16 A die is loaded so that the probability of obtaining each number is as shown in the table.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$

The die is rolled twice and Y is the sum of the two results.

(a) Draw a table to show the probability distribution of Y. Express the probabilities as fractions in simplest form.

(b) Find $P(Y \geq 4)$. **(c)** Find $P(Y < 7)$. **(d)** Find $P(3 \leq Y \leq 9)$. **(e)** Find $P(Y \geq 6 \mid Y \leq 10)$.

17 Sharmela is an interior designer who uses up to six colours in her designs. She sometimes chooses colours randomly for a unique effect. The number of colours Sharmela uses in her designs is a discrete random variable X with a probability distribution formula $P(X = x) = \frac{k}{x}$.

(a) Calculate the value of k.

(b) Draw a table to show the probability distribution of X. Express the probabilities as fractions in simplest form.

(c) Add a third column to the table to display the probabilities in decimal form, correct to 2 decimal places, and then draw a dot plot of the distribution.

18 Analyse each of the following goal-scoring situations.

(a) Lou has a 71% chance of scoring a goal when she kicks a penalty kick for her football team. How many penalty kicks would she need to take to have a 95% chance of scoring at least one goal?

(b) Buddy plays full-forward in an AFL team. He has a 68% chance of scoring a goal when taking a kick from 45 m out from goal. How many shots at goal would he need to have so that he has a 95% chance of scoring at least one goal?

(c) What is the largest whole number percentage that your chance of scoring needs to be so that you would need more than three shots to have a 95% chance of scoring at least once?

19 Alexander is the manager of a shoe shop. The shop sells shoes only in the following sizes: 7, 8, 9, 10, 11. To stock the shop according to market demand, Alexander records the sizes of shoes he sells in a month. The table shows his results.

Shoe size	7	8	9	10	11
Number of pairs	24	a	72	a	8

(a) If Alexander sold a total of 200 pairs of shoes and the random variable X represents the shoe size sold, create a discrete probability distribution table for the sale of shoes given their size, by first calculating the value of a.

(b) What is the probability that the next customer who buys a pair of shoes will buy a pair of shoes of size less than 10 but greater than 7?

(c) Calculate the probability that the next two pairs of shoes that Alexander sells are a size 10 followed by a size 11. Assume the events are independent.

(d) Alexander has noticed that he made a mistake when he recorded the sales over the given month. In one sale he sold two pairs of shoes whose sizes added up to 18. Mistakenly he recorded two pairs of shoes of size 9, but when he checked the receipt again he realised that the two pairs were of sizes 7 and 11.

(i) Explain how this mistake changes the probabilities in the probability distribution table. Write down the revised values, correct to 3 decimal places.

(ii) Calculate the probability that the next two pairs of shoes that Alexander sells are a size 10 followed by a size 11 using the revised values. Assume the events are independent.

7.2 EXPECTED VALUE, VARIANCE AND STANDARD DEVIATION OF DISCRETE PROBABILITY DISTRIBUTIONS

In the past, you have calculated the mean of a set of values as a measure of central tendency for that set. Another way of describing the mean is to call it the **expected value**. The study of expected value was first related to gambling, where being able to decide if a game is fair is certainly an advantage. The following example shows how this can be done.

Consider a game in which you roll a die and receive as your prize the number of dollars showing on the die.
The following table shows the possible outcomes.

Number showing	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
Gain ($)	1	2	3	4	5	6

If you were running the game you would need to decide how much to charge players for a turn. So, you need to know the expected gain for a player. Since a player can expect to win $1 on $\frac{1}{6}$ of the rolls, $2 on $\frac{1}{6}$ of the rolls, etc., for any particular roll of the die the expected gain would be:

$$\$1\times\frac{1}{6}+\$2\times\frac{1}{6}+\$3\times\frac{1}{6}+\$4\times\frac{1}{6}+\$5\times\frac{1}{6}+\$6\times\frac{1}{6}$$
$$=\$\left(\frac{1}{6}+\frac{2}{6}+\frac{3}{6}+\frac{4}{6}+\frac{5}{6}+\frac{6}{6}\right)$$
$$=\$\frac{21}{6}$$
$$=\$3.50$$

This means you would expect the player to gain, on average, \$3.50 on each roll of the die. You should note that this average outcome could not happen itself, since there is no result which will return \$3.50 to the player. As the game operator, you could charge \$3.50 to play, in which case the game would be considered fair, or you could charge more than \$3.50 if you wanted to make a profit. It is important to note that gambling games in businesses like casinos are designed so that the casino always has a slight advantage to ensure they make a profit.

This example leads to the definition for the expected value of a discrete random variable.

Expected value

The expected value or mean, $E(X)$ or μ (mu), of a discrete random variable, X, is the sum of each possible value multiplied by its probability.

In symbolic form:

Let X take the values $x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ with associated probabilities $p_1, p_2, p_3, \ldots, p_i, \ldots, p_n$.

Then:

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \ldots + x_ip_i + \ldots + x_np_n = \sum_{i=1}^{n} x_i p_i$$

This sum can also be expressed as $\sum_{x=1}^{n} x\,P(X = x)$.

Remember that Σ is the Greek letter sigma and stands for 'the sum of'.

This is similar to the formula used to find the mean of a data set that has been expressed in a frequency table.

Therefore: mean $= \dfrac{\Sigma xf}{\Sigma f}$

$= \dfrac{105}{30}$

$= 3.5$

x	f	xf
1	4	4
2	7	14
3	3	9
4	6	24
5	6	30
6	4	24
	$\Sigma f = 30$	$\Sigma xf = 105$

Summation notation

There are several different notations used for summations of this type.

You can leave out the index from the summation sign when it is a sum over all possible values.

Another notation is $\sum_x$ which means to use all possible values of x. This form is especially useful when the values of x are not consecutive.

Mode

In statistics, the mode is the most common score.

For a discrete probability distribution, the mode will be the outcome that has the largest chance of occurring, that is, the outcome with the largest probability.

Obviously, in a uniform distribution, every outcome is the mode, or you can say that it is multimodal.

In Example **6** below, $P(X = 4) = \dfrac{3}{8}$ is the largest probability so the mode is 4.

Example 6

Consider the following probability distribution table.
Find the expected value of the probability distribution.

x	4	8	12	16
$P(X = x)$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$

Solution

Use the rule: $E(X) = \sum x_i p_i$

$$= \left(4 \times \frac{3}{8}\right) + \left(8 \times \frac{1}{4}\right) + \left(12 \times \frac{1}{4}\right) + \left(16 \times \frac{1}{8}\right)$$

$$= \frac{3}{2} + 2 + 3 + 2$$

$$= 8\frac{1}{2}$$

Hence $\sum x_i p_i = E(X) = 8\frac{1}{2}$

Example 7

The following table represents a probability distribution.
The expected value $E(X) = 3.4$.

x	1	2	3	4	5	6
$P(X = x)$	0.1	a	0.3	0.2	0.2	b

Find the values of a and b.

Hence write down the mode of the distribution.

Solution

Write an equation using the fact that the probabilities must add to 1: $0.1 + a + 0.3 + 0.2 + 0.2 + b = 1$

$$a + b = 0.2$$

Write an equation using $E(X)$: $E(X) = 3.4$

$$0.1 + 2a + 0.9 + 0.8 + 1 + 6b = 3.4$$

$$2a + 6b = 0.6$$

$$a + 3b = 0.3$$

Solve the equations simultaneously: $a + b = 0.2$ [1]

$a + 3b = 0.3$ [2]

[2] − [1]: $2b = 0.1$

$b = 0.05$

Substitute into [1]: $a + 0.05 = 0.2$

$a = 0.15$

Hence $a = 0.15$, $b = 0.05$

The largest probability is 0.3, so the mode is 3.

Sometimes you may want to know information about the random variable for which you have the probability distribution. In these cases, apply the function rule to the values the variable can take.

Example 8

The following table represents a probability distribution. Find the expected value of $3X - 2$.

x	2	3	4	5
$P(X = x)$	0.4	0.15	0.25	0.2

Solution

Apply the function to each value the variable can take. (This has no effect on the probability of each outcome): $f(x) = 3x - 2$

$$f(2) = 3(2) - 2 = 4$$
$$f(3) = 3(3) - 2 = 7$$
$$f(4) = 3(4) - 2 = 10$$
$$f(5) = 3(5) - 2 = 13$$

Rewrite the probability distribution table replacing the x values with the newly calculated values for $f(x)$:

x	4	7	10	13
$P[f(X) = f(x)]$	0.4	0.15	0.25	0.2

Multiply $f(x_i)$ by p_i and add the products:

(This is just a modified form of $E(X) = \sum x_i p_i$ where $f(x_i)$ replaces x_i.)

$$E(3X - 2) = 4 \times 0.4 + 7 \times 0.15 + 10 \times 0.25 + 13 \times 0.2$$
$$= 1.6 + 1.05 + 2.5 + 2.6$$
$$= 7.75$$

Hence $E(3X - 2) = 7.75$

From the previous example you can conclude $E(aX + b) = aE(X) + b$.

The following activity provides a verification of this result in a non-probability context. However, the ideas are the same for the probability application.

MAKING CONNECTIONS

Expected value $E(X)$

Use a spreadsheet to see how $E(aX + b) = aE(X) + b$.

If a probability distribution has an expected value $E(X) = 2.1$, then $E(3X - 5) = 1.3$, $E(5X + 1) = 11.5$ and $E(2X + 6) = 10.2$.

It is also worth noting that for two random variables, X and Y: $E(X + Y) = E(X) + E(Y)$.

Example 9

Consider the situation where two spinners are spun.

One spinner, X, has the following probability distribution:

x	2	3	4	5
$P(X = x)$	0.3	0.25	0.15	0.3

The other, Y, has the following probability distribution:

y	1	2	3	4
$P(Y = y)$	0.2	0.3	0.4	0.1

Show that $E(X + Y) = E(X) + E(Y)$.

Solution

The first step is to find the value of $E(X)$:

$$E(X) = 2 \times 0.3 + 3 \times 0.25 + 4 \times 0.15 + 5 \times 0.3$$
$$= 0.6 + 0.75 + 0.6 + 1.5$$
$$= 3.45$$

Now find the value of $E(Y)$:

$$E(Y) = 1 \times 0.2 + 2 \times 0.3 + 3 \times 0.4 + 4 \times 0.1$$
$$= 0.2 + 0.6 + 1.2 + 0.4$$
$$= 2.4$$

Then add these together to find the value of $E(X) + E(Y)$:

$$E(X) + E(Y) = 3.45 + 2.4$$
$$= 5.85$$

Draw a table that shows the sample space, and associated probabilities, for $X + Y$:

$Y \downarrow$	$X \rightarrow$	2	3	4	5
1		3 : 0.06	4 : 0.05	5 : 0.03	6 : 0.06
2		4 : 0.09	5 : 0.075	6 : 0.045	7 : 0.09
3		5 : 0.12	6 : 0.1	7 : 0.06	8 : 0.12
4		6 : 0.03	7 : 0.025	8 : 0.015	9 : 0.03

This table can then be used to draw up the probability distribution of $X + Y$:

$x + y$	3	4	5	6	7	8	9
$P(X + Y)$	0.06	0.14	0.225	0.235	0.175	0.135	0.03

Use the values in this probability distribution table to find $E(X + Y)$:

$$E(X + Y) = 3 \times 0.06 + 4 \times 0.14 + 5 \times 0.225 + 6 \times 0.235 + 7 \times 0.175 + 8 \times 0.135 + 9 \times 0.03$$
$$= 0.18 + 0.56 + 1.125 + 1.41 + 1.225 + 1.08 + 0.27$$
$$= 5.85$$

The values of $E(X + Y) = E(X) + E(Y)$ are equal, therefore: $E(X + Y) = E(X) + E(Y)$.

You should be able to see that it is much easier to add together the two separate expected values than to work out the distribution for the addition and then find its expected value.

Variance

The study of statistics includes **measures of central tendency** (mean, median, mode) and **measures of spread** (range, interquartile range, variance, standard deviation). You have already seen a measure of central tendency, the expected value or mean, $E(X)$ or μ. Consider now a measure of spread, the **variance**.

The variance is usually denoted by σ^2 ('sigma squared') or by $\text{Var}(X)$. In simple terms, the variance describes how far the values of a data set are spread out. The larger the variance, the more spread out the data.

The variance is defined as the expected value of the square of the difference from the mean $\text{Var}(x) = E(x - \mu)^2$.

For a finite probability distribution, to find the variance first subtract the mean μ from each value of x_i and square the result. Multiply that value by the corresponding probability p_i. The variance is the sum of these products. The further away from the mean the observed values, the greater the value of the variance.

$$\text{Var}(X) = \sigma^2$$
$$= \sum (x_i - \mu)^2 \, p_i \text{ where } \mu = E(X)$$

As $E(X) = \sum x_i p_i$, and recognising $(x_i - \mu)^2$ as a function of X: $\text{Var}(X) = E(X - \mu)^2$

This version of the variance is not always convenient, especially if $E(X)$ is already known.

Expand the perfect square:	$E(X-\mu)^2 = E(X^2 - 2X\mu + \mu^2)$
Use the property that $E(X+Y) = E(X) + E(Y)$:	$= E(X^2) - E(2X\mu) + E(\mu^2)$
As μ is a constant:	$= E(X^2) - 2\mu E(X) + E(\mu^2)$
And as $E(X) = \mu$:	$= E(X^2) - 2\mu^2 + \mu^2$
	$= E(X^2) - \mu^2$
And again, as $\mu = E(X)$:	$= E(X^2) - [E(X)]^2$

$$\begin{aligned}\text{Var}(X) &= \sigma^2\\ &= E(X-\mu)^2\\ &= \sum (x_i - \mu)^2 p_i\\ &= \sum x^2 P(x) - \mu^2\\ &= E\left(X^2\right) - \mu^2\\ &= E\left(X^2\right) - [E(X)]^2\end{aligned}$$

Use the format that is most convenient for the situation.

Example 10

For the following probability distribution:

(a) find the expected value

(b) find the variance.

x	1	2	3	4	5	6
$P(X=x)$	0.2	0.15	0.35	0.05	0.1	0.15

Solution

(a) Use $E(X) = \sum x_i p_i$:

$$\begin{aligned}E(X) &= 1\times 0.2 + 2\times 0.15 + 3\times 0.35 + 4\times 0.05 + 5\times 0.1 + 6\times 0.15\\ &= 0.2 + 0.3 + 1.05 + 0.2 + 0.5 + 0.9\\ &= 3.15\end{aligned}$$

(b) Rewrite the probability distribution table, replacing the x values with the $f(x) = x^2$ values.

x^2	1	4	9	16	25	36
$P(X=x)$	0.2	0.15	0.35	0.05	0.1	0.15

Find $E(X^2)$ by applying the rule $E[f(X)] = \sum f(x_i)p_i$:

$$\begin{aligned}E(X^2) &= 1\times 0.2 + 4\times 0.15 + 9\times 0.35 + 16\times 0.05 + 25\times 0.1 + 36\times 0.15\\ &= 0.2 + 0.6 + 3.15 + 0.8 + 2.5 + 5.4\\ &= 12.65\end{aligned}$$

Calculate $[E(X)]^2$:

$$\begin{aligned}E(X) &= 3.15\\ [E(X)]^2 &= (3.15)^2\\ &= 9.9225\end{aligned}$$

Calculate $\text{Var}(X) = E(X^2) - [E(X)]^2$:

$$\begin{aligned}\text{Var}(X) &= 12.65 - 9.9225\\ &= 2.7275\end{aligned}$$

You could also use a vertical table structure to find the value of $E(X^2)$. For the probability distribution used in this example you would have:

x	$P(X=x)$	x^2	$x^2P(X=x)$
1	0.2	1	0.2
2	0.15	4	0.6
3	0.35	9	3.15
4	0.05	16	0.8
5	0.1	25	2.5
6	0.15	36	5.4

$\sum x^2 P(X=x) = 12.65$

Var(X) must be positive. If you obtain a negative value, then there is an error.

Standard deviation

The measure of spread used most frequently is the standard deviation, measured in the same units as the variable itself. Fortunately, this value is simple to calculate if you know the variance.

The standard deviation of X is written as σ.

Recall that $\text{Var}(X) = \sigma^2$

Then the standard deviation of X, $\sigma = \sqrt{\sigma^2}$

$= \sqrt{\text{Var}(X)}$

Only the positive square root of the variance is useful, because σ is a measure of spread, which is not dependent on direction.

Example 11

The variable X has the following probability distribution:

x	–1	0	1	2	3
$P(X=x)$	0.2	0.15	0.25	0.3	0.1

Find the standard deviation correct to 2 decimal places.

Solution

Rewrite the table and add columns for $xP(X=x)$, x^2, $x^2P(X=x)$:

x	$P(X=x)$	$xP(X=x)$	x^2	$x^2P(X=x)$
–1	0.2	–0.2	1	0.2
0	0.15	0	0	0
1	0.25	0.25	1	0.25
2	0.3	0.6	4	1.2
3	0.1	0.3	9	0.9
$\sum$	1	0.95		2.55

From the table: $E(X) = 0.95$ $E(X^2) = 2.55$

Hence $[E(X)]^2 = 0.9025$

Calculate Var(X).

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$
$$= 2.55 - 0.9025$$
$$= 1.6475$$

Calculate σ using $\sigma = \sqrt{\mathrm{Var}(X)}$

$$\sigma = \sqrt{1.6475}$$
$$= 1.28 \text{ (to 2 decimal places)}$$

Just as there is a relationship between a function and the expected value based on the probability distribution, there is a relationship between a function and the standard deviation.

The connection that exists between the standard deviation of a discrete probability distribution and the standard deviation of a function based on that discrete probability distribution is summarised as:

$\sigma(aX + b) = |a|\sigma(X)$

MAKING CONNECTIONS

Standard deviation $\sigma(X)$

Use a spreadsheet to see how $\sigma(aX + b) = |a|\sigma(X)$.

One useful characteristic of the standard deviation is that, for many variables, about 95% of their values will lie within two standard deviations of the mean. The following example illustrates this.

Example 12

The following table represents the probability distribution of Y, the number of consultations per hour by a dentist.

y	0	1	2	3	4	5	6	7
$P(Y = y)$	0.015	0.05	0.255	0.321	0.179	0.155	0.016	0.009

For this distribution it has been calculated that $E(Y) = 3.173$ and $\sigma = 1.283$.

Find the probability of the values being within two standard deviations of the expected value, i.e. $P(\mu - 2\sigma \le Y \le \mu + 2\sigma)$, and comment on the result.

Solution

Calculate $\mu - 2\sigma$: $\mu - 2\sigma = 3.173 - 2 \times 1.283$
$= 0.607$

Calculate $\mu + 2\sigma$: $\mu + 2\sigma = 3.173 + 2 \times 1.283$
$= 5.739$

Rewrite the probability interval using the values just calculated: $P(0.607 \le Y \le 5.739)$

Calculate the probability from the table. (Remember, these are discrete values):

$$P(0.607 \le Y \le 5.739)$$
$$= P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) + P(Y = 5)$$
$$= 0.05 + 0.255 + 0.321 + 0.179 + 0.155$$
$$= 0.96$$

About 95% (96%) of the values lie within two standard deviations of the expected value.

EXERCISE 7.2 EXPECTED VALUE, VARIANCE AND STANDARD DEVIATION OF DISCRETE PROBABILITY DISTRIBUTIONS

1 Find **(i)** $E(X)$; **(ii)** the mode, for each of the following probability distributions.

(a)

x	1	3	5	7	9
$P(X=x)$	0.2	0.3	0.25	0.15	0.1

(b)

x	−1	0	1	2	3
$P(X=x)$	0.4	0.15	0.2	0.05	0.2

(c)

x	2	3	4	5	8
$P(X=x)$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{18}$

(d)

x	−2	−1	1	3	5
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$

2 Use the given value of $E(X)$ to solve for the unknowns.

(a) Find the values of a and b in the following probability distribution, given that $E(X) = 3.8$.

x	2	3	4	5	6
$P(X=x)$	0.4	0.1	a	0.1	b

(b) Find the values of i and j in the following probability distribution, given that $E(X) = 0.15$.

x	−3	−2	−1	0	1
$P(X=x)$	0.05	i	0.05	j	0.55

(c) Find the values of i and j in the following probability distribution, given that $E(X) = 6\frac{2}{5}$.

x	2	4	6	8	10
$P(X=x)$	$\frac{3}{20}$	$\frac{5}{20}$	i	$\frac{7}{20}$	j

(d) Find the values of a and b in the following probability distribution, given that $E(X) = 1\frac{4}{15}$.

x	−3	−1	1	3	5
$P(X=x)$	a	$\frac{1}{5}$	$\frac{1}{15}$	b	$\frac{1}{5}$

3 For the following probability distribution, find:

x	0	1	2	3	4
$P(X=x)$	0.25	0.3	0.2	0.15	0.1

(a) $E(X)$ **(b)** $E(2X-3)$ **(c)** $E(X^2-5)$ **(d)** $E(X^2+X-4)$

4 For the following probability distributions find the expected value and variance.

(a)

x	1	2	3	4	5	6
$P(X=x)$	0.1	0.3	0.25	0.05	0.15	0.15

(b)

x	5	6	7	8	9
$P(X=x)$	0.15	0.35	0.1	0.25	0.15

5 Find the standard deviation of the variable X which has the following probability distribution. Give your answer correct to 2 decimal places.

x	−2	−1	0	1	2	3
$P(X=x)$	0.1	0.15	0.3	0.15	0.2	0.1

6 For the following probability distribution, find:

y	-3	-2	-1	0	1	2
$P(Y = y)$	0.02	0.03	0.25	0.35	0.3	0.05

(a) the standard deviation of Y **(b)** $P(\mu - 2\sigma \le Y \le \mu + 2\sigma)$.

7 In each of the following probability distributions, find the values of k and $E(X)$.

(a)

x	1	2	3	4	5
$P(X = x)$	k	$2k$	$3k$	$12k$	$6k$

(b)

x	-3	-1	1	3	5
$P(X = x)$	$12k$	$2k$	$3k$	k	$2k$

8 For the following probability distribution, find:

(a) $E(W)$ **(b)** $E(3W - 4)$
(c) $E(2W + 5)$ **(d)** $E(W^2 - 7)$

w	2	4	6	8
$P(W = w)$	$\frac{5}{16}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{16}$

9 Variable X has the following probability distribution:

x	-2	-1	0	1
$P(X = x)$	0.3	0.4	0.1	0.2

What is the value of $E(X^2 - 4)$?

A 0.2 **B** -5.4 **C** 2.2 **D** -2.2

10 Variable Y has the following probability distribution:

y	0	1	2	3
$P(Y = y)$	$\frac{1}{9}$	$\frac{7}{18}$	$\frac{1}{3}$	$\frac{1}{6}$

What is the value of $\text{Var}(Y)$?

A $1\frac{5}{9}$ **B** $1\frac{153}{162}$ **C** $\frac{65}{81}$ **D** $2\frac{5}{38}$

11 Eric just got a new job selling cars. He is offered a choice of two salary packages. In the first package he receives a weekly retainer of \$200 and an additional \$650 for every car sold. In the second package his retainer would be \$400, but he would only receive \$400 for every car sold. Past sales patterns indicate that the probability distribution for the number of cars sold per week is as follows:

Number of vehicles	0	1	2	3	4	5
Probability	0.45	0.35	0.1	0.05	0.04	0.01

Which salary package would Eric be better off taking?

