NEW FORMAT

AJ Sadler

Mathematics Methods

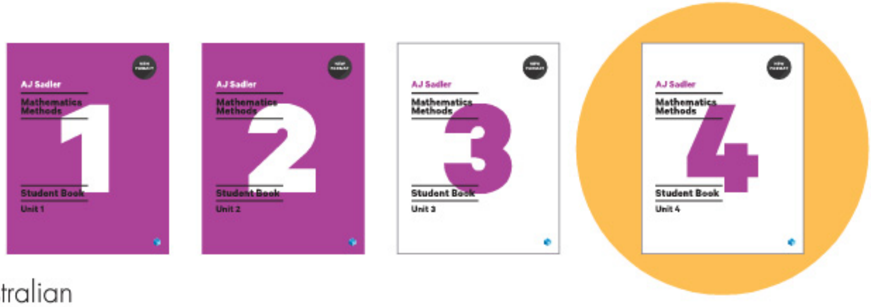
4

Student Book

Unit 4

PREFACE

This text targets Unit Four of the West Australian course *Mathematics Methods*, a course that is organised into four units altogether, the first two for year eleven and the last two for year twelve.



This West Australian course, *Mathematics Methods*, is based on the Australian Curriculum Senior Secondary course *Mathematical Methods*. Apart from small changes to wording, the unit fours of these courses are closely aligned. Hence this book would also be suitable for students following unit four of the Australian Curriculum course *Mathematical Methods*.

The book contains text, examples and exercises containing many carefully graded questions. A student who studies the appropriate text and relevant examples should make good progress with the exercise that follows.

The book commences with a section entitled **Preliminary work**. This section briefly outlines work of particular relevance to this unit that students should either already have some familiarity with from the mathematics studied in earlier years, or for which the brief outline included in the section may be sufficient to bring the understanding of the concept up to the necessary level.

As students progress through the book they will encounter questions involving this preliminary work in the **Miscellaneous exercises** that feature at the end of each chapter. These miscellaneous exercises also include questions involving work from preceding chapters to encourage the continual revision needed throughout the unit.

Some chapters commence with a '**Situation**' or two for students to consider, either individually or as a group. In this way students are encouraged to think and discuss a situation, which they are able to tackle using their existing knowledge, but which acts as a forerunner and stimulus for the ideas that follow. Students should be encouraged to discuss their solutions and answers to these situations and perhaps to present their method of solution to others. For this reason answers to these situations are generally not included in the book.

At times in this series of books I have found it appropriate to go a little outside the confines of the syllabus for the unit involved. In this regard readers will find that in this text that when considering sampling I include mention of 'capture – recapture' as an example of sampling, a technique not specifically mentioned in the syllabus, and when considering random sampling I found it appropriate to consider a few simulation activities. When introducing the idea of an interval estimate of a population proportion I also consider the point estimate.

Alan Sadler

ISBN 9780170395205

CONTENTS

PRELIMINARY WORK vii

1

LOGARITHMIC FUNCTIONS 2

2

CALCULUS INVOLVING LOGARITHMIC FUNCTIONS 28

3

CONTINUOUS RANDOM VARIABLES 42

ISBN 9780170395205

IMPORTANT NOTE

This series of texts has been written based on my interpretation of the appropriate *Mathematics Methods* syllabus documents as they stand at the time of writing. It is likely that as time progresses some points of interpretation will become clarified and perhaps even some changes could be made to the original syllabus. I urge teachers of the *Mathematics Methods* course, and students following the course, to check with the appropriate curriculum authority to make themselves aware of the latest version of the syllabus current at the time they are studying the course.

Acknowledgements

As with all of my previous books I am again indebted to my wife, Rosemary, for her assistance, encouragement and help at every stage.

To my three beautiful daughters, Rosalyn, Jennifer and Donelle, thank you for the continued understanding you show when I am 'still doing sums' and for the love and belief you show.

To the delightfully supportive team at Cengage – I thank you all.