12 A biased die has the following probability distribution:

d	1	2	3	4	5	6
$P(D = d)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$

(a) If you rolled this die twenty times, what would be the expected (mean) value of the data, stated in mixed number form?
(b) State the range of $\mu \pm 2\sigma$ correct to 2 decimal places.
(c) Does the range of $\mu \pm 2\sigma$ cover all possible results when rolling the die? Should any of the values be considered unusual?

13 A random variable, T, has the following probability distribution:

t	$w - 3$	$w - 2$	$w - 1$	w	$w + 1$
$P(T = t)$	0.2	0.5	0.1	0.05	0.15

(a) Given that $E(T) = 8.45$, find the value of w.
(b) Find $\text{Var}(T)$.
(c) Find the standard deviation of T correct to 2 decimal places.
(d) Find $\text{Var}(2T - 6)$ correct to 2 decimal places.
(e) Find $\text{Var}(5 - 3T)$ correct to 2 decimal places.

14 The probability distribution of G is given by:

$$P(G=g)=\begin{cases}k(6-g), & \text{if } g\in\{0,1,2,3,4\}\\ 0, & \text{for all other values of } g\end{cases}$$

Find the following values: **(a)** k **(b)** $E(G)$ **(c)** $\mathrm{Var}(G)$ **(d)** σ

15 Two spinners are spun. One spinner, X, has the following probability distribution:

x	1	2	3	4
$P(X=x)$	0.2	0.2	0.1	0.5

The other spinner, Y, has the following probability distribution:

y	2	3	4	5
$P(Y=y)$	0.2	0.3	0.4	0.1

Show that $E(X+Y)=E(X)+E(Y)$.

16 Enrico runs a game of chance at the Sydney Show. The player pays a fee to play, and draws a card at random from a normal pack of 52 playing cards. If the card is black, the player gets \$1 and their fee is refunded. If the card is a diamond, the player gets \$5 and their fee is refunded.

(a) Find Enrico's loss for the following conditions:
  (i) the card the player chooses is black
  (ii) the card the player chooses is a diamond.
(b) Describe the event for which Enrico keeps the game fee.
(c) Let Enrico's profit for the event described in part **(b)** be \$$p$, and draw up a probability distribution table that shows the three possible outcomes from Enrico's point of view.
(d) Find the value of p so that the expected value for the distribution is 0.
(e) If Enrico wants to make a profit, what is the minimum whole dollar amount he should charge to play the game?

17 Dubravko and Erina often play a best-of-three-sets match of tennis. From past experience they know that the probability that Dubravko will win a set is $\frac{3}{5}$.

(a) Draw a tree diagram to show the possible outcomes for their three-set match.
(b) Use the results from your tree diagram to draw a probability distribution table for their matches. Let X stand for the number of sets played.
(c) Find the expected number of sets they play in a match.
(d) Comment on your answer to part **(c)** in real-life terms.

18 A spinner has nine equal sections, of which five are yellow, three are blue and one is red. If the spinner lands on yellow, you receive \$1. If it lands on blue you receive \$3 and if it lands on red you receive \$5. Let X stand for the amount of money you receive.

(a) Draw up a probability distribution table for this game.
(b) What is the expected value of X?
(c) If the game is to be fair, how much should you pay to play?
(d) Comment on your answer to part **(c)** in real-life terms.

19 The game of 'Take a Chance' requires the player to roll three dice, of which one is blue, one is red and one is white. If a 1 shows on the blue die the player receives \$1, if a 1 shows on the red die the player receives \$2 and if a 1 shows on the white die the player receives \$5. In all other circumstances the player receives nothing. A player can receive more than one prize.

(a) Find the probability that the player receives the following amounts:
  (i) \$1 **(ii)** \$2 **(iii)** \$3 **(iv)** \$5 **(v)** \$6 **(vi)** \$7 **(vii)** \$8 **(viii)** \$0
(b) What is the expected return on this game?
(c) How much should the operator charge to play if the game is fair?

20 The number, Y, appearing on a spinner showing the numbers 1, 2, 3, 4, has a probability distribution as follows:

y	1	2	3	4
$P(Y = y)$	0.1	0.2	0.3	0.4

For another spinner, the number appearing, X, has a probability distribution as follows:

x	0	1	2	3	4
$P(X = x)$	0.6	0.1	0.1	0.1	0.1

(a) Find $E(Y)$.
(b) Find $E(X)$.
(c) The numbers appearing on each of the two spinners are added together to give the variable M.
  (i) Draw a table to show the probability distribution of M.
  (ii) Find $E(M)$.
(d) What connection is there between $E(X)$, $E(Y)$ and $E(M)$?

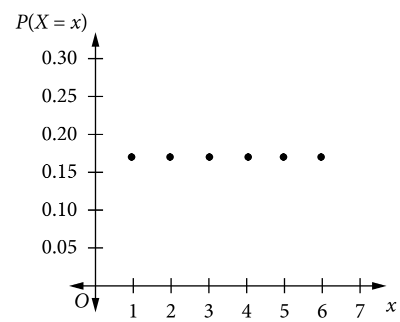
7.3 THE UNIFORM DISTRIBUTION

There are some discrete distributions that have special properties worthy of separate investigation—**uniform distribution** and binomial distribution. In this section, you will look at uniform distribution.

A discrete probability distribution is said to be uniform if all values of the random variable are equally likely.

A common example of a uniform distribution is the random variable, X, that is the value of the face showing when a normal, six-sided die is rolled.

$$P(X = x) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{for all other values of } x \end{cases}$$



The graph of this distribution is shown on the right:

The expected value is: $E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$

$$= \frac{21}{6}$$

$$= 3\frac{1}{2}$$

This can also be determined from the symmetry of the graph of the distribution.

The variance of the distribution is given by $\text{Var}(X) = E(X^2) - [E(X)]^2$:

$$\text{Var}(X) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) - \left(\frac{7}{2}\right)^2$$

$$= \frac{35}{12}$$

In this text a uniform distribution will be assumed to be based on the first n positive integers. This is so some general expressions can be established. The general expression for a discrete uniform distribution with n values is:

$$P(X = x) = \begin{cases} \frac{1}{n} & \text{if } x \in \{1,2,\ldots,n\} \\ 0 & \text{for all other values of } x \end{cases}$$

If you use the fact that the sum of the first n positive integers $1 + 2 + \ldots + n = \frac{n}{2}(n + 1)$, then the expected value is given by:

$$E(X) = 1 \times \frac{1}{n} + 2 \times \frac{1}{n} + 3 \times \frac{1}{n} + \ldots + n \times \frac{1}{n}$$

$$= \frac{1}{n}(1 + 2 + 3 + \ldots + n)$$

$$= \frac{1}{n} \times \frac{n}{2}(n + 1)$$

$$= \frac{n+1}{2}$$

The variance of X can be obtained by evaluating $\text{Var}(X) = E(X^2) - [E(X)]^2$ to get: $\text{Var}(X) = \frac{n^2-1}{12}$

Technology can be used to establish the expected value and variance of a discrete uniform distribution with n values. The probability function for this distribution is $P(X = x) = \frac{1}{n}$ for $x \in \{1, 2, 3, \ldots, n\}$ and zero otherwise.

The general form for the mean of a finite discrete distribution is: $E(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n = \sum_{i=1}^{n} x_ip_i$

For the uniform distribution with n values this becomes: $\sum_{i=1}^{n} x_ip_i = \sum_{i=1}^{n} i\frac{1}{n} = \sum_{i=1}^{n} \frac{i}{n}$

EXPLORE FURTHER

Expected value and variance of a uniform distribution

Use a spreadsheet to find the expected value and variance of the first n natural numbers.

For a discrete uniform probability distribution with n values, 1 to n:

$E(X) = \frac{n+1}{2}$ $\quad\quad$ $\text{Var}(X) = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$

Example 13

A roulette wheel in the United States usually has 38 equal-sized spaces showing the numbers 1 to 36 as well as 0 and 00. When the wheel is spun, a ball will land in one of the 38 spaces at random. For this question assume that 0 and 00 represent the 37th and 38th possible outcomes.

(a) Find the mean of the number of the space the ball lands in.

(b) Find the variance of the number of the space the ball lands in.

Solution

(a) Use the rule for $E(X)$:

$$E(X) = \frac{n+1}{2} = \frac{38+1}{2} = \frac{39}{2} = 19\frac{1}{2}$$

(b) Use the rule for $\text{Var}(X)$:

$$\text{Var}(X) = \frac{n^2-1}{12} = \frac{38^2-1}{12} = \frac{1443}{12} = 120\frac{1}{4}$$

This section has focused on uniform distributions where the values of x are 1 to n. However, it is still quite easy to find the expected value and variance for other uniform distributions where the values the distribution takes are consecutive numbers. This is because such a distribution is a lateral shift of the uniform distribution where x takes the values 1 to n, so the rules $E(X + b) = E(X) + b$ and $\text{Var}(X + b) = \text{Var}(X)$ can be applied.

EXERCISE 7.3 THE UNIFORM DISTRIBUTION

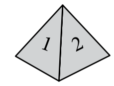
1 A number from 1 to 16 is chosen at random. The random variable, R, represents the value chosen. Find the following values:

(a) $E(R)$ $\quad\quad$ **(b)** $\text{Var}(R)$

2 A cleaner has nine similar-looking keys on a key chain. He tries them in turn, until he finds the one that opens the lock.

(a) What is the expected number of attempts for the cleaner to open the lock?
(b) What is the variance of the number of attempts?

3 A die in the shape of a tetrahedron (a solid with four triangular faces) is rolled. The four faces are numbered 1 to 4. Let F be a random variable that represents the value that is face down on the table.



(a) What is the expected value, $E(F)$? **(b)** What is the variance, $\mathrm{Var}(F)$?

4 A spinner is equally divided into n segments and each segment contains a value from 1 to n.

(a) If the expected value is 11.5, then n is equal to which of the following values?
A 20 **B** 21 **C** 22 **D** 23

(b) If the variance is 14 then n is equal to which of the following values?
A 9 **B** 10 **C** 12 **D** 13

5 Consider the random variable X that has the following probability distribution:

x	1	2	3	4	5
$P(X = x)$	0.2	0.2	0.2	0.2	0.2

(a) Find the following values:
(i) $E(X)$ **(ii)** $\mathrm{Var}(X)$

(b) Now consider the random variable Y that has the following probability distribution:
Find, from first principles, the following values:

y	2	3	4	5	6
$P(Y = y)$	0.2	0.2	0.2	0.2	0.2

(i) $E(Y)$ **(ii)** $\mathrm{Var}(Y)$

(c) What can you say about the values of $E(X)$ and $E(Y)$?
(d) What can you say about the values of $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$?
(e) Now consider the random variable Z that has the following probability distribution:
Without doing any additional calculations, determine the following values:

z	6	7	8	9	10
$P(Z = Z)$	0.2	0.2	0.2	0.2	0.2

(i) $E(Z)$ **(ii)** $\mathrm{Var}(Z)$

6 You enter a room that contains a digital clock, such as the one shown. Let T represent the minutes shown on the clock.

(a) Show, using symmetry, that $E(T) = 29.5$.
(b) Now calculate $\mathrm{Var}(T)$ and $\sigma(T)$, correct to 3 decimal places.
(c) If the numbers shown were 1 to 60, instead of 0 to 59, how would this effect $\mathrm{Var}(T)$ and $\sigma(T)$?

7 A six-sided die numbered 1 to 6 is rolled twice and the values obtained are added together.

(a) Construct a probability distribution table for this event, using S to represent the variable.
(b) Draw a bar chart to illustrate the distribution.
(c) How would you describe this distribution?
(d) Calculate the following values:
(i) $E(S)$ **(ii)** $\mathrm{Var}(S)$
(e) Recall that for a single roll of such a die: $E(X) = 3\frac{1}{2}$ and $\mathrm{Var}(X) = 2\frac{11}{12}$. Comment on the relationship between the values for $E(X)$, $\mathrm{Var}(X)$, $E(S)$ and $\mathrm{Var}(S)$.
(f) Consider a spinner that has four equally-sized sections labelled 1 to 4. Let V be the value the spinner stops on. Calculate the following values:
(i) $E(V)$ **(ii)** $\mathrm{Var}(V)$
(g) The spinner is now spun twice. Let T be the sum of the two values obtained. Calculate the following values:
(i) $E(T)$ **(ii)** $\mathrm{Var}(T)$
(h) Comment on the relationship between the values for $E(V)$, $\mathrm{Var}(V)$, $E(T)$ and $\mathrm{Var}(T)$.
(i) Can you now make a general comment on the values obtained for the mean (expected value) and variance when two identical uniform variables, with values between 1 and n, are added?

Example 14

Two standard dice are rolled and the variable X represents the number of sixes obtained.

Find the expected number of sixes obtained.

Solution

Write the sample space:

$$\begin{Bmatrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6) \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{Bmatrix}$$

There are 36 outcomes in the sample space.
Draw a table showing the probability distribution of the random variable.
Leave the probabilities in fraction form with the same denominator to make calculations easier:

x	0	1	2
$P(X=x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

Calculate $E(X)$ from first principles. Express the final answer as a fraction in simplest form:

$$\begin{aligned} E(X) &= 0\times\frac{25}{36}+1\times\frac{10}{36}+2\times\frac{1}{36} \\ &= \frac{12}{36} \\ &= \frac{1}{3} \end{aligned}$$

The expected number (or average number) of sixes in two rolls of a standard die is $\frac{1}{3}$.

Example 15

One of the games at the local sporting club's 'Vegas Night' involves rolling a standard six-sided die. If an even number is shown, there is no game charge and the player wins the number of dollars shown on the face of the die. If an odd number is shown, the cost of playing is the number of dollars shown on the face of the die. Let Z stand for the number of dollars received by the player.

(a) Draw up a probability distribution table.

(b) What type of distribution is this?

(c) Find $E(Z)$.

(d) Is the game fair?

Solution

(a) A win could be represented by a positive value and a loss by a negative value.
Draw up the probability distribution table:

z	−1	2	−3	4	−5	6
$P(Z=z)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(b) The probabilities are all equal, so this is a uniform distribution.

(c) The rule for $E(Z)$ cannot be used since the values are not $(1, 2, 3, \ldots, n)$. Express the final answer as a fraction in simplest form:

$$E(Z) = -1 \times \frac{1}{6} + 2 \times \frac{1}{6} - 3 \times \frac{1}{6} + 4 \times \frac{1}{6} - 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$$
$$= \frac{3}{6}$$
$$= \frac{1}{2}$$

(d) The game is not fair. It is actually in favour of the player since the expected return is $0.50.

EXPLORE FURTHER

Mean, variance and standard deviation of a discrete distribution

Use a spreadsheet to find the mean, variance and standard deviation for a discrete probability distribution.

EXERCISE 7.4 DISCRETE DISTRIBUTIONS IN PRACTICAL SITUATIONS

1 A die, labelled 1 to 6, is rolled until the total of the scores is 4 or greater. Answer each of the following, giving all answers as exact values, in simplest fraction form.

(a) Find the probability distribution of the number of rolls X required to achieve this total.
(b) Find the expected number of rolls required.
(c) Find the variance for the number of rolls required.

2 One of the games at the local sporting club's 'Vegas Night' involves rolling a standard six-sided die. If a non-prime number is shown, there is no game charge and the player wins the number of dollars shown on the face of the die. If a prime number is shown, the cost of playing is the number of dollars shown on the face of the die. Let Z stand for the number of dollars received by the player.

(a) Draw a table to show the probability distribution of the variable.
(b) Is this best described as a uniform or a non-uniform distribution?
(c) Find the value of $E(Z)$.
(d) Is this game fair? If not, is it biased in favour of the operator or the player?

3 Four cards are labelled from 1 to 4. Two cards are dealt at random, without replacement. Let X represent the larger of the two numbers shown on the cards.

(a) How many values can X take?
(b) Find $P(X = 2)$.

4 For the discrete random variable X, the probability distribution is given by:

$$P(X = x) = \begin{cases} kx, & x = 1,2,3,4 \\ k(9 - x), & x = 5,6,7,8 \end{cases}$$

(a) Find the value of k.
(b) Complete the following table to show the probability distribution of X.

x	1	2	3	4	5	6	7	8
$P(X = x)$								

(c) Find the value of $E(X)$.
(d) Find the value of $\text{Var}(X)$.

5 The ratio boys : girls in a particular town was found to be 11 : 10, where the gender of one child in the family is independent of the gender of any other child in the family, and all the children are either boys or girls.

(a) State all possible combinations of boys and/or girls for a family with three children.
(b) What is the probability that a family with three children will have at least one boy?
(c) What proportion of families with exactly 4 children will have at least 3 girls?
(d) What proportion of families with exactly 4 children will have 2 girls and 2 boys?

6 Tomino has written the following as his answer for the probability distribution of the random variable X:

$$P(X = x) = \begin{cases} \frac{4-x}{5}, & \text{for } x = 1, 2, 3, 4, 5 \\ 0, & \text{otherwise} \end{cases}$$

Explain to Tomino why this cannot be correct.

7 The discrete random variable X has the probability distribution shown in the following table.

x	1	2	3
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(a) Find $E(X)$.
(b) Find $\text{Var}(X)$.
(c) A second random variable Y has the same distribution as X, and the two variables are independent. Draw a table to show the probability distribution of $X + Y$.
(d) Find $E(X + Y)$.
(e) Find $\text{Var}(X + Y)$.
(f) How is the value of $E(X + Y)$ related to the values of $E(X)$ and $E(Y)$?
(g) How is the value of $\text{Var}(X + Y)$ related to the values of $\text{Var}(X)$ and $\text{Var}(Y)$?

8 A standard six-sided die is rolled twice.

(a) How many ordered pairs make up the sample space?
(b) Assign Z to be the maximum number in each ordered pair. Draw a table to show the probability distribution of Z. Write your answers in fraction form using a denominator of 36.
(c) Now assign Y to be the minimum number in each ordered pair. Draw a table to show the probability distribution of Y. Write your answers in fraction form using a denominator of 36.
(d) How is the distribution of Z related to the distribution of Y?
(e) Find the expected value for each of the following:
(i) Z **(ii)** Y
(f) How far from the greatest value of Z is $E(Z)$?
(g) How far from the least value of Y is $E(Y)$?
(h) Comment on your answers to part **(f)** and part **(g)**.
(i) Given $\text{Var}(Z) = 1\frac{926}{1296}$, what can you say about $\text{Var}(Y)$?

9 Stephan has made a game in which the probability of randomly picking a number from 0 to 5 is given by the probability distribution shown in the following table. Answer each of the following, giving all answers correct to 3 decimal places.

X	0	1	2	3	4	5
$P(X = x)$	0.002	0.076	0.293	0.268	a	0.098

(a) Calculate the expected value for this random variable.
(b) Leanne made a game similar to Stephan's, but the probability of randomly picking a number from 0 to 5 is given by the following probability distribution.

Y	0	1	2	3	4	5
$P(Y = y)$	0.005	0.029	0.047	0.219	0.386	0.314

Calculate the expected value for this random variable.

(c) If one value in Stephan's game and one value in Leanne's game are chosen at random, calculate the probability that:
  (i) they are the same value (ii) they are different (iii) their sum is greater than 8.

10 The discrete random variable X has the following probability distribution.

x	1	2	3	4
$P(X=x)$	0.25	0.1	0.45	0.2

(a) Find $E(X)$.
(b) Verify that $E(2X) = 2E(X)$.
(c) Find $\text{Var}(X)$.
(d) Find the following values:
  (i) $\text{Var}(2X)$ (ii) $\text{Var}(3X)$
(e) State the relationship between $\text{Var}(X)$ and $\text{Var}(kX)$.

11 The probability distribution table of a random variable X is shown. Answer each of the following, giving all answers correct to 3 decimal places where necessary.

X	n	$n+1$	$n+2$	$n+3$
$P(X=x)$	0.80	0.12	0.05	0.03

(a) Show that $E(X) = n + 0.31$.
(b) If two independent values of X are chosen at random, calculate the probability of choosing two consecutive values.
(c) If two independent values of X are chosen at random, calculate the probability that the sum of the two values is even.
(d) If four independent values of X are chosen at random, calculate the probability that they are one of each type.

12 A biased four-sided die is rolled. The following table gives the probability of each score.

Score	1	2	3	4
Probability	$\frac{7}{16}$	$\frac{5}{16}$	3^k	3^{2k+1}

What is the probability of rolling 3?

A $\frac{1}{4}$ B $\frac{1}{6}$ C $\frac{1}{8}$ D $\frac{1}{5}$

13 At a market, a stallholder is hosting free-to-play mini games.

If the player wins against the stallholder, they will receive \$25 as a reward. If they lose, they must pay \$5 to the stallholder. If the game ends in a draw, no money is exchanged.

The random variable, X, represents the amount of money a player wins per mini game, where the probability of winning a game is independent of the result of the previous game.

The distribution of X is given in the table below.

x	-5	0	25
$P(X=x)$	$0.29-2k^2$	0.87	$k-0.28$

(a) Show that $2k^2 - k + 0.12 = 0$.
(b) Find the value of k, giving reasons for your answer.
(c) Find $E(X)$.
(d) A player decides to play three mini games.
  (i) How much is the player expected to win or lose?
  (ii) What is the probability that the player will lose exactly two of the three games?

14 The table shows the probability distribution of a discrete random variable, X.

x	2	4	10	20
$P(X = x)$	k	0.05	0.35	$3k$

(a) Show that $k = 0.15$.
(b) State the mode.
(c) Find the expected value.

7.5 CONTINUOUS PROBABILITY DISTRIBUTIONS

When a variable can take any value in a particular interval, for example when you measure it rather than count it, you say you have a **continuous random variable**. Quantities which can be modelled using continuous random variables include height, weight, time and mass.

A continuous random variable is defined by its **probability density function** ('pdf') which is usually represented by $f(x)$. A continuous random variable is defined over the interval $(-\infty, \infty)$ although, in practice, it is common that the probability associated with much of this interval is 0. As a result, $f(x)$ is usually defined as a hybrid, or 'piece-wise', function.

To help understand continuous random variables, consider the following example.

Example 16

Dirk travels to work by train every day. He is interested in mathematics and decides to collect some data about the lateness of his train service. Over a ten-week period (50 work days) he found the following, where $0 - < 2$ means from 0 to less than 2 minutes late.

Minutes late	0–<2	2–<4	4–<6	6–<8	8–<10
Frequency	5	15	17	10	3

(Dirk's train service runs every 10 minutes so a train cannot be more than 10 minutes late.)

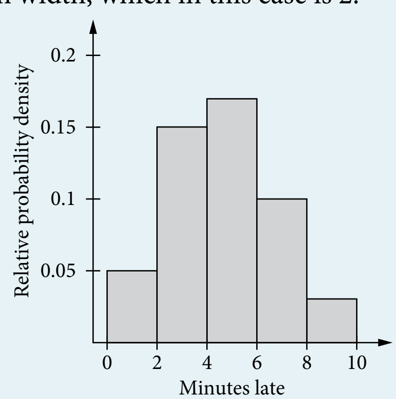
If Dirk wants to calculate the probability of his train being less than 4 minutes late, he can add together the 5 and 15 and say the probability is $\frac{20}{50} = \frac{2}{5}$.

This is written as: $P(\text{train is less than 4 minutes late}) = \frac{2}{5}$

This is a simple calculation. But how can Dirk calculate the probability of his train being less than 3 minutes late? The grouped data table is not useful for calculations within individual data bins.