Alan Sadler

PRELIMINARY WORK

This book assumes that you are already familiar with a number of mathematical ideas from your mathematical studies in earlier years.

This section outlines the ideas which are of particular relevance to Unit Four of the *Mathematics Methods* course and for which some familiarity will be assumed, or for which the brief explanation given here may be sufficient to bring your understanding of the concept up to the necessary level.

Read this 'preliminary work' section and if anything is not familiar to you, and you don't understand the brief mention or explanation given here, you may need to do some further reading to bring your understanding of those concepts up to an appropriate level for this unit. (If you do understand the work but feel somewhat 'rusty' with regards to applying the ideas some of the chapters afford further opportunities for revision as do some of the questions in the miscellaneous exercises at the end of chapters.)

- Chapters in this book will continue some of the topics from this preliminary work by building on the assumed familiarity with the work.
- The miscellaneous exercises that feature at the end of each chapter may include questions requiring an understanding of the topics briefly explained here.

Number

It is assumed that you are familiar with, and competent in the use of, positive and negative numbers, recurring decimals (e.g. 0.666 66…, written $0.\dot{6}$), square roots and cube roots and that you are able to choose levels of accuracy to suit contexts and distinguish between exact values, approximations and estimates.

Numbers expressed with positive, negative and fractional powers should also be familiar to you as should be the following index laws:

$$\begin{array}{rclrclrcl}
a^n \times a^m & = & a^{n+m} & a^n \div a^m & = & a^{n-m} & a^0 & = & 1 \\
a^{-n} & = & \dfrac{1}{a^n} & a^{\frac{1}{n}} & = & \sqrt[n]{a} & (a^n)^m & = & a^{n \times m} \\
(ab)^n & = & a^n \times b^n & \left(\dfrac{a}{b}\right)^n & = & \dfrac{a^n}{b^n} & & &
\end{array}$$

Note: The set of numbers that you are currently familiar with is called the set of **real numbers**. We use the symbol $\mathbb{R}$ for this set.

$\mathbb{R}$ contains many subsets of numbers such as the whole numbers, the integers, the prime numbers etc. (If you are also a student of *Mathematics Specialist* you will have encountered numbers beyond this real system. Such considerations are beyond the scope of this unit.)

The absolute value

The absolute value of a number is the distance on the number line that the number is from the origin. The absolute value of x is written $|x|$ and equals x when x is positive, and equals $-x$ when x is negative. Thus $|3| = 3$, $|-3| = 3$, $|4| = 4$, $|-4| = 4$.

Algebra

It is assumed that you are already familiar with:

manipulating algebraic expressions, in particular,
expanding,
simplifying,
factorising,

and *solving equations*, in particular, solving:
linear equations,
quadratic equations,
simultaneous equations,
exponential equations, e.g., $2^x + 3 = 35$,
trigonometric equations, e.g., $\sin x = 0.5$ for $0 \le x \le 2\pi$.

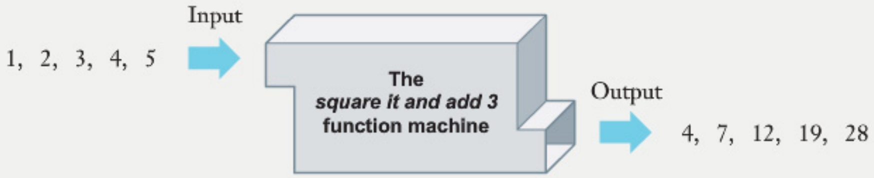
Function

It is assumed that you are familiar with the idea that in mathematics any rule that takes any input value that it can cope with, and assigns to it a particular output value, is called a **function**.

Familiarity with the function notation $f(x)$ is also assumed.

It can be useful at times to consider a function as a machine. A box of numbers (the **domain**) is fed into the machine, a certain rule is applied to each number, and the resulting output forms a new box of numbers, the **range**.

In this way $f(x) = x^2 + 3$, with domain $\{1, 2, 3, 4, 5\}$, could be 'pictured' as follows:



If we are not given a specific domain we assume it to be all the numbers that the function can cope with. This is the function's **natural domain** or **implied domain**.

It is assumed you are particularly familiar with the characteristic equations and graphs of **linear functions**, **quadratic functions** and with the graphs of

$$y = x^3, \qquad y = \sqrt{x} \qquad \text{and} \qquad y = \frac{1}{x},$$

$$y = \sin x, \qquad y = \cos x \qquad \text{and} \qquad y = e^x.$$

It is further assumed that the effect altering the values of a, b, c and d have on the graph of $y = af[b(x - c)] + d$ is something you have previously considered for various functions.

The idea of using the output from one function as the input of a second function should also be familiar to you. In this way we form a **composite function**, also referred to as a **function of a function**.

$$\begin{array}{llcll}
\text{For example, if} & f(x) & = & x^2 & \text{(the } \textit{square it} \text{ function)} \\
\text{and} & g(x) & = & x + 3, & \text{(the } \textit{add three} \text{ function)} \\
\text{then} & f(g(x)) & = & f(x+3) & \text{and} \quad g(f(x)) = g(x^2) \\
 & & = & (x+3)^2 & \qquad\qquad\quad\;\; = x^2 + 3.
\end{array}$$

The exponential function, e^x

In addition to being familiar with the various laws of indices, e.g. $a^n \times a^m = a^{n+m}$, you should also have encountered 'e', which can be defined as $\lim\limits_{n \to \infty} \left[\left(1 + \frac{1}{n} \right)^n \right]$, and be familiar with the function $f(x) = e^x$.

- If $y = e^x$ then $\frac{dy}{dx} = e^x$. The exponential function differentiates to itself!
- The constant e (≈ 2.718 28) allows us to describe many situations involving growth (or decay) mathematically.

  Many growth and decay situations involve some variable, say A, growing, or decaying continuously, according to a rule of the form $A = A_0 e^{kt}$

  where A is the amount present at time t,
  A_0 is the initial amount (i.e. the amount present at $t = 0$),
  and k is some constant dependent on the situation.

Summary statistics

Measures of central tendency

The **mean**, the **median** and the **mode** are all measures used to summarise a set of scores. The mean and the median each indicate a 'central score'. The mode is often included in these 'averages' but there is no guarantee that the mode is any 'central' measure.

Measures of spread (or dispersion)

The **range** of a set of scores is the difference between the highest score and the lowest score and gives a simple measure of how widely the scores are spread. Whilst the range is easy to calculate it is determined using just two of the scores and does not take any of the other scores into account. For this reason it is of limited use.

You should be familiar with **variance** and **standard deviation** as more sophisticated measures of dispersion. The variance is found by finding how much each of the scores differs from the mean, squaring these values and finding the average (mean) of the squared values. The standard deviation is the square root of the variance.

Consider the eight scores listed below, for which the mean is 18.

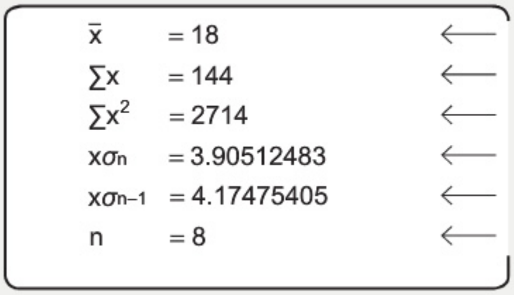
Scores:	12	15	16	16	18	20	22	25
Deviation from mean:	−6	−3	−2	−2	0	+2	+4	+7

$$\text{Variance of scores} = \frac{(-6)^2 + (-3)^2 + (-2)^2 + (-2)^2 + (0)^2 + (2)^2 + (4)^2 + (7)^2}{8}$$

$$= 15.25$$

$$\text{Standard deviation} = \sqrt{15.25} \text{ i.e. } 3.91 \text{ (correct to two decimal places).}$$