Dirk thinks a histogram might help, but decides to use the relative probability density values for the vertical axis. The relative probability values are found by dividing the probabilities by the bin width, which in this case is 2.

Minutes late	0–<2	2–<4	4–<6	6–<8	8–<10
Probability	0.1	0.3	0.34	0.2	0.06
Relative probability density	0.05	0.15	0.17	0.1	0.03



If you calculate the area of each of the rectangles in this histogram showing relative frequencies you will find that their sum is 1. This is because the total area represents the total probability of all possibilities.

Does this help Dirk find the probability of his train being less than 3 minutes late? No, but it gives Dirk another idea. He decides to model the data using a parabola, as he can imagine a negative parabola (concave down) running over the histogram. Dirk decides that the turning point for his parabola will be (5, 0.17) as this represents the median of the time values and corresponds to the highest point in the histogram. This gives an equation in the form $y = -k(x - 5)^2 + 0.17$.

Dirk now has to find the value of k. The area under the curve must be 1, so calculus can be used to express the area as an integral.

This can now be solved to find the value of k: $\int_0^{10} \left(-k(x-5)^2 + 0.17\right) dx = 1$

Evaluate the integral: $$\left[\frac{-k(x-5)^3}{3} + 0.17x\right]_0^{10} = 1$$

$$\frac{-k \times 5^3}{3} + 1.7 - \left(\frac{-k \times (-5)^3}{3} + 0\right) = 1$$

$$\frac{-125k}{3} + 0.7 - \frac{125k}{3} = 0$$

$$250k = 2.1$$

$$k = \frac{2.1}{250} = 0.0084$$

This gives the probability density function (that is, the parabola that models the data) as:

$$f(x) = \begin{cases} -0.0084(x-5)^2 + 0.17 & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

To check the accuracy of this Dirk can calculate the probability for each of the intervals in the original grouped data table.

$$\int_0^2 \left(-0.0084(x-5)^2 + 0.17\right) dx = \left[\frac{-0.0084(x-5)^3}{3} + 0.17x\right]_0^2 = 0.0656$$

$$\int_2^4 \left(-0.0084(x-5)^2 + 0.17\right) dx = \left[\frac{-0.0084(x-5)^3}{3} + 0.17x\right]_2^4 = 0.2672$$

$$\int_4^6 \left(-0.0084(x-5)^2 + 0.17\right) dx = \left[\frac{-0.0084(x-5)^3}{3} + 0.17x\right]_4^6 = 0.3344$$

$$\int_6^8 \left(-0.0084(x-5)^2 + 0.17\right) dx = \left[\frac{-0.0084(x-5)^3}{3} + 0.17x\right]_6^8 = 0.2672$$

$$\int_8^{10} \left(-0.0084(x-5)^2 + 0.17\right) dx = \left[\frac{-0.0084(x-5)^3}{3} + 0.17x\right]_8^{10} = 0.0656$$

Minutes late	0–<2	2–<4	4–<6	6–<8	8–<10
Probability	0.1	0.3	0.34	0.2	0.06
Calculated probability	0.07	0.27	0.33	0.27	0.07

Dirk is satisfied with this model as the values are close, so he uses it to answer his original question:

$$P(\text{train is less than 3 minutes late}) = \int_0^3 \left(-0.0084(x-5)^2 + 0.17\right) dx = \left[\frac{-0.0084(x-5)^3}{3} + 0.17x\right]_0^3 = 0.1824$$

Correct to 2 decimal places, P(train is less than 3 minutes late) = 0.18.

The model in the example above cannot be used to calculate the probability that, for example, the train will be *exactly* 3 minutes late, as the upper and lower limits in the integral would be the same and hence its value would be zero. Models like this must deal with intervals, even if the interval is very small.

> If X is a continuous random variable then $P(X = x) = 0$, for all possible values of x.
> The probability is represented by area, so zero width results in zero area.

The function f is the probability density function ('pdf') of the variable X. It is important to note that $f(x)$ is not the probability. You need to integrate between two limit values to obtain a probability.

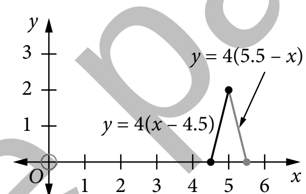
> As $P(X = x) = 0$, all of the following expressions have the same value:
>
> $P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < x < b)$

As seen above, you can find the probabilities associated with continuous probability density functions by integrating the function between the values that specify an interval.

A quadratic model will not suit every data set. It may be that a linear model will work best, or a model based on some other mathematical function, perhaps even a piece-wise (hybrid) function.

An example of this type of function would be one that follows a 'triangular' distribution. Consider the following piece-wise function and a sketch of its graph:

$$f(x) = \begin{cases} 4(x - 4.5), & 4.5 < x \le 5.0 \\ 4(5.5 - x), & 5.0 < x \le 5.5 \\ 0, & \text{otherwise} \end{cases}$$



An inspection of the graph will reveal that the area under the graph is 1, as required for the function to represent a probability density function.

> For a function $f(x)$ to be a probability density function:
>
> - $f(x) \ge 0$ for all values of x
> - $\int_{-\infty}^{\infty} f(x)dx = 1$, i.e. the area enclosed by the graph $y = f(x)$ and the x-axis is equal to 1.

Note that a probability density function can take on values greater than 1. You must remember that it is the *area* bounded by the curve and the x-axis that must be 1.

Cumulative distribution function

This then leads to the Cumulative Distribution Function, $F(x) = p(X \le x) = \int_0^x f(x)dx$, where $f(x)$ is a probability distribution function.

As well, $P(a \le x \le b) = F(b) - F(a)$

The Cumulative Distribution Function, $F(x)$, is a non-decreasing function for all x in its domain as its least value is 0 and its greatest value is 1.

Thinking back to Dirk and his investigation into the lateness of his train (in Example **16**), to find P(train less than three minutes late) he needed to calculate $\int_0^3 f(x)\,dx$ and to find

P(train less than five minutes late) the calculation would be $\int_0^5 f(x)\,dx$.

To save recalculating $\int f(x)\,dx$ each time, a new function, the **cumulative distribution function** ('cdf'), can be defined. The cdf is usually designated as $F(x)$. For Dirk's train investigation the following would apply:

$f(x) = -0.0084(x-5)^2 + 0.17$, or after anti-differentiating the function:

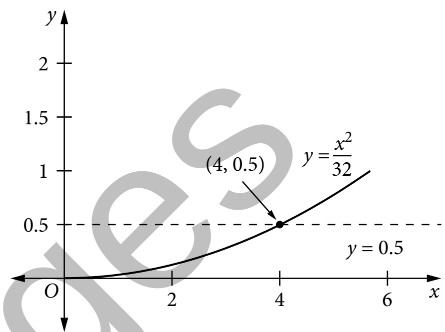
$$F(x) = \begin{cases} -0.0028x^3 + 0.042x^2 - 0.04x, & 0 \le x \le 10 \\ 0, & \text{elsewhere} \end{cases}$$

Note that $0 \le F(x) \le 1$, as the probability must be between 0 and 1.

The graphs of $F(x)$ and $y = 0.5$ can be graphed on the same set of axes as an alternative way of finding the median.

For the probability density function in the previous example, the corresponding cumulative density function would be:

$$F(x) = \begin{cases} \dfrac{x^2}{32}, & 0 \le x \le 4\sqrt{2} \\ 0, & \text{otherwise} \end{cases}$$



The value of the median is the x-coordinate of the point of intersection, which is the same value as found earlier.

Similarly, for the first quartile, find the x-coordinate of the point of intersection of $y = F(x)$ and $y = 0.25$.

For the third quartile, find the x-coordinate of the point of intersection of $y = F(x)$ and $y = 0.75$.

Calculating a value associated with a continuous density function

Example 17

A particular continuous random variable has the following probability density function:

$$f(x) = \begin{cases} -\dfrac{3}{2}(x-1)^2 + \dfrac{3}{2}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the cumulative distribution function, CDF.

(b) Sketch its graph.

(c) On the graph show the point that represents $P(X \le 0.7)$.

(d) Find $P(X \le 0.7)$.

Solution

(a) Expand the quadratic expression: $-\dfrac{3}{2}(x-1)^2 + \dfrac{3}{2} = -\dfrac{3}{2}\left(x^2 - 2x + 1 - 1\right)$

$$= -\frac{3}{2}\left(x^2 - 2x\right)$$

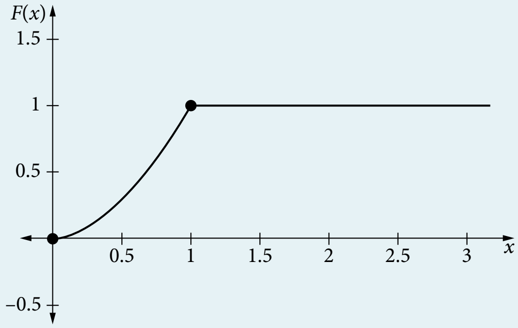
$$= \frac{3}{2}\left(2x - x^2\right)$$

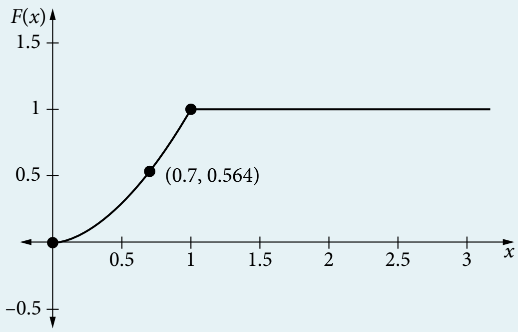
$$F(x) = \int_0^x \left(\frac{3}{2}(2t - t^2)\right) dt$$

$$= \frac{3}{2}\left[t^2 - \frac{t^3}{3}\right]_0^x$$

$$= \frac{3x^2}{2} - \frac{x^3}{2}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{3x^2}{2} - \dfrac{x^3}{2} & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

(b)


(c)


(d) Find the required integral: $\int_0^{0.7} \frac{3}{2}\left(2x - x^2\right) dx$

$$= \frac{3}{2}\left[x^2 - \frac{x^3}{3}\right]_0^{0.7}$$

$$= [1.5 \times (0.7)^2 - 0.5 \times (0.7)^3] - 0$$

$$P(X \leq 0.7) = 0.5635 \approx 0.564$$

There will be times when the pdf will not be able to be integrated to find the CDF. In these situations approximate methods of integration like the trapezoidal rule or a graphing application like FX graph may be used.

Example 18

A particular continuous random variable has the pdf given by $f(x) = xe^x$, for $0 \leq x \leq 1$.

(a) Write down the integral for the cumulative distribution function.

Using graphing software, find:

(b) $P(X \leq 0.8)$, correct to 3 decimal places

(c) the mean of the distribution

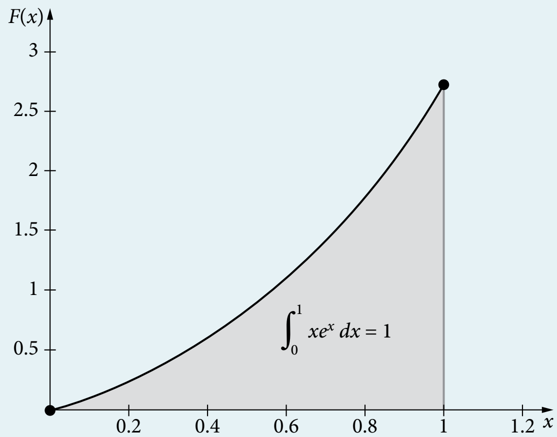
(d) the standard deviation of the distribution.

Solution

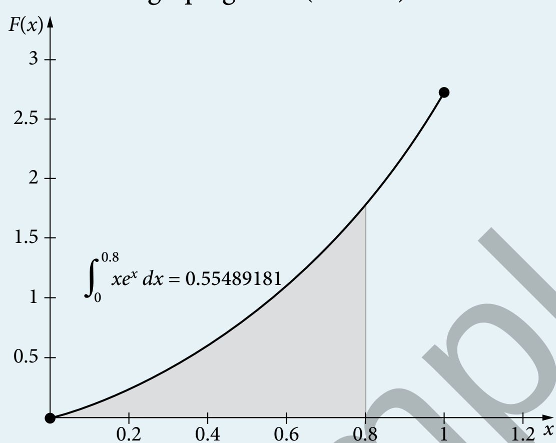
(a) $f(x) = xe^x, 0 \le x \le 1$: $\quad F(x) = \int_0^x te^t dt, 0 \le x \le 1.$

(b) Graphing $F(x)$, the following is obtained.

The first graph verifies that $f(x)$ is a pdf.



The second graph gives $P(X \le 0.8)$.

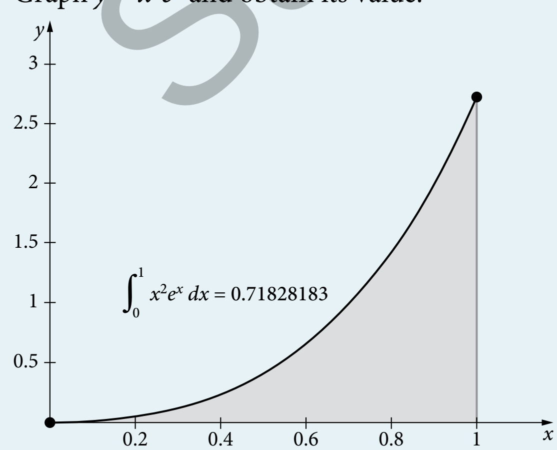


$P(X \le 0.8) = 0.555$

(c) Find the expression for $xf(x)$: $\quad xf(x) = x \times xe^x = x^2e^x$ for $0 \le x \le 1$.

$E(x) = \int_0^1 x^2e^x dx$

Graph $y = x^2e^x$ and obtain its value.

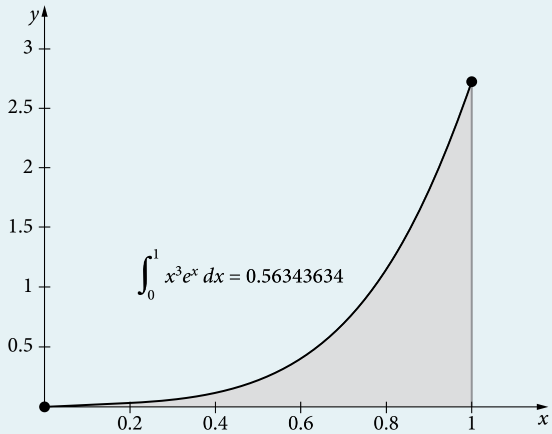


$\mu = E(X) = 0.718$

(d) Find the value of the integral $\int_0^1 x^2 f(x)dx$.

$\int_0^1 x^2 f(x)dx = \int_0^1 x^3 e^x dx$

Graph $y = x^3 e^x$ and obtain its value.



$\int_0^1 x^2 f(x)dx = \int_0^1 x^3 e^x dx = 0.563$

$\mu = 0.718$ so $\mu^2 = 0.5155$

$\sigma^2 = 0.563 - 0.5155 = 0.0475$

$\sigma = \sqrt{0.0475} = 0.2179 \approx 0.22$

Calculating a value associated with a continuous density function for a piece-wise function

Example 19

A particular continuous random variable has the following probability density function:

$$f(x) = \begin{cases} 4(x-4.5), & 4.5 < x \le 5.0 \\ 4(5.5-x), & 5.0 < x \le 5.5 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find $P(X \le 4.7)$.

(b) Find $P(X \le 5.2)$.

Solution

Define the piece-wise function: $f(x) = \begin{cases} 4(x-4.5), & 4.5 < x \le 5.0 \\ 4(5.5-x), & 5.0 < x \le 5.5 \\ 0, & \text{otherwise} \end{cases}$

(a) Find the definite integral for $P(X \le 4.7)$: $\int_{4.5}^{4.7} 4(x-4.5)dx$

Evaluate this integral: $= [2(x^2 - 9x)]_{4.5}^{4.7}$

$= 2[4.7^2 - 9 \times 4.7 - (4.5^2 - 9 \times 4.5)]$

$= 0.08$

(b) Find the definite integral for $P(X \le 5.2)$: $\int_{4.5}^{5.0} 4(x-4.5)dx + \int_{5.0}^{5.2} 4(5.5-x)dx$

$$= [2(x^2 - 9x)]_{4.5}^{5} + [2(11x - x^2)]_{5}^{5.2}$$

$$= 2[25 - 45 - (4.5^2 - 40.5) + 57.2 - 5.2^2 - (55 - 25)]$$

$$= 0.82$$

If any of your integrals evaluate to more than 1 for any probability density function, then you can be sure your answer is incorrect. The entire area under the curve must be 1, so part of it cannot be greater than this.

Similarly, if any of your integrals give a negative value, then you can be sure this is incorrect.

The value of the integral represents probability, and probability cannot be less than 0 or greater than 1.

As for discrete probability distributions, common calculations with continuous probability density functions involve finding the mean, variance and standard deviation.

You should recall that for any discrete probability distribution, the expected value (mean) is found by adding the sum of the products of the values and individual probability values ($\Sigma x_i p_i$) and that the variance can be found from $E(X^2) - [E(X)]^2$. The integral is the continuous distribution equivalent of summation, Σ, so the following formulae should not surprise you.

The mean of a continuous probability density function is found using the following formula:

$$\mu = \int_{-\infty}^{\infty} x f(x)dx$$

The variance of the continuous probability density function is found using the following formula:

$$\sigma^2 = E\left(X^2\right) - \left[E(X)\right]^2$$

$$= \left(\int_{-\infty}^{\infty} x^2 f(x)dx\right) - \mu^2$$

As usual, the standard deviation is the square root of the variance.

Although the formulae say to find the integrals from $-\infty$ to ∞, in practice you calculate the integrals over the interval where the function is non-zero.

Example 20

For a particular Infant Welfare Centre the probability density function of the age of the children (x years) brought to the centre is given by:

$$f(x) = \begin{cases} \frac{3}{4}x(2-x), & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Find the following correct to three decimal places:

(a) mean

(b) standard deviation.

Solution

(a) Find the expression for $xf(x)$: $xf(x) = x \times \frac{3}{4}x(2-x)$

$$= \frac{3}{4}\left(2x^2 - x^3\right)$$

Find the mean μ by calculating the integral $\int_{-\infty}^{\infty} xf(x)\,dx$:

(Remember, you actually calculate the integral over the interval for which the function is non-zero.)

$$\begin{aligned}
\mu &= \frac{3}{4}\int_0^2 \left(2x^2 - x^3\right)dx \\
&= \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 \\
&= \frac{3}{4}\left(\frac{16}{3} - 4 - 0\right) \\
&= 1
\end{aligned}$$

(b) Find the value of the integral $\int_{-\infty}^{\infty} x^2 f(x)\,dx$:

(Remember, you actually calculate the integral over the interval for which the function is non-zero.)

$$\begin{aligned}
\int_{-\infty}^{\infty} x^2 f(x)\,dx &= \frac{3}{4}\int_0^2 x^2\left(2x - x^2\right)dx \\
&= \frac{3}{4}\int_0^2 \left(2x^3 - x^4\right)dx \\
&= \frac{3}{4}\left[\frac{x^4}{2} - \frac{x^5}{5}\right]_0^2 \\
&= \frac{3}{4}\left(8 - \frac{32}{5}\right) \\
&= \frac{6}{5}
\end{aligned}$$

From **(a)**, μ^2=1.

Find the value of μ^2 (you know the value of μ from **(a)**): $\mu^2 = 1$

Calculate the variance using $\sigma^2 = \left(\int_{-\infty}^{\infty} x^2 f(x)\,dx\right) - \mu^2$: $\sigma^2 = \frac{6}{5} - 1$

$$\begin{aligned}
&= 0.2 \\
\sigma &= \sqrt{0.2} \\
&= 0.447
\end{aligned}$$

You should check that your answers are plausible. For example, is the mean value within the interval for which the probability is non-negative? Also, as a general rule, the range of the distribution should be roughly five times the standard deviation. Knowing this will assist you when trying to sketch some graphs. For the example above, the mean is near the middle of the non-negative interval and five standard deviations is $5 \times 0.447 = 2.235$, compared to the actual range of 2.

Median and quartiles

You can also find the median of a continuous probability density function. As should be expected, the median m is the value such that $P(X < m) = 0.5$. This means you need to set up and solve an equation of the form:

$\int_0^a f(x)\,dx = 0.5$, assuming the non-zero part of the piece-wise function starts at 0.

To find the first quartile, solve the equation $\int_0^a f(x)\,dx = 0.25$, assuming that the non-zero part of the piece-wise function starts at 0.

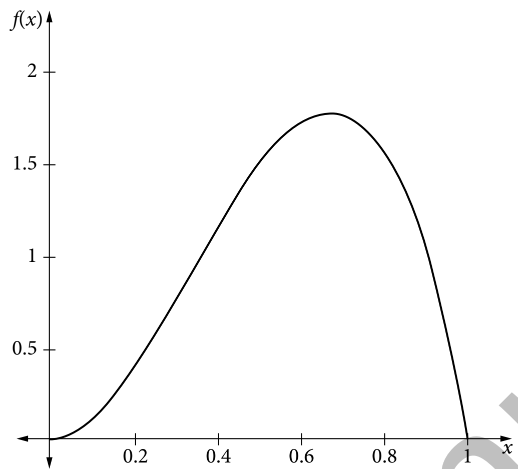
To find the third quartile, solve the equation $\int_0^a f(x)\,dx = 0.75$, assuming that the non-zero part of the piece-wise function starts at 0.

The lower limit in each of these integrals will be the start of the non-zero part of the piece-wise function.

Mode

The diagram shows the graph of the probability density function given by

$$f(x) = \begin{cases} 12x^2(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}.$$



The highest point of this graph is somewhere between $x = 0.6$ and $x = 0.8$. Hence the mode of this distribution is between 0.6 and 0.8 .

The highest point on this this graph occurs at its turning point, where $f'(x) = 0$.

Find $f'(x)$: $f(x) = 12\left(x^2 - x^3\right)$

$$f'(x) = 12\left(2x - 3x^2\right) = 12x(2-3x)$$
$$f'(x) = 0 : \quad 12x(2-3x) = 0$$
$$x = 0 \text{ or } \frac{2}{3}$$

Since $f(0) = 0$, then the mode is $\frac{2}{3}$.

If the probability density function is a linear function, then $f'(x) = 0$ for all values for which the function has been defined. In this case, the mode will be the largest value in the domain of the function.

Example 21

A particular continuous random variable X has the following probability density function:

$$f(x) = \begin{cases} \dfrac{x}{16}, & 0 \le x \le 4\sqrt{2} \\ 0, & \text{otherwise} \end{cases}$$

Find:

(a) the median
(b) the first quartile
(c) the mode.

Solution

(a) Write an equation that states what you are required to solve: $\int_0^a f(x)\,dx = 0.5$

Write an expression for $\int f(x)\,dx$: $\int f(x)\,dx = \int \frac{x}{16}\,dx$

Call the upper boundary a, and write the definite integral in the equation: $\int_0^a \frac{x}{16}\,dx = 0.5$

Find the primitive: $\left[\frac{x^2}{32}\right]_0^a = \frac{1}{2}$

Evaluate and solve the equation:

$$\frac{a^2}{32} = \frac{1}{2}$$

$$a^2 = 16$$

$a = 4$, taking the positive square root as $a > 0$.

The median is 4.

(b) First quartile:

$$\int_0^b f(x)\,dx = 0.25$$

$$\int_0^b \frac{x}{16}\,dx = 0.25$$

$$\left[\frac{x^2}{32}\right]_0^b = \frac{1}{4}$$

$$\frac{b^2}{32} - 0 = \frac{1}{4}$$

$$b^2 = 8$$

$$b = 2\sqrt{2} \text{ as } 0 \le b \le 4\sqrt{2}$$

The first quartile is $2\sqrt{2}$.

(c) $f(x) = \frac{x}{16}$ for $0 \le x \le 4\sqrt{2}$.

$f'(x) = \frac{1}{16}$ for all x in the domain so the largest value of $f(x)$ occurs at the largest value of x in the domain.
The mode is $4\sqrt{2}$.

Continuous uniform random variable

Consider the probability function for a uniform distribution given by $f(x)=\begin{cases}\dfrac{1}{b-a}, & \text{if } a\le x\le b\\ 0, & \text{otherwise}\end{cases}$.

Consider: $\displaystyle\int_a^b f(x)\,dx=\int_a^b \frac{dx}{b-a}$

$$\begin{aligned}&=\left[\frac{x}{b-a}\right]_a^b\\&=\frac{b}{b-a}-\frac{a}{b-a}\\&=\frac{b-a}{b-a}\\&=1\end{aligned}$$

This confirms that $f(x)$ is a probability function.

Mean

The expected value, or mean of this probability distribution function is given by:

$$\begin{aligned}E(X)=\mu=\int_a^b xf(x)\,dx&=\int_a^b \frac{x}{b-a}dx\\&=\left[\frac{x^2}{2(b-a)}\right]_a^b\\&=\frac{1}{2(b-a)}\left(b^2-a^2\right)\\&=\frac{a+b}{2}\end{aligned}$$

Variance

The variance of this probability distribution function is given by:

$$\begin{aligned}\text{Var}(X)=\sigma^2=E\left(X^2\right)-[E(X)]^2&=\int_a^b \frac{x^2}{b-a}dx-\left(\frac{a+b}{2}\right)^2\\&=\frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b-\frac{(a+b)^2}{4}\\&=\frac{1}{3(b-a)}\left(b^3-a^3\right)-\frac{(a+b)^2}{4}\\&=\frac{(b-a)\left(b^2+ab+a^2\right)}{3(b-a)}-\frac{a^2+2ab+b^2}{4}\\&=\frac{4\left(b^2+ab+a^2\right)-3\left(a^2+2ab+b^2\right)}{12}\\&=\frac{b^2-2ab+a^2}{12}\\&=\frac{(b-a)^2}{12}\end{aligned}$$

Probabilities associated with a continuous uniform distribution are given by:

$P(c\le X\le d)=\displaystyle\int_c^d \frac{dx}{b-a}=\frac{d-c}{b-a}$, where $a\le c<d\le b$.

Example 22

Trains from Newcastle to Sydney run every sixty minutes throughout the day.

(a) Find the probability density function for X.
(b) Find the expected waiting time and the standard deviation.
(c) Find the probability of waiting at least 15 minutes for the next train after arriving at Newcastle station.
(d) What is the probability that the wait is less than 15 minutes?

Solution

(a) $f(x) = \dfrac{1}{60-0} = \dfrac{1}{60}$

(b) $E(X) = \dfrac{0+60}{2} = 30$ minutes

$\sigma^2 = \dfrac{(60-0)^2}{12} = 300$, so $\sigma = 10\sqrt{3}$ minutes.

(c) $P(x \geq 15) = \dfrac{60-15}{60-0} = 0.75$

(d) $P(x < 15) = \dfrac{15-0}{60-0} = 0.25$

EXERCISE 7.5 CONTINUOUS PROBABILITY DISTRIBUTIONS

1 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{3}{13}\left(x^2+4\right), & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

find, correct to four decimal places:
(a) $P(X \leq 0.75)$ (b) $P(X \leq 0.9)$ (c) $P(X \geq 0.25)$ (d) $P(X \geq 0.65)$

2 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{15\,000(x-50)}{x^4}, & x \geq 50 \\ 0, & \text{otherwise} \end{cases}$$

find the exact values for the following:
(a) $P(X \leq 60)$ (b) $P(X \leq 75)$ (c) $P(X \geq 65)$ (d) $P(X \geq 99)$

3 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{3}{13}\left(x^2+4\right), & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

find, correct to four decimal places, the following values:
(a) the mean
(b) the standard deviation.

4 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{15\,000(x-50)}{x^4}, & x \geq 50 \\ 0, & \text{otherwise} \end{cases}$$

find the value of the mean.

5 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{3}{13}\left(x^2+4\right), & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

find the median, correct to four decimal places.

6 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{15\,000(x-50)}{x^4}, & x \geq 50 \\ 0, & \text{otherwise} \end{cases}$$

Find:

(a) the median, correct to the nearest whole number

(b) the mode, correct to the nearest whole number.

7 If $f(x) = \begin{cases} \dfrac{x}{45}+k, & 0 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases}$

defines a probability density function, then the value of k is:

A 0 B $-\frac{1}{10}$ C $\frac{1}{10}$ D $\frac{1}{2}$

8 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{2x}{25}, & 0 \leq x \leq 5 \\ 0, & \text{otherwise} \end{cases}$$

find, correct to four decimal places:

(a) $P(1.5 \leq X \leq 2.5)$

(b) $P(2 \leq X \leq 4.5)$

(c) $P(1.75 \leq X \leq 3.15)$

9 If $f(x) = \begin{cases} \dfrac{\pi}{12}\sin\left(\dfrac{\pi x}{6}\right), & 0 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases}$

then $P(X \leq 3)$ is equal to:

A 0.45 B 0.5 C 0.55 D 0.6

10 For the uniform continuous random variable with probability density function:

$$f(x) = \begin{cases} k, & 0 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases}$$

find the following values:

(a) k (b) $P(X \leq 3)$ (c) $P(X \leq 5)$

11 For the continuous random variable with probability density function:

$$f(x) = \begin{cases} \dfrac{\sin(x)}{2}, & 0 \le x \le \pi \\ 0, & \text{otherwise} \end{cases}$$

Find the exact value of:

(a) the median

(b) the first quartile

(c) the third quartile

(d) the mode.

12 In the javelin competition at a primary schools athletics carnival, it is found that the distance s metres that the javelin is thrown is a continuous random variable with probability density function:

$$f(s) = \begin{cases} \dfrac{1}{486}\left(81 - s^2\right), & 0 \le x \le 9 \\ 0, & \text{otherwise} \end{cases}$$

Find:

(a) the mean distance the javelin is thrown

(b) the standard deviation for the distance the javelin is thrown

(c) the median distance the javelin is thrown.

13 Let the duration, X minutes of a telephone conversation be represented by a continuous random variable with probability density function:

$$f(x) = \begin{cases} Ce^{-\frac{x}{10}}, & 0 \le x \le \infty \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the exact value of C.

(b) Find the following values, correct to four decimal places:

(i) $P(X \le 2.5)$ **(ii)** $P(X \ge 6.4)$ **(iii)** $P(2 \le X \le 7)$

(c) Find $E(X)$.

(d) Find $\text{Var}(X)$.

(e) What is $P(X \le E(X))$, correct to 4 decimal places?

14 A physical therapist uses a particular practical test to determine the reaction time of her patients. From experience, she has determined that the reaction time, t seconds, can be modelled by a continuous random variable with probability density function:

$$f(t) = \begin{cases} k\left(t - \dfrac{1}{4}\right)\left(\dfrac{5}{4} - t\right)^2, & \dfrac{1}{4} \le x \le \dfrac{5}{4} \\ 0, & \text{otherwise} \end{cases}$$

Unless otherwise stated, find the following answers correct to 3 decimal places.

(a) Find the exact value of k.

(b) Find the mean reaction time.

(c) Find the median reaction time.

(d) Find the proportion of patients who react in less than one second.

(e) Find the proportion of patients who react in a time between $\frac{2}{5}$ and $\frac{4}{5}$ of a second.

(f) Find the proportion of patients who take longer than $\frac{3}{4}$ of a second to react.

15 The local council needs to decide what to do about collecting recyclable materials. Each household has a 'recyclable materials' bin that is collected fortnightly. Each truck that collects these bins will contain an amount X of materials that are *not* recyclable. (X is measured in units of 100 kg.) It is known that the random variable X has a probability density function:

$$f(x) = \begin{cases} kx(4-x)^2, & 0 \le x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that $k = \frac{3}{64}$.

(b) Find the mean and standard deviation of X.

(c) Find the probability that a bin chosen at random has more than 3 units of non-recyclable material in it.

The council has two choices as to how to proceed with the collection of recyclable material. The first is as follows.

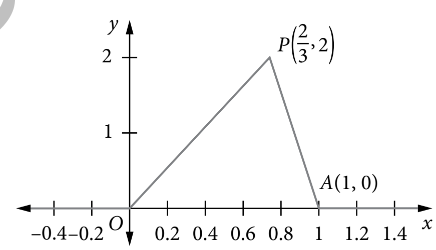
If the truck contains less than 3 units of non-recyclable material, then they will be able to sell the materials for \$500. However, if the quantity is more than this, then they will only receive \$300.

(d) Find the expected value of X for a truck full of material.

The second choice for the council is to sort the material before selling it. By doing this they will ensure that there are virtually no non-recyclable materials being sold. As a result, they will get \$650 per load. However, this means they cannot process as many trucks, so they only collect the material from households every 3 weeks.

(e) Is it financially viable for the council to follow this second approach?

16 Consider the function *f* whose graph is shown.



(a) Show that the function could be used to represent a probability density function. Explain your answer using mathematical reasoning.

(b) Specify the rule of *f*.

(c) Consider a continuous random variable, X, with probability density function *f*. Calculate $P\left(X < \frac{2}{3}\right)$ and $P\left(X > \frac{2}{3}\right)$ using the areas of the triangles.

(d) Consider a continuous random variable X with probability density function *f*. Calculate $P\left(X < \frac{2}{3}\right)$ and $P\left(X > \frac{2}{3}\right)$ using integration. Comment on the results from parts **(c)** and **(d)**.

(e) Calculate $P\left(\frac{1}{2} < X < \frac{5}{6}\right)$, using integration.

17 The probability density function, $f(x)$, is given by $f(x) = \begin{cases} \frac{1}{12}\left(8x - x^3\right), & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$.

The median, m, of this function satisfies the equation:

A $-m^4 + 4m^2 - 6 = 0$

B $m^4 - 16m^2 = 0$

C $m^4 - 16m^2 + 24 = 0.5$

D $m^4 - 16m^2 + 24 = 0$

18 The time in minutes taken for Amelie to score her first point in a tennis match can be represented by a continuous random variable X with probability density function given by:

$$f(x) = \begin{cases} ke^x, & \text{for } 0 \le x \le 4 \\ ke, & \text{for } 4 < x \le 10 \\ 0, & \text{for all other } x \end{cases}$$

(a) Find the exact value of the constant k.

(b) The median time taken for Amelie to score her first point is less than 4 minutes. Find this median time taken for Amelie to score her first point, giving your answer correct to the nearest second.

19 A continuous random variable, X, has a probability density function $f(x)$ given by:

$$f(x)=\begin{cases}12x^2(1-x), & \text{for } 0\le x\le 1\\ 0, & \text{for all other values of } x\end{cases}.$$

(a) Find the mode of X.

(b) Find the cumulative distribution function for the given probability density function.

20 A continuous random variable X has the following probability density function:

$$f(x)=\begin{cases}3^{-x}\ln 3, & \text{for } x\ge 0\\ 0, & \text{for all other values of } x\end{cases}.$$

(a) Find the cumulative distribution function $F(x)$.

(b) Find the exact value of the third quartile.

21 The queuing time, X minutes, of a teacher waiting on the phone with the Department of Education has a probability density function $f(x)=\begin{cases}\frac{3}{32}x(k-x), & 0\le x\le k\text{, where } k \text{ is a constant}\\ 0, & \text{otherwise}\end{cases}.$

(a) Show that $k = 4$.

(b) Find the cumulative distribution function, $F(x)$.

(c) Find the mode of the probability density function.

(d) Find the probability that the phone will be answered within the first minute.

22 In a hotel with an elevator that will take you to your required floor, it is known that from the time that you push the button to call the elevator, it takes between 10 and 40 seconds for you to arrive at your floor.

(a) Find the probability density function.

(b) Calculate the expected value, the variance and the standard deviation.

(c) Find the probability that it takes more than 30 seconds to arrive at your floor.

23 At a bus stop on a particular day, a bus arrives every 10 minutes. The time spent waiting for the next bus (X minutes) is a continuous random variable with the probability density function:

$$f(x)=\begin{cases}k, & x\in[0,10]\\ 0, & x\in(-\infty,0)\cup(10,\infty)\end{cases}.$$

Kota goes to the bus stop twice, once in the morning and again in the afternoon.

(a) What is the probability that his morning bus will require a wait time of more than 7 minutes?

(b) What is the probability that at least one of his buses will require a wait time of more than 7 minutes?

24 A continuous random variable, X, has a probability density function $f(x) = 2x$ for $0 \le x \le 1$. Find the cumulative distribution function, $F(x)$ and sketch its graph.

25 The cumulative distribution function, $F(x)$, is given by $F(x) = 1 - e^{-x}$, for $x \ge 0$.

Find $P(1 \le x \le 2)$.

26 A pdf is defined as $f(x)=\begin{cases}0, & x<1\\ x\ln x, & 1\le x\le 2.2295\\ 0, & x>2.2295\end{cases}.$

(a) Use graphing software to verify that $f(x)$ is a pdf, correct to 3 decimal places.

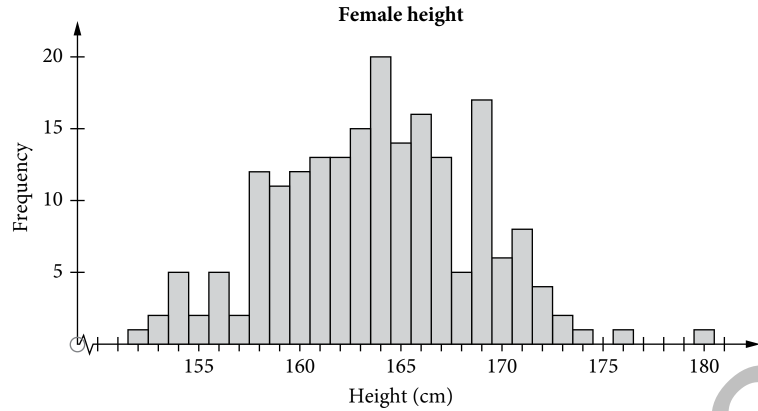
(b) Write down the integral for $F(x)$.

(c) Use graphing software to find $P(1\le X < 2)$, correct to 3 decimal places.

27 For the pdf in question 25, find the mean and the standard deviation of the distribution.

7.6 THE NORMAL DISTRIBUTION

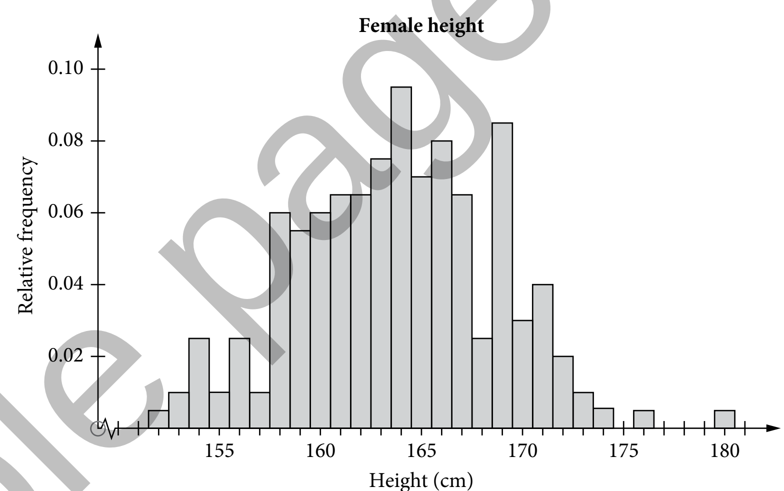
The following histogram shows the height of 200 randomly selected Australian women, measured in cm accurate to one decimal place.



Can you find a function that will model the data, just like Dirk did with the train data in the previous module?

The total area of the rectangles in the histogram is 200, the size of the sample, so the first thing that needs to be done is to make the sum of the areas of the rectangles equal to 1.

To do this, divide the vertical scale values by 200. As the bin width (the width of the columns in the histogram) is set to 1, this means that the area of each bin represents the probability of a woman having a height in that 1 cm range. More importantly, the total area of the bins is now 1, so the situation is like a probability density function.

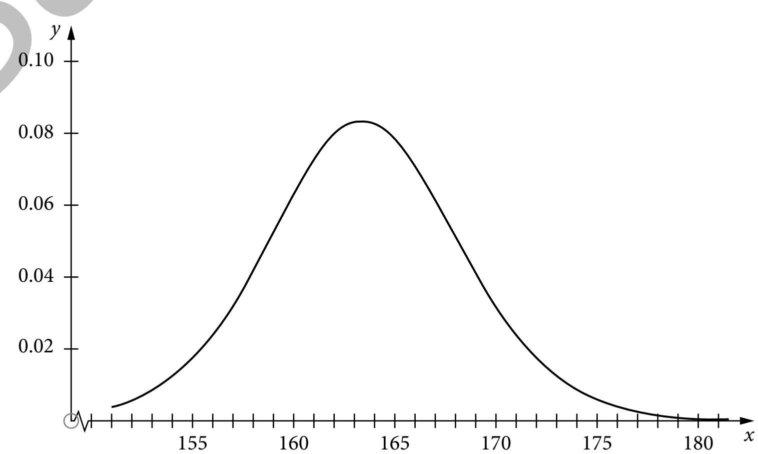


From the raw data, the mean and standard deviation can be calculated:

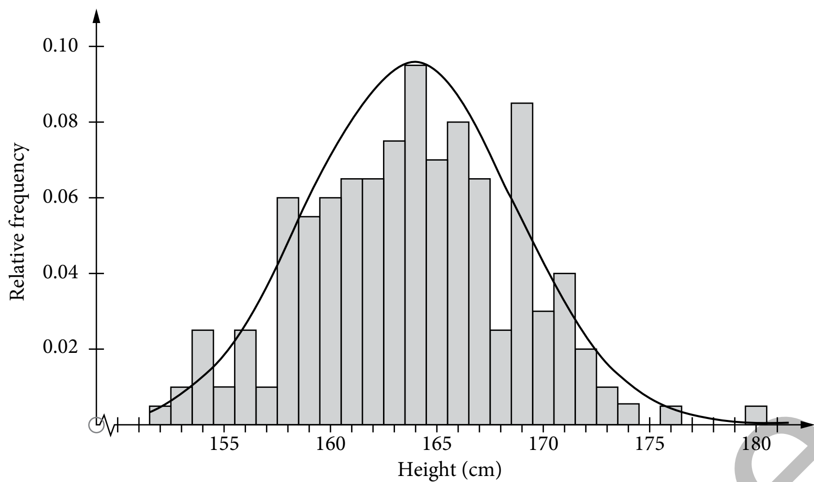
Mean = 163.8085

Standard deviation = 4.858 855

What sort of function might be suitable as a model for this data? The shape appears similar to a parabola, but the ends seem to drop away too quickly for that to be a successful model. Perhaps a shape such as the one shown below might be suitable.

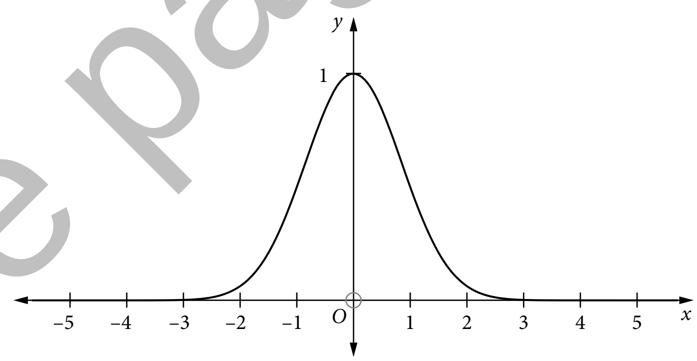


If the two images are put together it can be seen that this function certainly has some potential as a model for the distribution of heights.



This matches the general shape quite well, although as with all models, it is not perfect. What function might produce this graph? The way the graph falls off suggests an exponential function of some kind, but unlike a typical exponential it falls off on both sides. Instead, try $f(x) = 2^{-x^2}$.

This shape looks like a good match, but is it a probability density function? The values of the function are non-negative, so it satisfies that condition for a probability density function. The graph almost touches the x-axis at −3 and 3, so the area could be approximated using a triangle with a base of 6 and a height of 1. That gives an area of 3, which is too large a value. The graph of the function will need to be adjusted, possibly by using a vertical dilation, for example $f(x) = k \times 2^{-x^2}$ for some suitable value of k. This can be determined by using a definite integral to obtain a more accurate approximation.



Note that while the domain of the function is $\mathbb{R}$ (real numbers), the value of the function is increasingly close to 0 to the left and right of the graph, for example, $x < -3$ and $x > 3$.

Using a suitable definite integral, find an approximation for the area beneath the graph and hence find the value of $\int_{-3}^{3} 2^{-x^2}\,dx$. The integration of this function is not covered in this course, but if you have access to graphing software you can use that to evaluate this integral. The integral could also be approximated using the trapezoidal rule.

MAKING CONNECTIONS

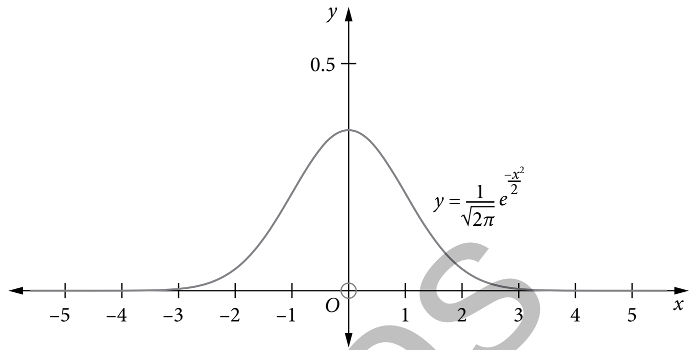
Definite integral of an approximate distribution curve

Use technology to find the value of the definite integral $\int_{-3}^{3} 2^{-x^2}\,dx$.

Hence $\int_{-3}^{3} 2^{-x^2}\,dx = 2.128\,0567\,6757 \approx 2.128\,06$.

This shows that using $k = 2.128\,06$, and hence $f(x) = \dfrac{1}{2.128\,06}2^{-x^2}$, will produce a probability density function with the desired shape. However, it will also need to be translated and dilated to match the data presented at the beginning.

Mathematicians have found that a base of e rather than 2 works well, and in this case instead of $\dfrac{1}{2.128\,06}$ the value $\dfrac{1}{\sqrt{2\pi}}$ should be used. This gives the function with domain $\mathbb{R}$ and rule $f(x) = \dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$, defined using constants you are already familiar with. This graph is shown at right and has the same general shape.



Although it looks like the graph meets the x-axis, it is asymptotic to the x-axis in both directions. That is, as $x \to \infty$, $f(x) \to 0$ and as $x \to -\infty$, $f(x) \to 0$ where $f(x) > 0$ for all $x \in \mathbb{R}$.

You can evaluate $\int_{-\infty}^{\infty} \dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx$ using graphing software.