You should be able to determine the mean, median, mode, range, standard deviation and variance of a set of scores when the scores are presented in various forms (e.g. as a list, as a frequency table, as a dot frequency graph etc.), using the ability of your calculator to determine these statistical quantities as appropriate.



```
x̄       = 18             ⟵  The mean of the scores.
∑x      = 144            ⟵  The sum of the scores.
∑x²     = 2714           ⟵  The sum of the squares of the scores.
xσn     = 3.90512483     ⟵  The standard deviation of the scores.
xσn–1   = 4.17475405     ⟵  A different standard deviation – see note (2) below.
n       = 8              ⟵  The number of scores.
```

(1) The standard deviation is a measure of spread. For most distributions very few, if any, of the scores would be more than three standard deviations from the mean, i.e. the vast majority of the scores (and probably all of them) would lie between $(\bar{x} - 3\sigma)$ and $(\bar{x} + 3\sigma)$.

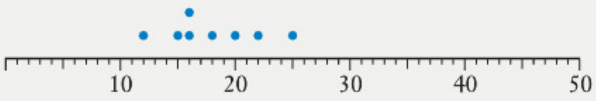
(2) The display shown above has two different standard deviations:
σ_n is the standard deviation of the eight scores.
σ_{n-1} gives an answer a little bigger than σ_n by dividing the sum of the squared deviations by $(n - 1)$ rather than n. This would be used if the eight scores were a sample taken from a larger population and we wanted to use the standard deviation of the sample to estimate the standard deviation of the whole population. Division by $(n - 1)$ rather than n compensates for the fact that there is usually less variation in a small sample than there is in the population itself. If the sample is large, then n will be large and there will be little difference between σ_n and σ_{n-1}.

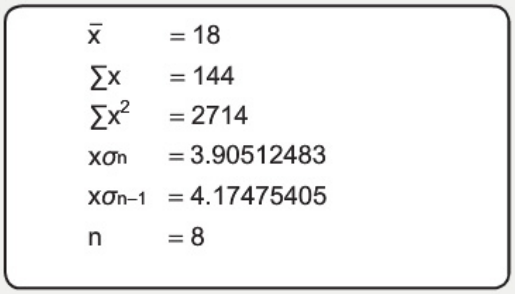
Change of scale and origin

Consider again the set of eight scores:

12 15 16 16 18 20 22 25

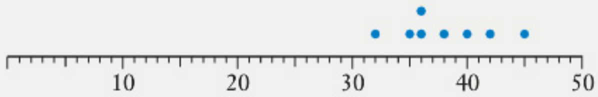
The scores are displayed below left as a dot frequency diagram:





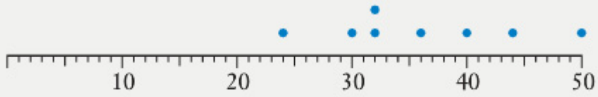
```
x̄       = 18
∑x      = 144
∑x²     = 2714
xσn     = 3.90512483
xσn–1   = 4.17475405
n       = 8
```

Now suppose we increase all of the scores by 20. This will see them all move 20 places to the right on the dot frequency diagram, i.e. a *change of origin*. With all of the scores increased by 20 we would expect the mean to increase by 20. However, the points are no more, or less, spread out, than they were before. Hence the standard deviation should be unchanged.



```
x̄       = 38
∑x      = 304
∑x²     = 11674
xσn     = 3.90512483
xσn–1   = 4.17475405
n       = 8
```

Suppose instead we were to multiply all of the original scores by 2, i.e. a *change of scale*. The scores would again all increase in value but would also become more spread out than the original set. We would expect the mean and the standard deviation of this new set of scores to be twice the mean and standard deviation of the original set.



```
x̄       = 36
∑x      = 288
∑x²     = 10856
xσn     = 7.81024967
xσn–1   = 8.34950811
n       = 8
```

Probability

The probability of something happening is a measure of the likelihood of it happening and is given as a number between zero (no chance of happening) to 1 (certain to happen).