MAKING CONNECTIONS

Definite integral of a normal distribution curve

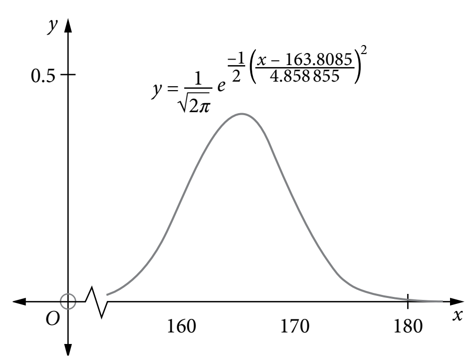
Use technology to find the value of the definite integral $\int_{-\infty}^{\infty} \dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx$.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx = \int_{-100}^{100} \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx = 1$$

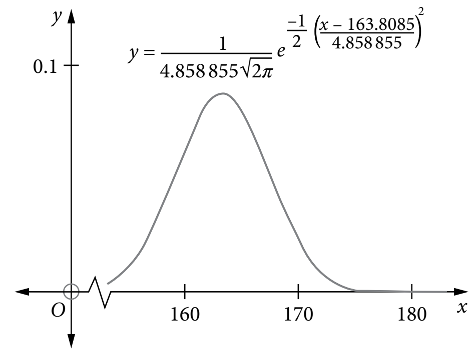
(The limits on the integral have been changed as ∞ cannot be used by all software, but [−100,100] is a close enough approximation to infinity; −50 to 50 also gives a value of 1.)

You now have a probability density function, but you still need to translate and dilate this to match your data. From your previous work of transformations of graphs of functions, you should recall that a horizontal translation by h units will transform the graph of $y = f(x)$ onto the graph of $y = f(x - h)$. You should be able to see that the required horizontal translation is equal to the mean; however, a horizontal dilation will also be required. What value might be related to the horizontal dilation? The effect of the horizontal dilation is to spread out the values. The standard deviation is a measure of the spread of the data, so that is a value worth trying.

In combination, a horizontal translation by h units combined with a horizontal dilation by b units will transform the graph of $y = f(x)$ onto the graph of $y = f\left(\dfrac{x-h}{b}\right)$. In this case, the value of h is given by the mean of your data, 163.8085, and the value of b is given by the standard deviation, 4.858 855. The function obtained is $f(x) = \dfrac{1}{\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-163.8085}{4.858\,855}\right)^2}$.



The area under the curve is given by:



$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-163.8085}{4.858\,855}\right)^2} dx = \int_{133}^{193} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-163.8085}{4.858\,855}\right)^2} dx$$

$$= 4.858\,855$$

The area should be related to the total probability (1), but it is too big by a factor of the standard deviation. Divide by this to obtain a vertical dilation which reduces the area to one, so that

$f(x) = \frac{1}{4.858\,855\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-163.8085}{4.858\,855}\right)^2}$ is the final result. This is a probability density function with a mean of 163.8085 and a standard deviation of 4.858 855.

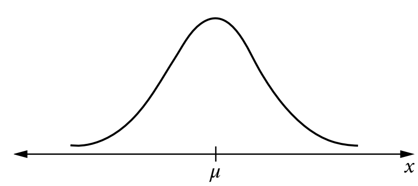
You would not want to go through this long process every time you have a data set to analyse. However, you can generalise the rule for this function to the form: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

The distribution represented by the probability density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ over domain $\mathbb{R}$, where μ is the mean of the distribution and σ is the standard deviation of the distribution, is known as the **normal distribution**. The special case where $\mu = 0$ and $\sigma = 1$ is called the standard normal distribution, as all other normal distributions can be obtained from it by translation and dilation.

Note that this distribution is defined over $\mathbb{R}$, that is from $-\infty$, to ∞. In practical situations which can be modelled by a normal distribution, there will usually be an interval $[a, b]$ for which the value of the probability density function can 'reasonably' be interpreted in that situation.

The normal distribution is the most useful of all the probability distributions for continuous random variables. As has been seen, its graph is characterised by a symmetrical 'bell' shape. This shape can be used as a model for the data collected from many naturally occurring variables, such as the height of a population, the intelligence quotient (IQ) of a population, or the distribution of errors in a production process. It is also frequently used as an approximation for the sums and averages of samples taken from fixed populations and can be essential when making inferences about populations from samples.

The symmetrical nature of the graph means that the mean, median and mode coincide for the normal distribution and this value becomes the axis of symmetry. It should also be noted that, this graph extends infinitely (in theory) in both positive and negative directions, but it is always located above the x-axis. Thus, the normally distributed variable can assume any value. The graph of the normal distribution is often referred to as a bell-shaped curve.

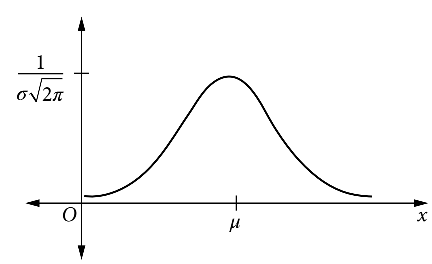


Looking at the graph it can be seen that the maximum value occurs when $x = \mu$. If you substitute this into the equation you get $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^0$. This simplifies to $\frac{1}{\sigma\sqrt{2\pi}}$, as shown in the graph on the next page.

(This maximum value is useful when you are setting the window size for your graphing software.)

You need to check that the conditions for a probability density function have been met. By inspection, $f(x) \geq 0$ for all values of x and $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$.

However, the proof of this is beyond the scope of this course.



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

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

$$= -\left(\frac{x-\mu}{\sigma^3\sqrt{2\pi}}\right) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f'(x) = 0$ when $x = \mu$.

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

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

When $x = \mu$, $f''(\mu) = -\left[\frac{e^{-\frac{1}{2}\left(\frac{0}{\sigma}\right)^2}}{\sigma^2\sqrt{2\pi}}\right]\left(1 - \frac{(0)^2}{\sigma^2}\right) = -\frac{e^0}{\sigma^2\sqrt{2\pi}} < 0$

Hence the maximum value of the function occurs when $x = \mu$.

If $f''(x) = 0$, then $-\left[\frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma^2\sqrt{2\pi}}\right]\left(1 - \frac{(x-\mu)^2}{\sigma^2}\right) = 0$

Since $e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} > 0$, then $1 - \frac{(x-\mu)^2}{\sigma^2} = 0$

Hence $(x-u)^2 = \sigma^2$

So, $x = \mu \pm \sigma$

Thus, the maximum value of the function occurs at $x = \mu$ and the points of inflection occur at $x = \mu \pm \sigma$, that is at one standard deviation from the mean.

MAKING CONNECTIONS

Transformations of the normal distribution

Move the sliders to explore the effects of changing μ and σ on the graph of a normal distribution.

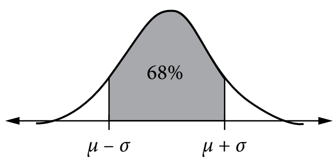
These investigations show that:

- increasing μ shifts the graph to the right, i.e. μ affects the location of the curve.
- increasing σ makes the curve flatter and wider, but does not alter its axis of symmetry. If the size of the standard deviation is increased, then the range $\mu \pm 3\sigma$ also increases, thus demonstrating the increased width required for the graph.

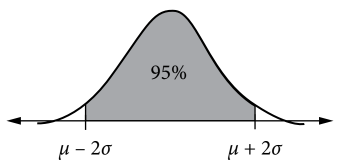
The common feature for all of the graphs is that the area under the curve is always 1.

Remember that the total area under the curve is 1 (a condition of being a probability density function). Probability is related to the proportion of the total area which is being considered in a particular example.

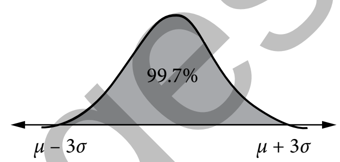
One feature of the normal distribution that is frequently used, especially in situations where technology cannot be used, is related to the percentage of observations that you would expect to be within a certain number of standard deviations of the mean.



Approximately 68% of the observations will lie within one standard deviation of the mean.



Approximately 95% of the observations will lie within two standard deviations of the mean. It can be said that it is *very likely* that any particular observation will lie within this range.



Approximately 99.7% of the observations will lie within three standard deviations of the mean. It can be said that it is *almost certain* that any particular observation will lie within this range.

Empirical rule

For normal distributions:

- approximately 68% of data lie within one standard deviation of the mean
- approximately 95% of data lie within two standard deviations of the mean
- approximately 99.7% of data lie within three standard deviations of the mean.

For instance, if you know that the height of a particular population is normally distributed with a mean of 170 cm and a standard deviation of 6 cm, you would expect observations roughly as follows:

- 68% in the range $\mu \pm \sigma$, i.e. 170 ± 6, which gives the range 164 to 176 cm
- 95% in the range $\mu \pm 2\sigma$, i.e. 170 ± 12, which gives the range 158 to 182 cm
- 99.7% in the range $\mu \pm 3\sigma$, i.e. 170 ± 18, which gives the range 152 to 188 cm.

You can use the 68%, 95% and 99.7% rules to help estimate and check values in many situations.

There is a short way of writing that a random variable has a normal distribution:

A continuous random variable that has a normal distribution with mean μ and variance σ^2 is written as: $X \sim N(\mu, \sigma^2)$.

Example 23

The time taken in seconds, X, for all competitors to finish a 100 m race at the school athletics carnival was found to follow a normal distribution where $X \sim N(15, 4)$. Find the following values:

(a) the time range in which you would expect to find the middle 68%

(b) the percentage of students you would expect to take more than 19 seconds.

Solution

(a) Write the mean μ and the standard deviation σ: $\mu = 15$, $\sigma^2 = 4$, $\sigma = \sqrt{4} = 2$
The middle 68% describes the values within one standard deviation of the mean, i.e. $\mu \pm \sigma$:
15 ± 2 gives the range 13 to 17 seconds.
You would expect the middle 68% of participants to take between 13 and 17 seconds to complete the race.

(b) Identify the value in terms of the mean and the standard deviation: $19 = \mu + 2\sigma$
(You may need to use a guess-and-check process.)
Use the symmetry of the curve to answer the question.
(In this case you are using only one tail, so you need to halve the percentage.)
You would expect 5% outside the range $\mu \pm 2\sigma$, so you would expect 2.5% of competitors to take more than 19 seconds.

EXPLORE FURTHER

Calculating probabilities using the normal distribution

Use technology to explore normal distribution probabilities within 1, 2 and 3 standard deviations from the mean.

Beware:

The formal description of the normal distribution $N(\mu, \sigma^2)$ uses the variance σ^2 as a parameter. However, worded questions often give the standard deviation σ. Make sure you read the questions carefully. When using technology you will also usually need to use the standard deviation and not the variance.

EXERCISE 7.6 THE NORMAL DISTRIBUTION

1 The time taken for all competitors to finish the 50 m freestyle at the school swimming carnival, X seconds, was found to follow a normal distribution where $X \sim N(45, 9)$. Find the following values.

- **(a)** The time range in which you would expect to find the middle 95% of results.
- **(b)** The percentage of students you would expect to take more than 48 seconds.
- **(c)** The percentage of students you would expect to take less than 36 seconds.

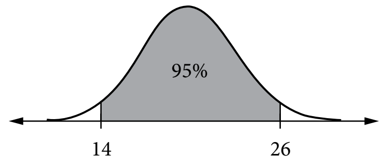
2 $X \sim N(10, 4)$.

- **(a)** What is the range of x values in which you would expect to find the middle 68%?
- **(b)** What is the range of x values in which you would expect to find the middle 95%?
- **(c)** What is the range of x values in which you would expect to find the middle 99.7%?

3 The height X cm of a population is known to be distributed as $X \sim N(170, 81)$.

- **(a)** What is the percentage of the population expected to be found in the range 152–188 cm?
- **(b)** What is the percentage of the population expected to be taller than 197 cm?
- **(c)** What is the percentage of the population expected to be shorter that 170 cm?
- **(d)** What is the percentage of the population expected to be found in the range 161–197 cm?

4 The graph represents a continuous random variable which has a normal distribution.

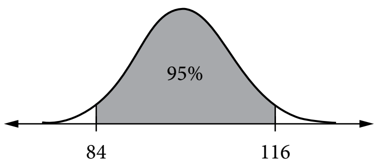


The distribution is best represented by:

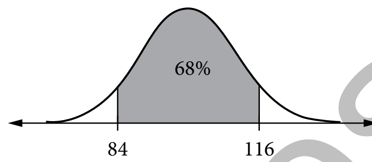
A $N(14, 26)$ **B** $N(26, 12)$ **C** $N(20, 9)$ **D** $N(20, 3)$

5 Given X is normally distributed with a mean of 100 and a variance of 16, which of the following graphs correctly represents the distribution?

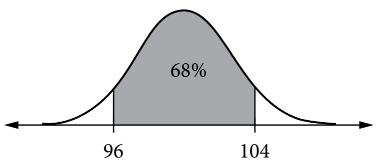
A



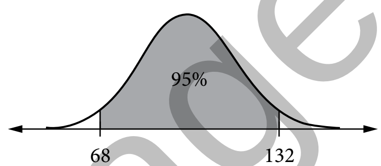
C



B



D



6 The marks, X, obtained by students in an examination were normally distributed with a mean of 85 and a standard deviation of 4. If the top 2.5% of students received a prize, find the minimum whole number score possible to receive a prize.

7 The mass, M, grams of a batch of commemorative coins is such that $M \sim N(50, 9)$. Each coin is weighed before packaging and will be rejected if its mass is less than 47 g. What is the percentage of coins expected to be rejected?

8 Packets of 'Greatstart' breakfast cereal are labelled as having a mass of 500 g. However, the machine that fills the packets actually follows a normal distribution with a mean of 510 g and a standard deviation of 5 g.

What percentage of packets, correct to two decimal places, will have a mass less than 500 g?

9 'Statzone' potato chips are packed by two different machines. Machine A fills the packets following a normal distribution with a mean of 100 g and a standard deviation of 3 g. Machine B fills the packets following the normal distribution $N(104, 16)$.

(a) **(i)** Between what values would you expect to find the middle 68% of packets for Machine A?
(ii) Between what values would you expect to find the middle 68% of packets for Machine B?
(b) **(i)** Between what values would you expect to find the middle 95% of packets for Machine A?
(ii) Between what values would you expect to find the middle 95% of packets for Machine B?
(c) **(i)** Between what values would you expect to find the middle 99.7% of packets for Machine A?
(ii) Between what values would you expect to find the middle 99.7% of packets for Machine B?
(d) If you bought a packet of the chips, which machine would you prefer to have packed it? Explain your answer.

10 Seventy students enrolled in a university course had their height measured in metres, correct to 2 decimal places where necessary. Their heights were as follows:

1.54	1.57	1.63	1.65	1.68	1.69	1.7	1.7	1.7	1.7
1.7	1.7	1.7	1.7	1.7	1.7	1.72	1.72	1.73	1.73
1.73	1.73	1.73	1.73	1.75	1.75	1.75	1.75	1.75	1.75
1.76	1.76	1.76	1.76	1.77	1.78	1.78	1.78	1.78	1.78
1.78	1.78	1.78	1.78	1.78	1.78	1.8	1.8	1.8	1.8
1.8	1.8	1.8	1.82	1.83	1.83	1.83	1.83	1.83	1.84
1.85	1.85	1.85	1.85	1.85	1.85	1.85	1.88	1.93	1.95

(a) Group the results using 1.5–<1.55, 1.55–<1.6, … as the intervals and draw a histogram to represent the data.
(b) Comment on the shape of the histogram and whether it would be reasonable to suggest that the sample follows a normal distribution.

(c) Find the mean and standard deviation, correct to three decimal places, of the original ungrouped data.
(d) What percentage, correct to one decimal place, of the results fall within:
(i) one standard deviation of the mean
(ii) two standard deviations of the mean
(iii) three standard deviations of the mean?
(e) Do you think it would be reasonable to say that the sample does follow a normal distribution?

11 Ivana sells freshly cut Christmas trees. She has 50 trees on her property ready to be cut and delivered. Prices depend on the height of the tree. She has recorded the heights of the trees, in metres, as shown.

1.61	1.66	1.68	1.68	1.71	1.71	1.71	1.72	1.73	1.73
1.73	1.76	1.76	1.77	1.77	1.77	1.77	1.79	1.79	1.79
1.81	1.81	1.81	1.81	1.82	1.83	1.83	1.83	1.83	1.84
1.86	1.86	1.86	1.86	1.86	1.86	1.87	1.87	1.88	1.91
1.91	1.91	1.93	1.93	1.94	1.94	1.96	1.96	1.96	2.01

(a) Calculate the average height of the Christmas trees.
(b) Display this data in a frequency table using class intervals of 0.05, using 1.60–1.65 as the first interval.
(c) Display this data on a histogram and comment on its shape.
(d) Calculate the standard deviation for this set of data, correct to four decimal places.

12 Eliza runs a business selling packets of stickers. The number of stickers per packet is normally distributed with a mean of 200 and a standard deviation of 2. To balance customer satisfaction and profits, Eliza only sells packets that contain between 198 and 204 stickers.
(a) If Eliza produces 400 packets of stickers per month, how many of these can she sell?
(b) Given that a packet is unable to be sold, what is the probability that it contains fewer stickers than average? Give your answer to the nearest per cent.

13 Packets of coffee beans are labelled with a net weight of 300 g. It is found that the weight of a packet can be modelled by a normal distribution with mean 306 g and standard deviation 3 g.
(a) Use the empirical rule to determine the probability that the weight of one packet of coffee beans is less than the advertised weight of 300 g.
(b) In a shipment of 40 boxes, each with 100 packets of coffee beans, how many packets would be expected to be underweight?
(c) Manufacturers aim to ensure that the expected number of underweight packets in the shipment will be less than 20. The machine is adjusted to give a mean weight of 309 g with the standard deviation of 3 g remaining. Will they meet their target? Justify your answer.

14 (a) Using graphing software draw the graph of $f(x) = e^{-x^2}$.
(b) Use software to evaluate $\int_{-4}^{4} e^{-x^2}\,dx$.

15 (a) Using graphing software, draw the graph of $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, for:
(i) $\mu = 10, \sigma = 3$
(ii) $\mu = 0, \sigma = 1$
(b) Use the integration tool in the software to evaluate $\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$ in each case in part **(a)**.

16 Use the trapezoidal rule with six sub-intervals to find the approximate value of the following:
(a) $\int_{1}^{19} \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-10)^2}{18}}\,dx$ (b) $\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$
Hint: Consider the symmetry of the function when constructing your table.
(c) Compare your answers to those obtained in question **13(b)**.

7.7 THE STANDARD NORMAL DISTRIBUTION

Although it is usually easy to use technology to find values associated with any normal distribution, it is often useful, especially when comparing distributions, to use what is called the **standard normal distribution**. This is a normal distribution that has a mean of 0, a variance of 1 and probability density function $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ with domain $\mathbb{R}$. Due to its importance this has a special letter, Z, reserved for the random variable of the standard normal distribution.

The standard normal distribution is such that $Z \sim N(0, 1)$, i.e. it has a mean of 0 and a variance of 1.

In the previous section you saw that the probability density function for $X \sim N(\mu, \sigma^2)$ can be obtained from that of $Z \sim N(0, 1)$ by the transformation $x = \sigma z + \mu$. Thus, to transform a normal distribution into the standard normal distribution the transformation $z = \frac{x - \mu}{\sigma}$ is applied. That is, the mean μ is subtracted from the observed value, X, and the result is divided by the standard deviation σ. As stated above, the resulting standard normal distribution is usually referred to by the letter Z.

If $X \sim N(\mu, \sigma^2)$, then $\frac{x - \mu}{\sigma} = z, Z \sim N(0, 1)$.

Consider the function $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ from the standard normal distribution.

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

$$= \frac{-x}{\sqrt{2\pi}} \times e^{\frac{-x^2}{2}}$$

$f'(x) = 0$ when $x = 0$, the mean of the standard normal distribution.

$$f''(x) = \frac{-1}{\sqrt{2\pi}} \left[1 \times e^{\frac{-x^2}{2}} + x\left(-xe^{\frac{-x^2}{2}}\right)\right]$$

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

$$= \frac{-e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}(1 - x)(1 + x)$$

When $x = 0$, $f''(0) = \frac{-e^0}{\sqrt{2\pi}} < 0$.

Hence the maximum value of the function occurs when $x = 0$.

If $f''(x) = 0$, then $\frac{-e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}(1 - x)(1 + x) = 0$.

Hence $x = \pm 1$.

Thus, the maximum value of the function occurs at $x = 0$, the mean, and the points of inflection occur at $x = \pm 1$, that is at one standard deviation from the mean.

Before digital technology was easily available, tables of values were used to calculate probabilities associated with normal distributions. This was one of the reasons it was so important to be able to change a distribution to the standard normal: it was the only distribution for which tables of values were easily available. In subjects such as this it is now used mainly for comparisons.

However, these standard z values (also called z-scores) have some important applications in science and elsewhere, such as in the field of paediatric health. Children grow at different rates, so it is difficult to use the more standard statistical tests when assessing particular children for serious health defects or conditions. For these applications to be useful, large statistical samples are needed for children of various ages, heights and weights. Then, comparisons can be made for a particular child against other children of the same age/height/weight, or some combination of these factors. For example, it may be suspected that a child has an inappropriately large dilation of a ventricle in their heart. If the dilation increases over time, this may not be a problem, because the dilation would always be expected to increase as the child grows older. However, if the z value increases over time, this is evidence that there may be a problem.

Calculating a *z* value

Example 24

X is a random variable following a normal distribution with mean 10 and variance 9 (i.e. $X \sim N(10, 9)$).

Find the z value that would be used to represent an x value of 14.

Solution

Identify σ and μ: $\mu = 10$, $\sigma = \sqrt{9} = 3$

Use the formula $z = \dfrac{x-\mu}{\sigma}$ to convert the x value to the equivalent z value: $z = \dfrac{14-10}{3}$

$$= \frac{4}{3}$$

$$\approx 1.33$$

This means that the observed value of X, 14, is 1.33 standard deviations greater than the mean.

Comparing *z* values

Example 25

Juan has been applying for scholarships. On one particular test he obtained 45 on a test that followed the distribution $X \sim N(40, 4)$ and on another he obtained 85 on a test that followed the distribution $Y \sim N(75, 25)$. On which test did Juan do better?

Solution

Find the z value for the first test: $\dfrac{X-\mu}{\sigma} = \dfrac{45-40}{2}$

$$= \frac{5}{2} = 2.5$$

Find the z value for the second test: $\dfrac{Y-\mu}{\sigma} = \dfrac{85-75}{5}$

$$= \frac{10}{5}$$

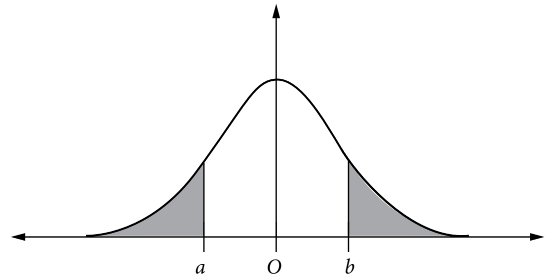
$$= 2$$

Juan did better on the first test, as his z value was further to the right.

The symmetry of the normal distribution helps with some calculations. This is most easily seen using the standard normal curve, where $\mu = 0$ and $\sigma^2 = 1$ (and therefore $\sigma = 1$).