With activities such as rolling a die or flipping a coin, whilst we are unable to consistently predict the outcome of a particular die roll or coin flip, when these activities are repeated a large number of times each has a predictable long-run pattern. For less predictable events the **long-term relative frequency** with which an event occurs is then our best guess at the probability of the event occurring. Probability based on experimental or observed data like this is called **empirical probability**.

You should be familiar with the various 'probability rules' listed on the next page.

For **complementary** events (A and A′):

$$\mathrm{P}(\mathrm{A}') = 1 - \mathrm{P}(\mathrm{A})$$

For **conditional probability** (B | A):

$$\mathrm{P}(\mathrm{B} \mid \mathrm{A}) = \frac{\mathrm{P}(\mathrm{A} \cap \mathrm{B})}{\mathrm{P}(\mathrm{A})}$$

For A **and** B (A ∩ B):

To determine the probability of A **and** B occurring we multiply the probabilities together, paying due regard to whether the occurrence of one of the events affects the likelihood of the other occurring:

$$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B} \mid \mathrm{A})$$

If A and B are **independent** events, P(B | A) = P(B) and so

$$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B})$$

For A **or** B (A ∪ B):

To determine the probability of A **or** B occurring we add the probabilities together and then make the necessary subtraction to compensate for the 'double counting of the overlap':

$$\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B}) - \mathrm{P}(\mathrm{A} \cap \mathrm{B})$$

If A and B are **mutually exclusive** events, P(A ∩ B) = 0 and so

$$\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B})$$

Random variables

Suppose a normal fair coin is flipped 3 times. The 8 equally likely outcomes are:

TTT HTT THT TTH THH HTH HHT HHH

If X represents the number of heads obtained, then X can take the values 0, 1, 2, 3.

$$\underbrace{\text{TTT}}_{X=0} \quad \underbrace{\text{HTT} \quad \text{THT} \quad \text{TTH}}_{X=1} \quad \underbrace{\text{THH} \quad \text{HTH} \quad \text{HHT}}_{X=2} \quad \underbrace{\text{HHH}}_{X=3}$$

The value X takes, 0, 1, 2 or 3, depends upon a random selection process.

We call X a **discrete random variable**.

The word **discrete** means 'separate' or 'individually distinct' which is the case here because X can only take the distinct values 0, 1, 2 or 3.

The possible values of a random variable must be numerical.

Discrete random variables commonly occur when we are *counting* events, for example the number of successes in a number of attempts.

In this unit we will extend our understanding of random variables to consider **continuous random variables**. These commonly occur when we are *measuring* something, for example heights, weights, times etc. The variable is not restricted to certain values but can now take any value (usually within certain limits of reasonableness).

Probability distribution of a discrete random variable

For the random variable X referred to on the previous page, the table below gives the probability associated with each value the variable X can take.

Number of heads (X)	0	1	2	3
Probability	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

We write $\quad \mathrm{P}(X=0)=\frac{1}{8}, \quad \mathrm{P}(X=1)=\frac{3}{8}, \quad \mathrm{P}(X=2)=\frac{3}{8}, \quad \mathrm{P}(X=3)=\frac{1}{8}.$

This table of probabilities completed for the three flips of a coin situation shows how the total probability of 1 is *distributed* amongst the possible values the variable X can take. The table gives the **probability distribution** for the random variable X.

The possible values the random variable can take must together cover all eventualities without overlap. We say they must be **exhaustive** and **mutually exclusive**.

The sum of the probabilities in a probability distribution must be 1.

From our understanding of probability it also follows that for each value of x,

$$0 \quad \leq \quad \mathrm{P}(X=x) \quad \leq \quad 1.$$

For each value the random variable, X, can take, the table assigns the corresponding probability of X taking that value. In mathematics we call a rule or relationship that assigns to each element of one set an element from a second set, a **function**. The pairs of values in the previous table show the **probability function** for the random variable X. We frequently use the notation f(x) to represent a function so we will sometimes use $f(x)$ for $\mathrm{P}(X=x)$.

Mean, variance and standard deviation of a discrete random variable

When working with random variables the mean value is sometimes referred to as the **expected value**. For the random variable X, the mean or expected value is sometimes written as E(X). Do not be misled by the use of the word 'expected'. It is not the value we expect to get with one roll of a die, for example, but is instead the number we expect our long-term average to be close to.

For the probability distribution shown at the top of this page, the mean or expected value

$$\begin{aligned} &= \quad 0 \times \frac{1}{8} \quad + \quad 1 \times \frac{3}{8} \quad + \quad 2 \times \frac{3}{8} \quad + \quad 3 \times \frac{1}{8} \\ &= \quad 1.5 \end{aligned}$$

For the probability distribution shown below:

x	1	2	3	4	5
P($X=x$)	0.1	0.2	0.2	0.4	0.1

$$\begin{aligned} \text{the mean or expected value} \quad &= \quad 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.2 + 4 \times 0.4 + 5 \times 0.1 \\ &= \quad 3.2 \end{aligned}$$

If the discrete random variable, X, has possible values x_i, with $\mathrm{P}(X = x_i) = p_i$ then

$$\mathrm{E}(X) = \Sigma(x_i\, p_i)$$

the summation being carried out over all of the possible values x_i.

If we use the Greek letter, μ, (mu, pronounced myew) to represent $\mathrm{E}(X)$, then the **variance**, sometimes written $\mathrm{Var}(X)$, is equal to $\Sigma[p_i(x_i - \mu)^2]$.

Prior to the ready availability of calculators, applying this formula could be a tedious process, especially if $\mathrm{E}(X)$ was not an integer. In such cases, the alternative formula $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ could be used.

The **standard deviation** is the square root of the variance.

The standard deviation of X is sometimes written $\mathrm{SD}(X)$.

The binomial probability distribution

A trial which can be considered to have just two mutually exclusive outcomes, sometimes referred to as *success* (1) and *failure* (0), is called a **Bernoulli trial**. If the probability of success is p then the long term mean $= p$ and variance $= p(1 - p)$.

If a Bernoulli trial is performed repeatedly, with the probability of success in a trial occurring with constant probability, i.e. the trials are **independent**, the distribution that arises by considering the number of successes is called a **binomial distribution**.

If a Bernoulli trial is performed n times, and the probability of success in each trial is p, the probability of exactly x successes in the n trials is

$$^nC_x\, p^x(1 - p)^{n-x}$$

The number of trials, n, and the probability of success on each trial, p, are called the **parameters** of the distribution. If we know that a random variable is binomially distributed and the parameters n and p are known, the probability distribution can be completely determined.

If the discrete random variable X is binomially distributed with parameters n and p this is sometimes written as:

$$X \sim \mathrm{b}(n, p), \qquad X \sim \mathrm{B}(n, p), \qquad X \sim \mathrm{bin}(n, p) \qquad \text{or} \qquad X \sim \mathrm{Bin}(n, p).$$

For example, suppose that each question of a multiple-choice test paper offers five answers, one of which is correct. If a student answers 7 questions by simply guessing which response is correct each time, and if we define the random variable X as how many of these seven questions the student gets correct, we have:

$$\begin{aligned} &\text{Number of trials} &&= 7,\\ \text{and}\quad &\text{P(success, i.e. gets question correct)} &&= 0.2.\\ \text{Hence}\quad &X &&\sim \mathrm{Bin}(7, 0.2).\\ \text{Thus}\quad &\mathrm{P}(X = 3) &&= {}^7C_3\, 0.2^3\, 0.8^4\\ & &&\approx 0.1147 \end{aligned}$$

For a binomial distribution involving n trials, with p the probability of success on each trial:

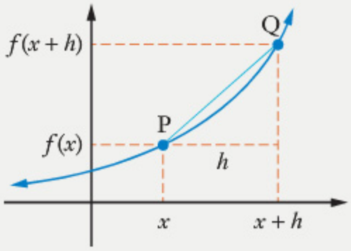
$$\begin{aligned} &\textbf{Mean} &&= \boldsymbol{np}\\ \text{and}\quad &\textbf{Standard deviation} &&= \sqrt{np(1-p)} \text{ or } \sqrt{npq} \end{aligned}$$

where $q = (1 - p)$, the probability of ‘failure’ on each trial.