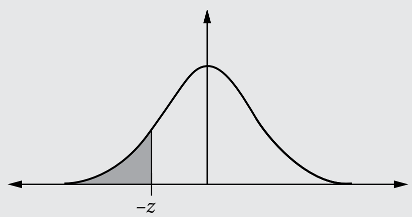
In the diagram the two shaded areas are equal. For this to be true, the distance between a and 0 (μ) must be the same as the distance between 0 and b. So $a = -b$.

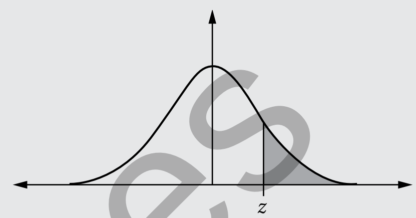
This result is useful as it shows that $P(X < a) = P(X > b)$. Also worth noting from this diagram is $P(X > b) = 1 - P(X < b)$ as the total probability is 1.



$P(Z < -z) = P(Z > z) = 1 - P(Z < z)$

A negative z value indicates that the observed value is less than the mean.



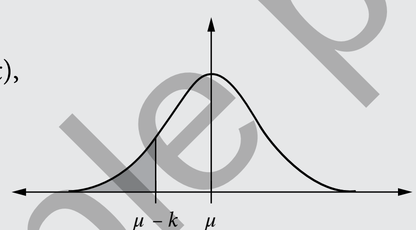


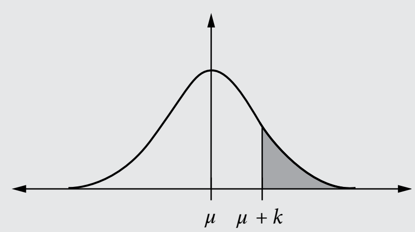
This result can be extended to any variable X that follows a normal distribution.

Assume $\mu = 20$. The value 5 units to the left of the mean is 15 and the value 5 units to the right of the mean is 25. This means, in this case, that $P(X < 15) = P(X > 25)$. This can also be expressed as $P(X < 15) = 1 - P(X < 25)$.

Expressing this result in general terms, for a variable X that follows a normal distribution with a mean of μ:

$P(X < \mu - k) = P(X > \mu + k)$ which leads to $P(X < \mu - k) = 1 - P(X < \mu + k)$, where k is the distance from the mean.





EXPLORE FURTHER

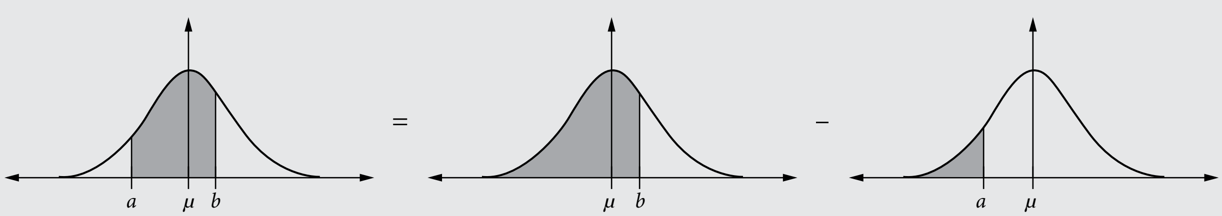
Symmetry of the normal distribution

Use technology to explore the symmetry of the normal distribution graph.

This result is useful when dealing with technology-free problems where the symmetry of the distribution is the only real information available to you.

For example, if you needed to find $P(10 < X < 15)$ where $\mu = 12$, you could find $P(X < 10)$ and subtract this from $P(X < 15)$.

Using general terms, to find $P(a < X < b)$:



Using the symmetry of the normal distribution

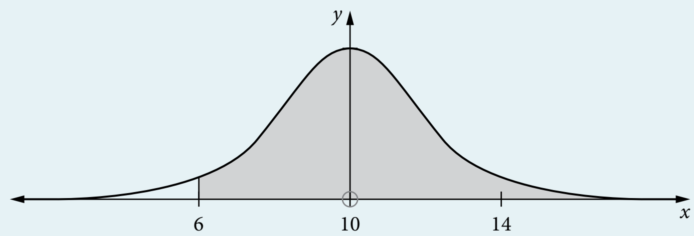
Example 26

A normal distribution graph is shown.

If $P(X > 14) = 0.22$ then $P(X > 6)$ is equal to:

A 0.22 **B** 0.44

C 0.56 **D** 0.78



Solution

Find $P(X > 6)$.

Identify the relevant symmetry aspects of the diagram in terms of probabilities: $P(X > 14) = P(X < 6)$

Write an expression using the required probability: $P(X > 6) = 1 - P(X < 6)$

Substitute known values and calculate the numerical answer: $P(X > 6) = 1 - 0.22 = 0.78$

Hence the answer is **D**.

Empirical rule

The results obtained earlier apply for the standard normal distribution:

For normally distributed random variables,
- approximately 68% of data will have z-scores between -1 and 1
- approximately 95% of data will have z-scores between -2 and 2
- approximately 99.7% of data will have z-scores between -3 and 3.

EXERCISE 7.7 THE STANDARD NORMAL DISTRIBUTION

1 If $X \sim N(15, 9)$, find the exact z values corresponding to the following x values:

(a) 18 **(b)** 21 **(c)** 22 **(d)** 16

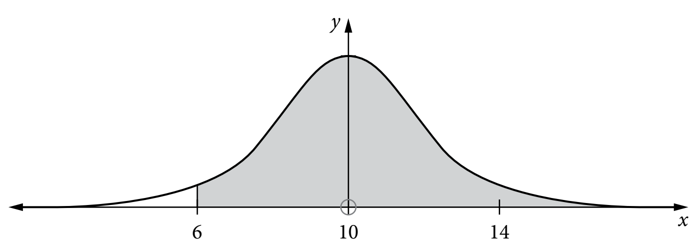
2 Felipe has been applying for scholarships. On one particular test that followed the distribution $X \sim N(40, 4)$, he obtained a 42; and on another test that followed the distribution $X \sim N(75, 25)$, he obtained an 82.

(a) Find the z value for the first test.
(b) Find the z value for the second test.
(c) On which test did Felipe do better?

3 A normal distribution graph is shown.

If $P(X > 14) = 0.35$, then $P(X > 6)$ is equal to:

A 0.35 **B** 0.55
C 0.65 **D** 0.7



4 For a particular normal distribution you know that $P(X < a) = 0.214$ and $P(X < b) = 0.496$, where $a < b$.

Find the following probabilities.

(a) $P(X > a)$ **(b)** $P(X > b)$ **(c)** $P(a < X < b)$

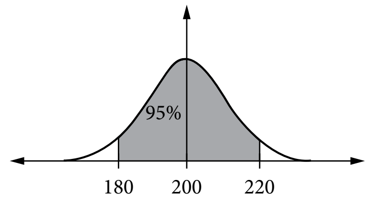
5 For $Z \sim N(0, 1)$ it is known that $P(Z < 0.85) = 0.8023$. Find:

(a) $P(Z > 0.85)$ **(b)** $P(Z < -0.85)$ **(c)** $P(Z > -0.85)$ **(d)** $P(-0.85 < Z < 0.85)$

6 For $Z \sim N(0, 1)$ it is known that $P(-1.5 < Z < 1.5) = 0.8864$. Find:

(a) $P(Z < -1.5)$ **(b)** $P(Z > -1.5)$ **(c)** $P(Z < 1.5)$

7 Given the diagram below for X which follows a normal distribution, which of the following expressions best represents the shaded area? (Z represents the standard normal distribution.)



A $P(Z < 2)$ B $P(-2 < Z < 2)$

C $P(180 < Z < 220)$ D $P(-0.2 < Z < 0.2)$

8 X is a random variable that follows a normal distribution with a mean of 35 and a standard deviation of 7. The values of a and b are such that $P(a < X < b) = 0.95$ where this represents the middle 95% of values. a and b are best represented by:

A $a = 23.49$, $b = 46.51$ B $a = 21.28$, $b = 48.72$

C $a = 29.11$, $b = 40.89$ D $a = 21$, $b = 49$

9 The distribution of weights for all 80.5 cm-tall girls in a population is such that the mean weight is 10.3 kg with a standard deviation of 0.8 kg.

(a) Adah is 80.5 cm tall and weighs 8.3 kg. What is Adah's z value for weight?

These z values are used as a definition for some forms of malnutrition. Moderate acute protein-energy malnutrition is defined as having a z value in the range $[-3.0, -2.0)$ and severe acute protein-energy malnutrition is defined as having a z value less than -3.0.

(b) Based on these definitions, what sort of acute protein-energy malnutrition would Adah be diagnosed with?

(c) Jamilah, who is also 80.5 cm tall, has been diagnosed with severe acute protein-energy malnutrition. What is Jamilah's weight, correct to one decimal place, less than?

(d) Another girl who is 80.5 cm tall, Xhosa, is not diagnosed with either form of acute protein-energy malnutrition. What is Xhosa's minimum weight?

10 The percentages obtained by a group of students in a Mathematics examination are represented by a random variable M and are normally distributed with a mean of 72 and a variance of 121. All percentages are rounded to the nearest whole percentage.

(a) Calculate the probability that a student obtained a mark of at least 50% (when rounded to the nearest whole percentage) in this examination, correct to four decimal places, and the number of standard deviations that this mark is below the mean.

(b) Determine the z-score of a student who obtained a mark of 45%. What is the expected mark of a student whose z-score has the same size but opposite sign from the student who scored 45%?

To obtain an A^{++} mark, a student has to be in the top 2.5% of the group of students who have undertaken this examination.

(c) Calculate the minimum mark a student should obtain in this examination to be awarded an A^{++} by first finding the corresponding z value.

The marks in the previous year's Mathematics examination were normally distributed with a mean of 70 and a variance of 144.

(d) Would a student who obtained a mark of 94% have been awarded an A^{++} grade? Use appropriate calculations in your explanation.

11 Anita's daily charges for gas usage in her home form a normal distribution with an average daily cost of \$7.65 and a variance of 1.44, where the random variable C represents the daily cost for the gas used.

(a) What is the probability that in any one day Anita's cost is more than \$6.45?

(b) Determine the number of standard deviations from the mean for a cost of \$8.05 and a cost of \$6.65.

(c) Plot the two z values from part **(b)** on the normal distribution curve $N(7.65, 1.2^2)$.

12 Leonie completed two class tests this week.

Her results are shown in the table below with the class mean and standard deviation for each subject.

Subject	Leonie's mark	Mean	Standard deviation
Mathematics	72	64	4
Chemistry	78	68	10

(a) In which test did Leonie perform better, relative to her class peers? You must show working to support your answer.

(b) Chloe is in Leonie's class and they sat the same class tests. Her z-score for Mathematics is −1. What mark did Chloe record for Mathematics?

7.8 USING TABLES OF THE STANDARD NORMAL DISTRIBUTION

Only having the percentage values for 1, 2 and 3 standard deviations from the mean restricts the usefulness of the standard normal distribution. Questions may be answered by evaluating the corresponding integral using technology or the trapezoidal rule, so a more efficient method is needed for when access to technology is not available. This is achieved using a table of values for the standard normal distribution.

Consider the following example.

Example 27

The monthly mean daily global solar exposure, X MJ/m^2, for Nelson Bay for the 2018–19 year, is such that $X \sim N(17.3, 30.4)$. Find the probability that:

(a) $X < 15$ **(b)** $X > 15$ **(c)** $X < 25$ **(d)** $11.8 < X < 20$.

Solution

$X \sim N(17.3, 30.4)$: $\mu = 17.3$, Var $= 30.4$ so $\sigma = \sqrt{30.4} \approx 5.5$

(a) $X < 15$

$$z = \frac{15 - 17.3}{5.5} = \frac{-2.3}{5.5} = -0.42$$

$P(X < 15) = P(z < -0.42)$

Using the integral would require you to find

$$P(z < -0.42) = \int_{-\infty}^{-0.42} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\, dz$$
$$= 0.337\,24 \text{ using Desmos}$$

Remember that you cannot specify ∞ or −∞ as limits when using Desmos, instead use 5 and −5, for example.

(b) $X > 15$:

$$z = \frac{15 - 17.3}{5.5} = \frac{-2.3}{5.5} = -0.42$$

$P(X > 15) = P(z > -0.42)$

Using the integral would require you to find

$$P(z > -0.42) = \int_{-0.42}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\, dz$$
$$= 0.662\,76 \text{ using Desmos}$$

(c) X < 25: $z = \frac{25 - 17.3}{5.5}$

$$= \frac{7.7}{5.5}$$
$$= 1.4$$

$P(X < 25) = P(z < 1.4)$

Using the integral would require you to find

$$P(z < 1.4) = \int_{-\infty}^{1.4} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\, dz$$
$$= 0.919\,24 \text{ using Desmos}$$

(d) $11.8 < X < 20$: $\frac{11.8 - 17.3}{5.5} < z < \frac{20 - 17.3}{5.5}$

$$-\frac{5.5}{5.5} < z < \frac{2.7}{5.5}$$
$$-1 < z < 0.49$$

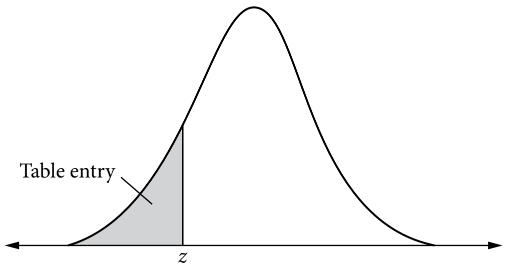
$P(11.8 < X < 20) = P(-1 < z < 0.49)$

Using the integral would require you to find

$$P(-1 < z < 0.49) = \int_{-1}^{0.49} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\, dz$$
$$= 0.529\,28 \text{ using Desmos}$$

This is a slow and cumbersome way to find a result. Prior to the current technology, mathematicians used a table of the standard normal distribution which gives the area to the left of a z score. You will now use a table to redo Example **26**.

Table A



The entries for z in Table A represent the area under the standard normal curve (bell curve) to the left of z.

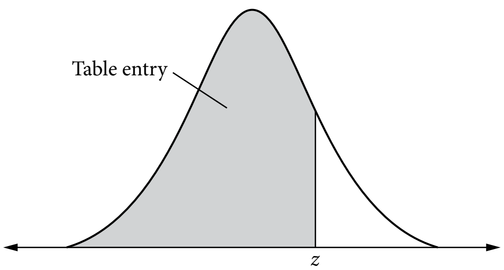
Negative values of z correspond to values that are less than the mean.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0016	.0016
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	0.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table A gives $P(z < k)$ for $-3.49 < k < 0$.

Thus $P(z < 0) = 0.5$, $P(z < -1) = 0.1587$, $P(z < -2.62) = 0.0044$, $P(z < -3.49) = 0.0002$.

Table B



The entries for z in this table represent the area under the standard normal curve (bell curve) to the left of z.

Positive values of z correspond to values which are greater than the mean.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.4885	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Table B gives $P(z < k)$ for $0 < k < 3.49$.

Thus $P(z < 0) = 0.5$, $P(z < 1) = 0.8413$, $P(z < 2.62) = 0.9956$, $P(z < 3.49) = 0.9998$.

Also, $P(z > 1) = 1 - P(z < 1) = 1 - 0.8413 = 0.1587 = P(z < -1)$, demonstrating the symmetry of the normal distribution and how Table B can be used on its own without the need for Table A.

Example 28

The monthly mean daily global solar exposure, X MJ/m2, for Nelson Bay for the 2018–19 year, is such that $X \sim N(17.3, 30.4)$.

Using the table of the standard normal distribution, find the probability that:

(a) $X < 15$ **(b)** $X > 15$ **(c)** $X < 25$ **(d)** $11.8 < X < 20$.

Solution

$X \sim N(17.3, 30.4)$: $\mu = 17.3$, Var $= 30.4$ so $\sigma = \sqrt{30.4} \approx 5.5$

(a) $X < 15$:

$$z = \frac{15 - 17.3}{5.5} = \frac{-2.3}{5.5} = -0.42$$

Use Table A: $P(X < 15) = P(z < -0.42) = 0.3372$

(b) $X > 15$:

$$P(X > 15) = P(z > -0.42) = 1 - P(z < -0.42) = 1 - 0.3372 = 0.6628$$

(c) $X < 25$:

$$z = \frac{25 - 17.3}{5.5} = \frac{7.7}{5.5} = 1.4$$

Use Table B: $P(X < 25) = P(z < 1.4) = 0.9192$

(d) $11.8 < X < 20$:

$$\frac{11.8 - 17.3}{5.5} < z < \frac{20 - 17.3}{5.5}$$

$$-\frac{5.5}{5.5} < z < \frac{2.7}{5.5}$$

$$-1 < z < 0.49$$

Use both tables: $P(11.8 < X < 20) = P(-1 < z < 0.49) = P(z < 0.49) - P(z < -1) = 0.6879 - 0.1587 = 0.5292$

From the symmetry of the standard normal distribution, as discussed earlier, the calculations in Example 27 could have been completed using only Table B.

Example 29

Using only Table B, find:

(a) $P(z < -0.42)$ **(b)** $P(z > -0.42)$ **(c)** $P(z < 1.4)$ **(d)** $P(-1 < z < 0.49)$.

Solution

(a) $P(z < -0.42) = 1 - P(z < 0.42)$

$= 1 - 0.6628$

$= 0.3372$

(b) By symmetry, $P(z > -k) = P(z < k)$

$P(z > -0.42) = P(z < 0.42)$

$= 0.6628$

(c) $P(z < 1.4) = 0.9192$

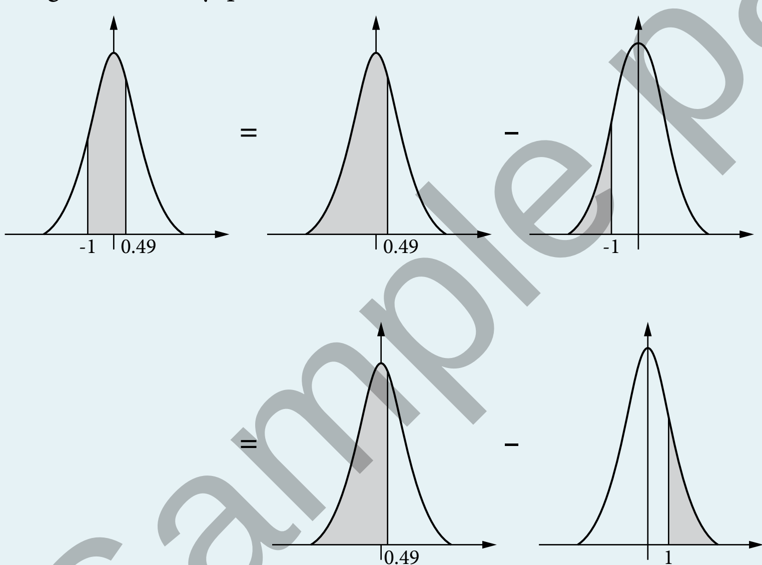
(d) $P(-1 < z < 0.49) = P(z < 0.49) - P(z < -1)$

$= P(z < 0.49) - \big(1 - P(z < 1)\big)$

$= 0.6879 - (1 - 0.8413)$

$= 0.5292$

Diagrammatically, part **(d)** can be shown as:



Example 30

Using the appropriate table of the standard normal distribution, find:

(a) $P(z < -1.2)$

(b) $P(z > -1.2)$

(c) $P(z < 2.3)$

(d) $P(z > 2.3)$

(e) $P(-1.2 < z < 2.3)$

(f) $P(z < -1.2 \text{ or } z > 2.3)$.

Solution

(a) $P(z < -1.2) = 0.1151$ (Using Table A)

$P(z < -1.2) = 1 - P(z < 1.2)$ (Using Table B)

$= 1 - 0.8849$

$= 0.1151$

(b) $P(z > -1.2) = 1 - P(z < -1.2)$

$= 1 - 0.1151$

$= 0.8849$

$P(z > -1.2) = P(z < 1.2)$ (Using Table B)

$= 0.8849$

(c) $P(z < 2.3) = 0.9893$ (Using Table B)

(d) $P(z > 2.3) = 1 - P(z < 2.3)$

$= 1 - 0.9893$

$= 0.0107$

(e) $P(-1.2 < z < 2.3) = P(z < 2.3) - P(z < -1.2)$

$= 0.9893 - 0.1151$

$= 0.8742$

(f) $P(z < -1.2 \text{ or } z > 2.3) = P(z < -1.2) + P(z > 2.3)$

$= 0.1151 + 0.0107$

$= 0.1258$

Using the standard normal distribution

Example 31

Use the standard normal distribution to find each of the following, correct to 4 decimal places:

(a) If $X \sim N(50, 100)$, find (i) $P(X < 70)$ (ii) $P(X < 45)$ (iii) $P(45 < X < 70)$

(b) If $X \sim N(2000, 40\,000)$, find (i) $P(X < 2100)$ (ii) $P(X < 2300)$ (iii) $P(2100 < X < 2300)$.

Solution

(a) $X \sim N(50, 100)$: $\mu = 50, \sigma = \sqrt{100} = 10$

(i) $X = 70$: $z = \dfrac{X - \mu}{\sigma} = \dfrac{70 - 50}{10} = 2$

$P(X < 70) = P(z < 2)$

$= 0.9772$

(ii) $X = 45$: $z = \dfrac{X - \mu}{\sigma} = \dfrac{45 - 50}{10} = -0.5$

$P(X < 45) = P(z < -0.5)$

$= 0.3085$

(iii) $P(45 < X < 70)$: $P(45 < X < 70) = P(X < 70) - P(X < 45)$

$= P(z < 2) - P(z < -0.5)$

$= 0.9772 - 0.3085$

$= 0.6687$

(b) $X \sim N(2000, 40\,000)$: $\mu = 2000, \sigma = \sqrt{40\,000} = 200$

(i) $X = 2100$: $z = \dfrac{X-\mu}{\sigma} = \dfrac{2100-2000}{200} = 0.5$

$P(X < 2100) = P(z < 0.5)$

$= 0.6915$

(ii) $X = 2300$: $z = \dfrac{X-\mu}{\sigma} = \dfrac{2300-2000}{200} = 1.5$

$P(X < 2300) = P(z < 1.5)$

$= 0.9332$

(iii) $P(2100 < X < 2300)$: $P(2100 < X < 2300) = P(z < 1.5) - P(z < 0.5)$

$= 0.9332 - 0.6915$

$= 0.2417$

Example 32

The weekly distances, M kilometres, travelled by a sales representative are normally distributed, given by $N \sim (365, 600)$.

Use the table of values for the standard normal distribution to find the probability that, in a random week, the sales representative will travel between 380 and 410 kilometres.

Solution

$\mu = 365, \text{Var}(M) = 600. \quad \sigma = \sqrt{600} = 10\sqrt{6}$

$$P(380 < M < 410) = P\left(\frac{380-365}{10\sqrt{6}} < Z < \frac{410-365}{10\sqrt{6}}\right)$$

$$= P\left(Z < \frac{45}{\sqrt{600}}\right) - P\left(Z < \frac{15}{\sqrt{600}}\right)$$

$$= P(Z < 1.837) - P(Z < 0.612) \approx P(Z < 1.84) - P(Z < 0.61)$$

$$= 0.9671 - 0.7291$$

$$= 0.238$$

To use the table of values for the standard normal distribution, values for Z need to be rounded to 2 decimal places.

EXERCISE 7.8 THE STANDARD NORMAL DISTRIBUTION

1 Find, correct to 4 decimal places:

(a) $P(z < -1.7)$ (b) $P(z < 1.7)$ (c) $P(z < -2.9)$ (d) $P(z < 0.3)$

2 Find, correct to 4 decimal places:

(a) $P(z > -1.9)$ (b) $P(z > 1.9)$ (c) $P(z > -2.7)$ (d) $P(z > 0.3)$

3 Find, correct to 4 decimal places:

(a) $P(z < -1.5)$ **(b)** $P(z > -1.5)$ **(c)** $P(z < 2.9)$

(d) $P(z > 2.9)$ **(e)** $P(-1.5 < z < 2.9)$ **(f)** $P(z < -1.5 \text{ or } z > 2.9)$

4 Use the standard normal distribution to find each of the following, correct to 4 decimal places:

(a) If $X \sim N(40, 100)$, find **(i)** $P(X < 60)$ **(ii)** $P(X < 35)$ **(iii)** $P(35 < X < 60)$.

(b) If $X \sim N(1000, 10\,000)$, find **(i)** $P(X < 1100)$ **(ii)** $P(X < 1300)$ **(iii)** $P(1100 < X < 1300)$.

5 X is a random variable that follows a normal distribution with a mean of 45 and a standard deviation of 8. The value of $P(31 < X < 59)$ is:

A 0.0401 **B** 0.0802 **C** 0.9198 **D** 0.9599

6 **(a)** Use the Trapezoidal rule with two sub-intervals to find the approximate value of

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx.$$

(b) Using the standard normal tables find $P(-1 < z < 1)$ and compare your answer to **(a)**.

(c) Comment on your answer in **(b)**.

7 Packets of 'Aunty Marys Rolled Oats' are labelled as having a mass of 1 kg. The machine used to fill these packets has been tested and found to follow a normal distribution with a mean of 1.02 kg and a standard deviation of 10 g.

(a) What percentage of packets will have a mass less than 1 kg?

(b) What percentage of packets will have a mass between 1 kg and 1.02 kg?

(c) Find the probability that a packet selected at random will have a mass more than 1 kg.

(d) The manufacturer expects only 5% of the packets to have a mass less than 1 kg. If they reset the mean mass of packets produced by the machine to 1.017 kg, keeping the same standard deviation, will this target now be met?

8 A certain variety of tangelo has mass, X g, such that $X \sim N(140, 400)$. Find:

(a) $P(X > 130)$

(b) $P(X < 145)$

The tangelos are classified as large if they have a mass greater than 170 g. Find:

(c) The proportion of large Tangellos.

(d) The probability of a large Tangello having a mass less than 190 g.

9 It is given that $X \sim N(20, 9)$. Using the standard normal distribution tables, verify that:

(a) about 68% of the results lie within one standard deviation of the mean

(b) about 95% of the results lie within two standard deviations of the mean

(c) about 99.7% of the results lie within three standard deviations of the mean.

10 The percentages obtained by a group of students in an Advanced Mathematics examination are represented by a random variable A and are normally distributed with a mean of 70 and a variance of 121. All percentages are rounded to the nearest percent.

(a) Calculate the probability that a student obtained a mark of at least 50% (when rounded to the nearest percentage) in this examination, correct to four decimal places.

(b) Determine the probability that a student obtained a mark less than 45% and then calculate the mark (m) for which $P(A > m)$ has this same probability.

(c) To obtain an A^+, a student must be in the top 10% of the cohort of students who have undertaken this examination. Calculate the minimum mark a student should have obtained to be awarded an A+, by first finding the corresponding z-value.

(d) The marks in the previous year's Advanced Mathematics examination were normally distributed with a mean of 70 and a variance of 144. Would a student who obtained a mark of 86% have been awarded an A^+ grade? Use appropriate calculations in your explanation.

11 In a particular city, the weekly rents for apartments are normally distributed with a mean of \$750 and a standard deviation of \$100.

(a) Find the probability that the weekly rent of an apartment chosen at random is less than \$600.

(b) Sam's budget for the weekly rent is \$600. She has refined her searches and produced a list such that she is only looking at properties that are less than the mean weekly rent. Find the probability that an apartment that Sam chooses at random from this list is within her budget.

12 A random variable is normally distributed with mean 0 and standard deviation 1.

Using the table of values, determine the probability that this random variable will lie between $z = -0.75$ and $z = 1.5$.

13 The arm spans (in metres) of a group of 2500 residents of a town are normally distributed with a mean of 1.63 metres and a standard deviation of 0.26 metres. How many residents will have an arm span greater than 1.5 metres?

14 A bag of jelly babies is labelled as having a net weight of 220 grams. The production process produces bags whose weights are normally distributed with a mean of 224.7 grams and a standard deviation of 2.35 grams.

(a) If 1200 bags of jelly babies are produced, how many bags would be expected to contain less than the labelled weight?

(b) In the batch of 1200 bags of jelly babies from part **(a)**, how many bags would be expected to have a weight of more than 231 grams, 5% over the labelled weight of 220 grams?

7.9 THE INVERSE STANDARD NORMAL DISTRIBUTION

Tables A and B may also be used to find the z values between which a given percentage of the values lie. This means that you must find the percentage in the table, as near as possible, and then read off the z value.

Example 33

For what value of a is:

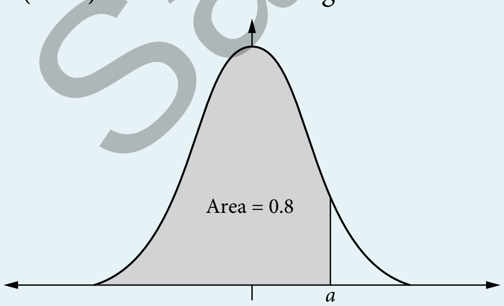
(a) $P(z < a) = 0.8$, which is the same as $P(z < a) = 80\%$

(b) $P(z > a) = 0.9$, which is the same as $P(z > a) = 90\%$

(c) $P(-a < z < a) = 0.6$, which is the same as $P(-a < z < a) = 60\%$?

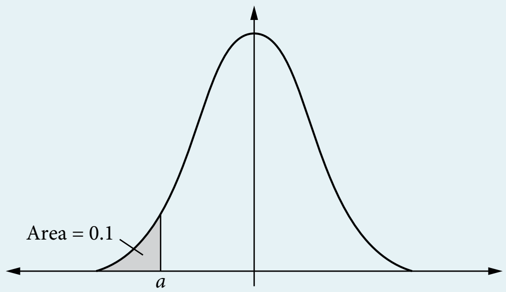
Solution

(a) $P(z < a) = 0.8$ means finding the value for a such that 80% of the values of z are less than a.



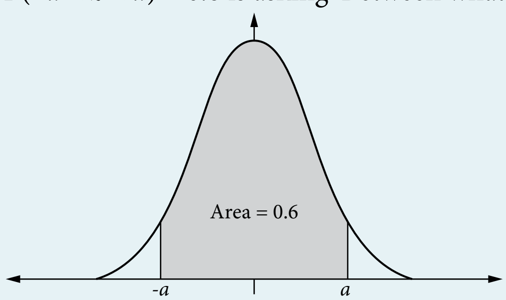
From Table B, $z = 0.84$ gives an area of 0.7995.
From Table B, $z = 0.85$ gives an area of 0.8023.
0.7995 is closer to 0.8 than 0.8023.
Hence $P(z < a) = 0.8$ gives $a = 0.84$.

(b) $P(z > a) = 0.9$: This becomes $P(z < a) = 0.1$ where a is negative.

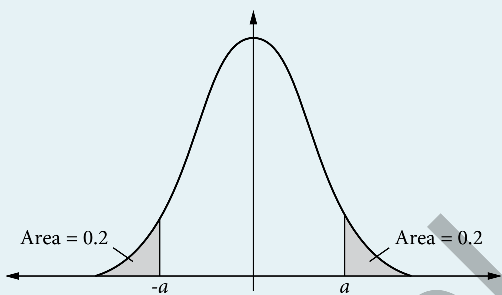


From Table A, $z = -1.29$ gives an area of 0.0985.
From Table A, $z = -1.28$ gives an area of 0.1003.
To 2 decimal places, the best approximation is $a = -1.28$.

(c) $P(-a < z < a) = 0.6$ is asking 'Between what values do the middle 60% of scores lie?'



This means that 40% or 0.4 is in the two tails, or 0.2 in each tail.



Thus the question has become $P(z < a) = 0.8$ since the area to the right of $z = a$ is 0.2.
Hence from **(a)**, $P(z < 0.84) = 0.7995$.
Hence $a = 0.84$ to 2 decimal places.

The above calculations become a lot easier if the following form of the table is available.

Table C

Values of a for selected values or $P(z \leq a)$							
a	0.842	1.036	1.282	1.645	1.960	2.326	2.576
$P(z \leq a)$	0.800	0.850	0.900	0.950	0.975	0.990	0.995

Table C removes the need to do many of the steps needed when using Table B in the previous examples.
Table C gives you the value of a for which $P(z < a)$ is given.
Thus if $P(z < a) = 0.800$ (80%), then Table C gives you directly that $a = 0.842$.

Example 34

Using Table C, find a if:

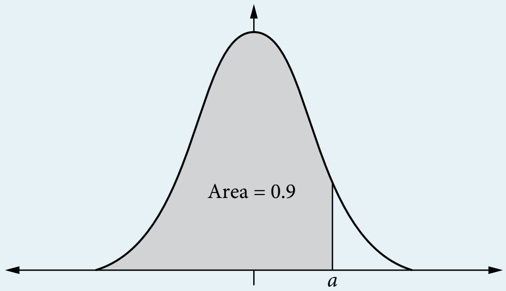
(a) $P(z < a) = 0.9$
(b) $P(z < a) = 0.1$
(c) $P(z > a) = 0.975$
(d) $P(-a < z < a) = 0.05$.

Solution

Consider a diagram to help interpret each part.

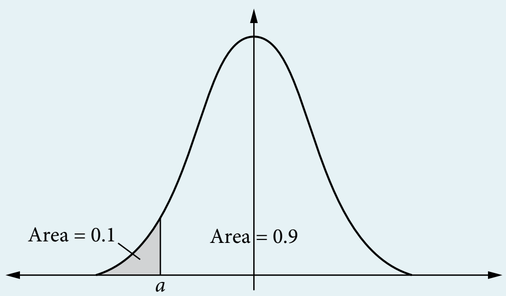
(a) $P(z < a) = 0.9$

$a = 1.282$ from Table C.



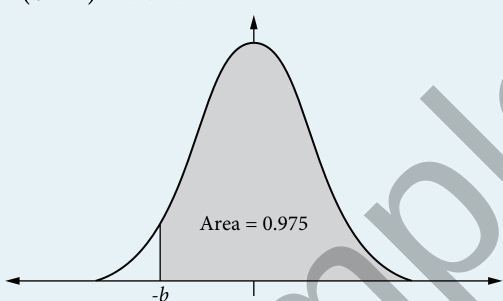
(b) $P(z < a) = 0.1$

From the symmetry of the curves in parts **(a)** and **(b)** it can be seen that a is negative and $a = -1.282$.



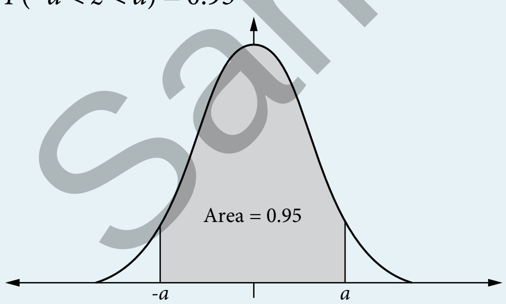
(c) $P(z > a) = 0.975$

Let $b = -a$ so you require $P(z > -b) = 0.975$.
From Table C, $P(z < 1.960) = 0.975$.
By symmetry, $b = 1.960$ and hence $a = -1.960$.



(d) $P(-a < z < a) = 0.95$

A diagram is needed as this will have two tails.
95% of the area lies between $-a$ and a, so this means that a total of 5% of the area lies in the two tails, that is 2.5% (0.025) in each tail.
Thus the area to the left of a is 0.975 and this is the value that you need to look for in Table C.
$P(-a < z < a) = P(-1.960 < z < 1.960)$



Example 35

The annual rainfall, X millimetres, in the town of Aquaville is such that $X \sim N(2000, 40\,000)$. Give the answers to the following questions correct to the nearest millimetre.

(a) Under how many millimetres would you expect 95% of the yearly results to be?

(b) Find k, such that $P(X > k) = 0.8$.

(c) Find k, such that $P(X < k) = 0.3$.

Solution

$X \sim N(2000, 40\,000)$: $\mu = 2000$, $\sigma = 200$

(a) $P(X < k) = 0.95$:

$$P\left(z < \frac{k - 2000}{200}\right) = 0.95$$

$$1.645 = \frac{k - 2000}{200}$$

$$329 = k - 2000$$

$$k = 2329$$

You would expect 95% of the years to have a rainfall under 2329 mm.

(b) $P(X > k) = 0.8$:

From the table, $P(z < 0.842) = 0.8$

By symmetry, $P(z > -0.842) = 0.8$

$$-0.842 = \frac{k - 2000}{200}$$

$$-168.4 = k - 2000$$

$$k = 1831.6$$

$$\approx 1832$$

You would expect 80% of the years to have a rainfall greater than 1832 mm.

(c) $P(X < k) = 0.3$:

From Table A, $z = -0.52$ (2 decimal places)

$$-0.52 = \frac{k - 2000}{200}$$

$$-104 = k - 2000$$

$$k = 1896$$

You would expect 30% of the years to have a rainfall less than 1896 mm.

Finding the mean using the inverse normal distribution

Example 36

The height, H cm, of a population of people is such that $H \sim N(\mu, 25)$. In addition it is known that $P(H < 180) = 0.990$. Find the value of the mean, μ, to the nearest integer.

Solution

$\mu = ?$, $\sigma = 5$, $P(H < 180) = 0.990$:

From Table C, $z < 2.326$

$$2.326 = \frac{180 - \mu}{5}$$

$$11.63 = 180 - \mu$$

$$\mu = 168.37$$

$$\mu \approx 168$$

Example 37

The height, H centimetres, of a population of people is such that $H \sim N(175, \sigma^2)$. In addition it is known that $P(H < 190) = 0.975$. Find the value of σ, correct to 2 decimal places.

Solution

$\mu = 175, \sigma = ?, P(H < 190) = 0.975$:

From Table C, $z < 1.960$

$$1.96 = \frac{190 - 175}{\sigma}$$

$$\sigma = \frac{15}{1.96}$$

$$= 7.653$$

$$\approx 7.65$$

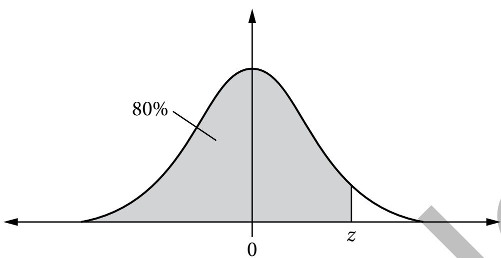
EXERCISE 7.9 THE INVERSE STANDARD NORMAL DISTRIBUTION

1 X is normally distributed with $\mu = 50$, $\sigma = 4$. If $P(X < a) = 0.95$, then a is closest to:

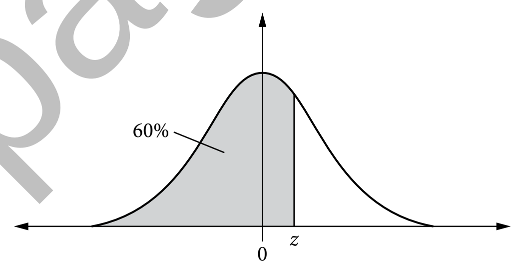
A 1.645 **B** 56.58 **C** 1.96 **D** 57.84

2 For each of the following diagrams, state the value of z, as accurately as the tables allow.

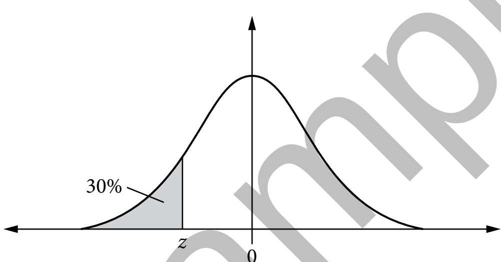
(a)



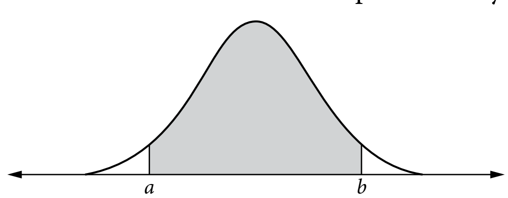
(b)



(c)



3 The diagram below represents the distribution of X, where $X \sim N(25, 8)$. The shaded area represents the middle 80% of observations. The values of a and b are best represented by:



A $a = -0.842, b = 0.842$ **B** $a = 22.62, b = 27.38$

C $a = -1.282, b = 1.282$ **D** $a = 21.38, b = 28.62$

4 The height, X metres, of a population of people is known to be normally distributed with $X \sim N(1.7, 0.0225)$. Give your answers correct to 2 decimal places and include the required units.

(a) Find the height under which you would expect to find 97.5% of the population.

(b) Find the value of k, such that $P(X > k) = 0.7$.

(c) Find the value of k, such that $P(X > k) = 0.35$.

5 The weight, X kilograms, of a population is known to be normally distributed, with $X \sim N(60, 90.25)$. Give your answers correct to 2 decimal places and include the required units.

(a) Find the weight under which you would expect to find 85% of the population.

(b) Find the value of k, such that $P(X < k) = 0.4$

6 **(a)** The height, H centimetres, of a population is such that $H \sim N(\mu, 25)$. If $P(H < 160) = 0.1587$, find the value of the mean.

(b) The height, H centimetres, of a population is such that $H \sim N(\mu, 36)$. If $P(H < 180) = 0.9525$, find the value of the mean.

(c) The arm length, L centimetres, of a population of people is such that $L \sim N(\mu, 16)$. If $P(L < 75) = 0.105$, find the value of the mean.

7 **(a)** The height, H centimetres, of a population of students is such that $H \sim N(165, \sigma^2)$. If $P(H < 160) = 0.1587$, find the value of the standard deviation.

(b) The height, H centimetres, of a population of students is such that $H \sim N(165, \sigma^2)$. If $P(H < 170) = 0.7967$, find the value of the standard deviation.

(c) The arm length, L centimetres, of a population of students is such that $L \sim N(80, \sigma^2)$. If $P(L < 85) =$ 0.9525, find the value of the standard deviation.

8 Given $Z \sim N(0, 1)$, find k in each case.

(a) $P(-k < Z < k) = 0.85$

(b) $P(Z < -k \cup Z > k) = 0.05$, which is another way of writing $P(Z < -k \text{ or } Z > k) = 0.05$

(c) $P(-k < Z < k) = 0.88$

(d) $P(Z < -k \cup Z > k) = 0.1$

9 Use Table A, B or technology to complete the following table, which is an expanded form of Table C.

Values of a for selected values of $P(z < a)$									
a					0			0.842	1.282
$P(z < a)$	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90

10 The scores, X, on an aptitude test are such that $X \sim N(75, 35)$. Find the following.

(a) The interquartile range of the scores (i.e. the middle 50% of scores).

(b) What scores did at most 20% of the participants obtain?

(c) What proportion of the scores were less than 50?

11 The monthly rainfall at the Solaris resort follows a normal distribution with a standard deviation of 12 mm. If the rainfall is less than 25 mm in 15% of the months, find the mean monthly rainfall.

12 The scores, X, in a national mathematics competition are such that $X \sim N(80, 20)$. All scores are integers. Find the following.

(a) The score needed to obtain a High Distinction certificate, if these are awarded to the top 3% of students.

(b) The minimum score needed to obtain a Distinction certificate if they are awarded to the next 10% of the candidates.

(c) The minimum integer score needed to obtain a Credit certificate if they if they are awarded to the next 22% of the candidates.

13 The amount of honey, X kilograms, taken from a hive follows a normal distribution with a mean of 2.1 kg and a standard deviation of $\sqrt{0.6}$ kg. Find, correct to three decimal places:

(a) the value, k, such that $P(X < k) = 0.92$

(b) the value, k, such that $P(X > k) = 0.62$

(c) the value, k, such that $P(X < k) = 0.45$

(d) the value, k, such that $P(X > k) = 0.16$

14 Let $Z \sim N(0, 1)$. Find the value of t, correct to three decimal places, if:

(a) $P(X < t) = 0.6789$

(b) $P(X > t) = 0.2357$

(c) $P(0 < X < t) = 0.1234$

(d) $P(t < X < 2.1) = 0.4680$

Now answer the following questions using only a scientific calculator.

(e) Given $P(0 < X < t) = 0.1234$, find $P(X < t)$

(f) Given $P(X < t) = 0.6789$, find $P(0 < X < t)$

(g) Given $P(X > \mathrm{t}) = 0.2357$, find $P(0 < X < \mathrm{t})$

15 A BestReact driving simulator was designed to check the reaction time of drivers. The random variable R representing the reaction times, in milliseconds, of teenagers is normally distributed with a mean of 250 and a variance of 2500.

(a) Calculate the probability that any one teenager had a reaction time of between 320 milliseconds and 350 milliseconds.

(b) What is the highest reaction time, k, such that the probability that any one teenager has a reaction time greater than k is 0.0580?

The same test was used to check the reaction time of drivers over 20. The random variable R representing the reaction times, in milliseconds, of these drivers is normally distributed as well with a mean of 280 ms and a variance of σ^2.

(c) Calculate the standard deviation for this set of data if the probability that any one driver over 20 has a reaction time greater than 300 ms is 0.1817.

16 A group of students had their times recorded in the running trials. These times, in seconds, are represented by the random variable X which is normally distributed with a mean μ and a variance 31.36.

(a) Calculate the mean of the normal distribution if $P(X < 57) = 0.2961$.

(b) Determine the interquartile range for this set of data, correct to one decimal place.

(c) Determine the time that places a student in the top 15% of fastest runners. (Hint: what sort of times do the fastest runners have?)

17 A farming company harvests oranges to sell to grocery stores. Oranges that weigh less than 150 g are rejected and used for juicing instead. Typically, 16% of all oranges are rejected. Also, oranges that are larger than 300 g are reserved for restaurants. Typically, 2.5% of all oranges are sold to restaurants.

If the weights of the oranges are normally distributed, find the mean and standard deviation of the normal distribution of oranges.

CHAPTER REVIEW 7

1 In each of the following find the value of k that makes the table a discrete probability distribution. In each case express the value in simplest fraction form.

(a)

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{8}$	$\frac{1}{16}$	k	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{3}{16}$

(b)

x	5	6	7	8	9	10
$P(X = x)$	$\frac{1}{12}$	k	$\frac{1}{6}$	$\frac{1}{3}$	$2k$	$\frac{1}{4}$

(c)

x	−1	0	1	2	3
$P(X = x)$	k	$2k$	$3k$	$4k$	$5k$

2 Two six-sided dice, numbered 3 to 8, are rolled. Let X be the total of the two dice.

(a) Complete the following table that shows the probability distribution of X. Express the probabilities as fractions in simplest form.

x	6	7	8	9	10	11	12	13	14	15	16
$P(X = x)$		$\frac{1}{18}$	$\frac{1}{12}$		$\frac{5}{36}$			$\frac{1}{9}$			$\frac{1}{36}$

(b) Find $P(X \geq 11)$.

(c) Find $P(X < 15)$.