Calculus

Differentiation

It is assumed that you are familiar with the idea of the **gradient**, or *slope*, of a line and in particular that whilst a straight line has the same gradient everywhere, the gradient of a curve varies as we move along the curve.



To find the gradient at a particular point, P, on a curve $y = f(x)$ we choose some other point, Q, on the curve whose x-coordinate is a little more than that of point P.

Suppose P has an x-coordinate of x and Q has an x-coordinate of $(x + h)$.

The corresponding y-coordinates of P and Q will then be $f(x)$ and $f(x + h)$.

Thus the gradient of PQ $= \dfrac{f(x+h)-f(x)}{h}$.

We then bring Q closer and closer to P, i.e. we allow h to tend to zero, and we determine the limiting value of the gradient of PQ.

i.e. Gradient at P $=$ limit of $\dfrac{f(x+h)-f(x)}{h}$ as h tends to zero.

This gives us the **instantaneous rate of change** of the function at P.

The process of determining the **gradient formula** or **gradient function** of a curve is called **differentiation**.

Writing h, the small increase, or increment, in the x coordinate, as δx, and writing $f(x + h) - f(x)$, the small increment in the y coordinate, as δy, we have:

$$\text{Gradient function} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x}$$

This **derivative** is written as $\dfrac{dy}{dx}$ and pronounced 'dee y by dee x'.

- If $y = f(x)$ then the derivative of y with respect to x can be written as $\dfrac{dy}{dx}$, $\dfrac{df}{dx}$ or $\dfrac{d}{dx}f(x)$.
- A shorthand notation using a 'dash' may be used for differentiation with respect to x. Thus if $y = f(x)$ we can write $\dfrac{dy}{dx}$ as $f'(x)$ or simply y' or f'.
- Whenever we are faced with the task of finding the gradient formula, gradient function, or derivative of some 'new' function, for which we do not already have a rule, we can simply go back to the basic principle:

$$\text{Gradient at } P(x, f(x)) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$$

ISBN 9780170395205

Applying this ‘limiting chord process’ leads to the following results:

If $y = ax^n$ then $\frac{dy}{dx} = anx^{n-1}$.

If $y = e^x$ then $\frac{dy}{dx} = e^x$.

If $y = \sin x$ (for x in radians) then $\frac{dy}{dx} = \cos x$.

If $y = \cos x$ (for x in radians) then $\frac{dy}{dx} = -\sin x$.

These facts, together with the rules that follow, allow us to determine the gradient function for many other functions.

With u and v each functions of x, then:

- If $y = u \pm v$, $\frac{dy}{dx} = \frac{du}{dx} \pm \frac{dv}{dx}$ Sum and difference rules.
- If $y = u \times v$, $\frac{dy}{dx} = v\frac{du}{dx} + u\frac{dv}{dx}$ The product rule.
- If $y = \frac{u}{v}$, $\frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$ The quotient rule.
- If $y = f(u)$ and $u = g(x)$, $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$ The chain rule.
- If $y = [f(x)]^n$, $\frac{dy}{dx} = n[f(x)]^{n-1}f'(x)$ From the chain rule.
- If $y = e^{f(x)}$, $\frac{dy}{dx} = f'(x)e^{f(x)}$ From the chain rule.

Antidifferentiation

You should also be familiar with the idea of **antidifferentiation** which, as its name suggests, is the opposite of differentiation.