(d) Find $P(7 \leq X \leq 11)$.

3 A spinner showing the numbers 1, 2, 3 and 4 is spun twice. Let X be the total of the two numbers obtained.

(a) Complete the table using fractions in simplest form.

x	2	3	4	5	6	7	8
$P(X = x)$							

(b) Use your table to find each of the following values:

(i) $E(X)$ **(ii)** $\text{Var}(X)$

4 A random variable, X, has the following probability distribution.

x	−2	−1	0	1	2
$P(X = x)$	0.15	0.2	0.1	0.35	0.2

Find the following, rounding your answers to 4 decimal places where necessary.

(a) $E(X)$ **(b)** $\text{Var}(X)$ **(c)** $E(3X - 2)$ **(d)** $E(X^2 - 2X)$ **(e)** $\text{Var}(3X - 2)$

5 A game at the local festival requires a player to toss two coins. If two heads are obtained, the player receives \$2. Otherwise, the player loses the cost of the game. For the game to be fair, what should it cost in dollars?

6 Technology is used to select a random number from 1 to 20 inclusive. The variable, R, represents the value chosen. Find the following values:

(a) $E(R)$ **(b)** $\text{Var}(R)$

7 Each of the numbers from 1 to n appearing on an n-sided spinner has an equal chance of appearing. If the variable S describes this situation, and $\text{Var}(S) = 18\frac{2}{3}$, what is the value of n?

8 A coin is biased in such a way that $P(\text{H}) = 3 \times P(\text{T})$. The coin is tossed 100 times. Let X stand for the number of tails obtained. Find the value of $E(X)$.

9 When Yehudi and Carlos play racquetball, the probability that Yehudi wins a point is 0.35.

(a) How many points would you expect Yehudi to win from the first 15 points? Give your answer to the nearest whole number of points.

(b) Choose the correct terms in the following statement.
If Yehudi won 10 out of the first 15 points I would [not be/be slightly/be very] surprised as the number is [about the same as/just above/well above] the expected number.

10 A box contains 10 items, of which three are defective. A sample of three items is selected at random from the box. Let Y represent the number of defective items selected.

(a) Complete the table to show the probability distribution of the variable. State the probabilities in simplest fraction form.

(b) Find the expected number of defective items, $E(Y)$.

y	0	1	2	3
$P(Y = y)$				

11 Which one of the following random variables is not associated with a discrete scale?

A The number of dim sims sold at the local fish and chip shop per day
B Your height, to the nearest centimetre, measured monthly for two years
C The time taken to fill your bath each night for one month
D The number of goals scored in each netball match by a team throughout the season

12 A jar contains seven white marbles, three green marbles and two blue marbles. Two marbles are drawn, with replacement, from the jar. What is the probability of drawing exactly one white marble?

A $\frac{7}{12} \times \frac{7}{12}$ **B** $\left(\frac{7}{12} \times \frac{5}{11}\right) + \left(\frac{5}{12} \times \frac{7}{11}\right)$ **C** $2 \times \frac{7}{12} \times \frac{5}{12}$ **D** $\frac{7}{12} \times \frac{5}{12}$

13 What is the value of t in the following probability distribution table?

x	0	1	2	3
$P(X=x)$	t	$2t$	$3t$	$4t$

A 10 **B** 0.1 **C** 6 **D** 0.6

14 Two six-sided dice are rolled. Let X be the total shown on the two dice.

(a) What is $P(X \geq 10)$?
A $\frac{11}{12}$ **B** $\frac{1}{6}$ **C** $\frac{1}{36}$ **D** $\frac{1}{12}$

(b) What is $P(5 < X \leq 9)$?
A $\frac{4}{9}$ **B** $\frac{1}{3}$ **C** $\frac{2}{3}$ **D** $\frac{5}{9}$

(c) What is the expected value, $E(X)$?
A 7 **B** 3 **C** 6.5 **D** 7.5

15 A probability distribution table is shown.

x	1	2	3	4	5
$P(X=x)$	0.2	0.1	0.35	0.05	0.3

(a) What is the expected value, $E(X)$?
A 0.2 **B** 1.34 **C** 1.8 **D** 3.15

(b) What is the expected value, $E(2X-3)$?
A 3 **B** 3.3 **C** 3.5 **D** 3.7

(c) What is the variance, $\text{Var}(X)$?
A 2.13 **B** 1.46 **C** 8.81 **D** 2.97

16 For the random variable X it is known that $E(X) = 2.3$. If the expected value $E(X^2) = 6.2$, what is the standard deviation σ?

A 0.8281 **B** 0.91 **C** 0.95 **D** 2.3

17 A variable Y has the probability distribution shown in the following table.

y	−2	−1	0	1
$P(Y=y)$	0.4	0.3	0.2	0.1

What is the variance, $\text{Var}(Y)$?

A 1 **B** −1 **C** 2.2 **D** 1.1

18 A spinner is equally divided into n segments. Each segment contains a different value from 1 to n

(a) If the expected value is 26.5, what is the value of n?
A 26 **B** 51 **C** 52 **D** 53

(b) If the expected value is 18, what is the value of n?
A 9 **B** 10 **C** 34 **D** 35

(c) If the expected value is 101, what is the value of n?
A 201 **B** 200 **C** 52 **D** 51

19 Audrey is playing a popular quest-style game on her games console. It randomly generates whole numbers in the range 3 to 8 each time she hits the *Play* button. She cannot start her quest until the total of the scores is 7 or greater. Give exact answers, in simplest fraction form, for this question.

(a) Find the probability distribution of the number of times Audrey hits *Play*, given the variable X, required to achieve this total.

(b) Find the expected number of hits of *Play* required.

(c) Find the variance for the number of hits of *Play* required. (If you cannot get the exact answer for this part write the answer as a decimal correct to 3 decimal places.)

20 The discrete random variable X can take only the values 1, 2, 3, 4, 5 and 6. The probability distribution of X is described by the following statements:

$P(X=1)=P(X=3)=P(X=5)=a$
$P(X=2)=P(X=4)=P(X=6)=b$
$a=3b$

(a) Find the values of a and b.
(b) Draw up a probability distribution table for X.
(c) Show that $E(X)=3\frac{1}{4}$.
(d) Show that $\text{Var}(X)=2\frac{41}{48}$.
(e) Find the probability that the sum of two independent observations from this distribution is greater than 9.

21 Freya rolls a normal six-sided die marked 1 to 6. If she obtains a 4 she rolls the die a second time, and in this case her score is the sum of 4 and the second number obtained. If she does not obtain a 4, her score is the number rolled. Freya has at most two rolls of the die. Let Z be the random variable representing Freya's score.

(a) Draw a table showing the probability distribution of Z.
(b) What is the expected value, $E(Z)$?
(c) What is $P(Z>6)$?
(d) What is $P(Z<7)$?
(e) What is $P(Z>4 \mid Z\le 8)$?

22 Use mathematical reasoning to explain why the function given by $P(X=x)=\frac{2x-3}{15}$ is a probability distribution for $x\in\{3, 4, 5\}$ but it is not a probability distribution for $x\in\{1, 2, 3, 4, 5\}$.

23 Many calculators can produce random numbers as whole-number values within a specified range.

(a) If the technology produced truly random numbers, what type of distribution would you expect the results to follow?
(b) Listed below are 96 random numbers (as displayed by a calculator) in the range 1 to 6, inclusive.

3	6	2	5	6	2	3	1	6	1	1	4
6	6	2	2	1	1	5	1	3	2	6	1
6	4	2	6	5	2	2	6	6	1	1	5
6	4	2	1	6	2	3	1	2	6	3	4
4	1	1	1	5	5	6	3	1	3	6	5
4	4	6	5	5	4	3	2	4	6	1	3
3	2	5	6	3	3	2	4	4	5	5	3
6	4	3	3	5	6	2	5	3	3	2	1

Create a probability distribution table based on these results. To make analysis easier, write the probabilities with a denominator of 96 and use X as the variable.

(c) What do these results suggest about whether or not the numbers are truly random? Refer to the size of the sample space in your answer.
(d) Calculate $E(X)$ and $\text{Var}(X)$ for the sample of 96 random values. Compare them to the theoretical values for the underlying distribution.
(e) Now use technology to generate a similar set of 96 numbers in the range 1 to 6. Calculate $E(X)$ and $\text{Var}(X)$ for the sample of 96 random values that you generated.
(f) Compare the three sets of statistics that you have now calculated.
(g) What do you think would happen if you generated 5000 random numbers in the range 1 to 6?

24 A standard six-sided die is rolled until an even number shows or five odd numbers in a row have shown.

(a) Draw up a probability table showing the number of rolls and the associated probabilities.
(b) Find the expected number of rolls. Give your answer correct to 1 decimal place.
(c) Consider tossing a coin where you will stop as soon as the coin lands on tails or if five heads in a row appear. Explain how you can easily state the expected number of tosses of the coin.

25 Raduhas developed a website where subscribers can pick up to six motivational videos per day to watch. The number of each video chosen is a discrete random variable X with a probability distribution formula $P(X = x) = k[30 - (x - 1)^2]$, where $x \in \{1, 2, 3, 4, 5\}$.

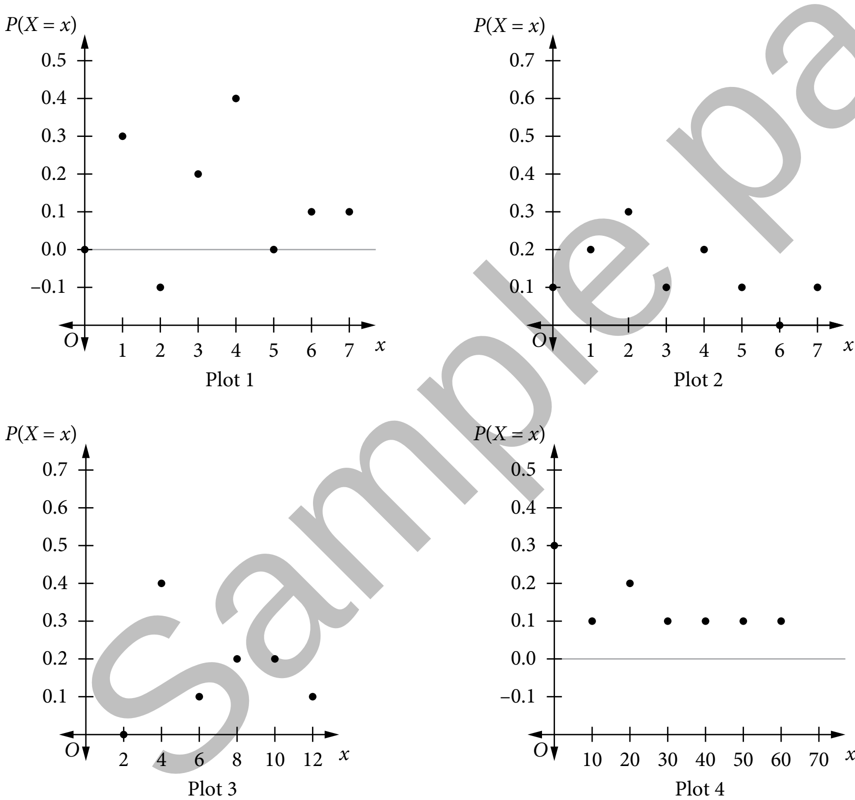
(a) Calculate the value of k.
(b) Display all probabilities as fractions in simplest form in a probability distribution table.
(c) Draw a scatterplot of the distribution.

26 Karina and Achim print and sell business cards in packs of 500. They sell a maximum of 4 packs per hour and incur an average of \$16 running costs per hour. The average price for a pack of 500 business cards is \$20. The probability distribution table shown represents the random variable X, the number of packs they sell in any given hour.

X	0	1	2	3	4
$P(X = x)$	0.012	0.097	0.138	0.325	0.428

(a) Calculate the number of packs of business cards they can expect to print and sell in any given hour.
(b) Determine whether Karina and Achim incur a loss or make a profit, and the amount on average, in any given hour.

27 Four scatterplots are shown.



(a) Establish which of the four scatterplots represents a probability distribution function. Explain your answer using mathematical reasoning.
(b) Calculate the mean and variance of the plots that represent probability distribution functions.

28 For the uniform continuous variable with probability density function $f(x) = \begin{cases} k, & 0 \le x \le 10 \\ 0, & \text{otherwise} \end{cases}$, find the following values.

(a) k (b) $P(X \le 3)$ (c) $P(X \le 5)$ (d) $P(2 \le X \le 9)$

29 Does the hybrid function $f(x) = \begin{cases} x^2 + 4x - \frac{10}{3}, & 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$ represent a probability density function?

30 For the continuous variable with probability density function $f(x) = \begin{cases} \frac{x^2}{12}, & 0 \le x \le \sqrt[3]{36} \\ 0, & \text{otherwise} \end{cases}$, find the exact value of the median.

31 Honey is packed in jars with a labelled mass of 500 g. The actual amount X g of honey in the jar is such that $X \sim N(500, 25)$.

(a) Find the probability that a jar chosen at random contains less than 495 g.

(b) Jars containing less than 490 g cannot be sold. In a batch of 1000 how many jars would you expect to be rejected?

32 If $X \sim N(12, 16)$, find the exact z values corresponding to the following x values.

(a) $x = 8$ **(b)** $x = 20$

33 If $X \sim N(28, 16)$, find the exact z values corresponding to the following x values.

(a) $x = 24$ **(b)** $x = 40$

34 If $f(x) = \begin{cases} \frac{2x}{15+k}, & 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$ defines a probability density function, then the value of k is:

A $\frac{2}{15}$ **B** -6 **C** $-\frac{2}{15}$ **D** $\frac{1}{5}$

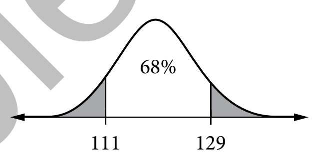
35 $X \sim N(10, 9)$. You would expect about 95% of observations to be in the range:

A -8 to 28 **B** 0 to 28 **C** 10 to 19 **D** 4 to 16

36 Consider the following graph.

The graph is best described by:

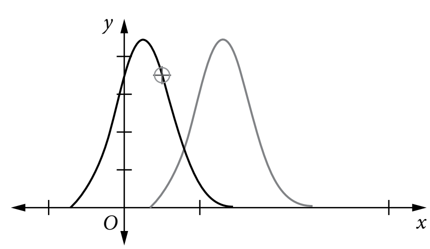
A $N(111, 129)$ **B** $N(120, 81)$
C $N(120, 3)$ **D** $N(120, 9)$



37 The graph shows two normal distributions drawn on the same set of axes. Graph 1 has a marker on it.

Compared to Graph 1, Graph 2 has:

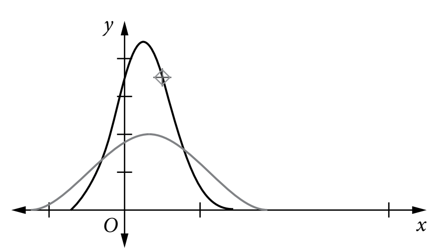
A the same mean but a larger variation.
B the same mean but a smaller variation.
C the same variation but a larger mean.
D the same variation but a smaller mean.



38 The graph shows two normal distributions drawn on the same set of axes. Graph 1 has a marker on it.

Compared to Graph 1, Graph 2 has:

A the same mean but a larger variation.
B the same mean but a smaller variation.
C the same variation but a larger mean.
D the same variation but a smaller mean.



In any of the following questions that involve evaluating an integral, you may need to use software to evaluate the integral if the integrand is outside the scope of this course, or use the trapezoidal rule.

39 An arrow is fired at a target. The distance, X metres, from the centre of the target that the arrow hits pierces follows the function $f(x) = \begin{cases} \dfrac{3}{\pi\left(1+x^2\right)}, & 0 \le x \le \sqrt{3} \\ 0, & \text{otherwise} \end{cases}$. You are given that $\int_0^{\sqrt{3}} \dfrac{dx}{1+x^2} = \dfrac{\pi}{3}$.

(a) Explain why it is reasonable to say that this represents a probability density function.
(b) Find:
(i) $P(X < 1)$ **(ii)** $P(X > 1.25)$ **(iii)** $P(0.25 < X < 1.1)$
(c) Find $E(X)$.
(d) Find $\text{Var}(X)$.

40 Andrea is waiting for a taxi at a popular venue at a busy time of the day. The continuous random variable T, in hours, that represents the time spent by people waiting for a taxi at this venue, is given by the probability density function: $f(t) = \begin{cases} \dfrac{5}{2}t\sqrt{t}, & 0 \le t \le 1 \\ 0, & \text{otherwise} \end{cases}$.

(a) Calculate the average time Andrea should expect to wait to catch a taxi. Give your answer correct to the nearest minute.
(b) Calculate the probability that Andrea will have to wait more than 15 minutes to catch a taxi, correct to two decimal places.
(c) What is the probability that Andrea's wait for the taxi is between 10 and 40 minutes?
(d) If 10 people are waiting for a taxi, how many would be expected to wait more than 30 minutes?

41 For their Mathematics assignment, Xandra and Zack have decided to record the times they spend completing the daily homework. Xandra's times are represented by the continuous random variable, X, in hours, with the probability density function:

$$f(x) = \begin{cases} -\dfrac{1}{9}\left(x^2 - 4x\right), & 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}.$$

Zack's times are represented by the continuous random variable, Z, in hours, with the probability density function:

$$f(z) = \begin{cases} \dfrac{2}{57}\left(z^2 - 9z + 20\right), & 0 \le z \le 3 \\ 0, & \text{otherwise} \end{cases}.$$

(a) Calculate the average time each student spends completing homework each day and compare the results.

Xandra has a school concert to attend and has to complete her set homework within 1.5 hours.

(b) Calculate the probability that Xandra will complete her homework in the required time.
(c) Calculate the variance for the two variables.

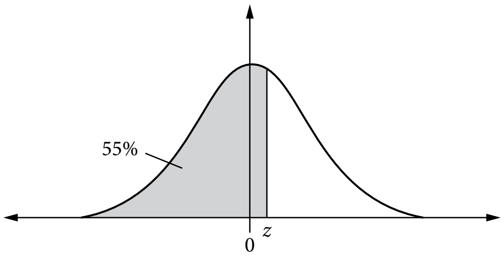
42 When fishing, fish shorter than a given length have to be thrown back into the water. Red snapper have a minimum allowed length of 30 cm and barramundi have a minimum allowed length of 55 cm. Let the variable R represent the lengths of red snapper caught, in cm, and the variable B represent the lengths of barramundi caught, in cm.

Observations of caught red snapper show that R is normally distributed with mean 36 and a variance of 9. Observations of caught barramundi show that B is normally distributed with a variance of 16.

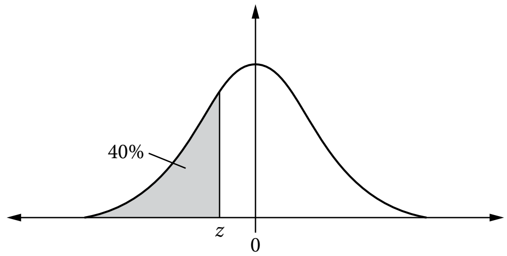
(a) Calculate the mean length of barramundi caught, correct to the nearest whole number, if 2.5% of the barramundi caught are less than 54 cm.
(b) Calculate the z-scores for the minimum allowed lengths for both variables, R and B, and interpret what this means in terms of which of the two species are more likely to be too small to keep.

43 For each of the following diagrams, state the value of z, as accurately as the tables allow.

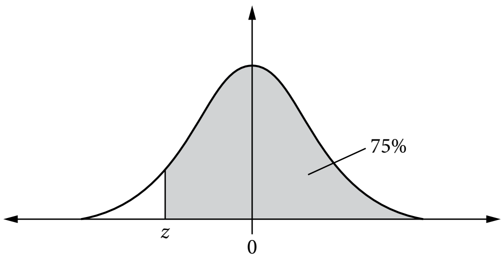
(a)



(b)



(c)



44 If $X \sim N(30, 12)$, find the following, correct to1 decimal place.

(a) The values between which you would expect the middle 60% of observations to lie.

(b) The value above which you would expect to find 5% of the observations.

45 If $X \sim N(120, 40)$, find k, correct to 2 decimal places.

(a) $P(X > k) = 0.15$ **(b)** $P(X < k) = 0.25$

46 The marks in an examination, X, are distributed such that $X \sim N(85, 49)$. All marks are integers.

(a) If 15% of the students failed to reach the pass mark, what is the pass mark?

(b) The test was out of 100. What percentage of students obtained full marks, correct to one decimal place?

(c) Students with a score within 1.5 standard deviations above the mean received a Credit. What is the range of scores that received a Credit?

47 Given $X \sim N(50, 16)$, then $P(X > 48)$ is closest to:

A 0.3085 **B** 0.5498 **C** 0.4502 **D** 0.6915

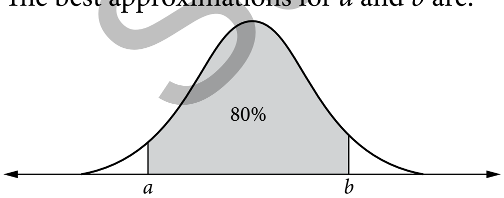
48 Given $X \sim N(50, 16)$, then $P(X < 60)$ is closest to:

A 0.9938 **B** 0.7333 **C** 0.2667 **D** 0.0062

49 Given $X \sim N(50, 16)$, then $P(43 < X < 61)$ is closest to:

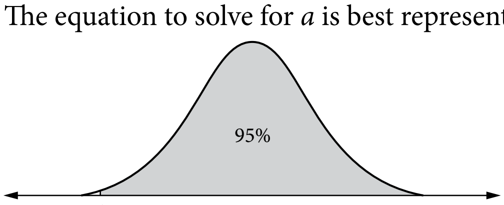
A 0.0431 **B** 0.9569 **C** 0.4236 **D** 0.5764

50 The graph is for the distribution $X \sim N(40, 13)$.
The best approximations for a and b are:



A $a = 29.06, b = 50.95$
B $a = 32, b = 48$
C $a = 35.38, b = 44.62$
D $a = 36.97, b = 40.03$

51 The graph is for the distribution $X \sim N(\mu, \sigma^2)$.
The equation to solve for a is best represented by:



A $a = \mu - 1.645\sigma^2$
B $a = \mu - 1.645\sigma$
C $a = \mu + 1.645\sigma^2$
D $a = \mu - 1.96\sigma$

52 The mass, X grams, of a packet of potato crisps is such that $X \sim N(150, 40)$. A packet with mass less than 143 g is rejected as being underweight.

(a) What proportion of packets, correct to four decimal places, is rejected as being underweight?

Packets with a mass of more than 162 g are rejected as being overweight.

(b) What proportion of packets, correct to four decimal places, are rejected for being overweight?

(c) From **(a)** and **(b)**, what is the probability of rejection, correct to 4 decimal places?

(d) Between which masses, correct to 1 decimal place, would you expect to find the middle 90% of production?

53 Topstuff Bakery makes Christmas cakes in varying stated sizes: 500 g, 1 kg and 2 kg. The machinery is such that the masses of the cakes are normally distributed with a mean which can be set on the machine and a standard deviation of 8 g.

(a) What should the mean be set at so that 95% of the cakes are above each of the advertised sizes? (3 answers needed.)

(b) Within what range would you expect to find the middle 70% of cakes labelled as 1 kg?

(c) In a batch of five hundred 2 kg cakes, how many would you expect to have a mass greater than 2.02 kg?