E.g. If $\frac{dy}{dx} = ax^n$ then antidifferentiation gives $y = \frac{ax^{n+1}}{n+1} + c$

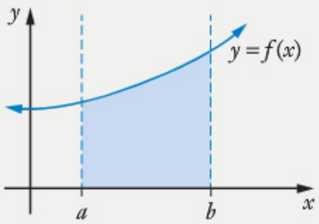
Remembered as: ‘*Increase the power by one and divide by the new power.*’

Antidifferentiation is also known as integration, which uses the symbol $\int$.

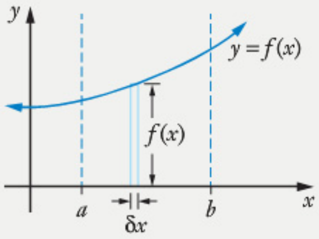
Hence $$\int ax^n dx = \frac{ax^{n+1}}{n+1} + c$$

Limit of a sum

The following should remind you of the idea of finding the area under a curve by summing strips of area, and of the fundamental theorem of calculus.



To determine the area between some function $y = f(x)$ and the x axis, from $x = a$ to $x = b$ (see the diagram on the right) we could divide the area into a large number of equal-width strips, each approximately rectangular, and sum the areas of the strips.



One such rectangular strip, of area $y\ \delta x$, is shown in the second diagram on the right.

The more strips, the smaller δx and the greater our accuracy.

If the exact area of the region is A then
$$A = \lim_{\delta x \to 0} \sum_{x=a}^{x=b} y\ \delta x$$

With a **s**ummation involved we use a 'stretched S' to represent this limit.

We write:
$$\lim_{\delta x \to 0} \sum_{x=a}^{x=b} y\ \delta x = \int_a^b y\, dx$$

A very useful mathematical fact is that this 'limit of a sum' process, called integration, can be determined using antidifferentiation. Indeed this explains why we freely use the same 'stretched S symbol', and the word 'integrate', when we are finding antiderivatives.

Hence to find the area under a curve, the limit of a sum we obtain by considering rectangles can be evaluated using antidifferentiation, a much easier process than summing the areas of many rectangles.

To evaluate $\int_a^b f(x)\, dx$:

(1) Antidifferentiate $f(x)$ with respect to x (and omit the '+ c').

(2) Substitute b into your answer from (1).

(3) Substitute a into your answer from (1).

(4) Calculate: (Part (2) answer) – (Part (3) answer).

In this way, evaluating $\int_a^b f(x)\, dx$ gives a specific answer, without any '+ c' being involved. Integrals of this form are called **definite integrals**.

ISBN 9780170395205

Hence the limit of a sum, which we call a definite integral, can be evaluated using antidifferentiation, because integration and differentiation are opposite processes. This is what the **fundamental theorem of calculus** is all about.

The two boxed results shown below show the opposite nature of this relationship between the definite integral and differentiation. They are the two parts of the **fundamental theorem of calculus**.

$$\int_a^b f'(x)dx = f(b) - f(a) \qquad \text{and} \qquad \frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x)$$

From the rule above left we see that integrating the derivative of a function 'gives us the function back'.

and from the rule above right, differentiating the integral of a function 'gives us the function back'.

Applications of calculus

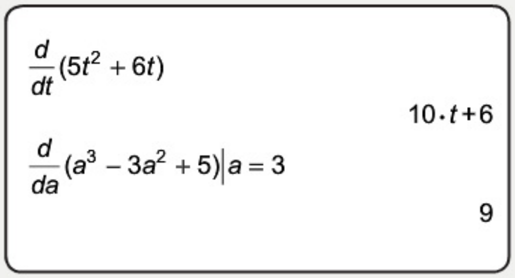
From your previous studies you should be familiar with applying calculus in relation to the following concepts:

Determining the area under a curve and between curves.
Locating turning points and points of inflection.
Optimisation.
Rectilinear motion.
Small changes and marginal rates of change.
Total change from rate of change.

Use of technology

You are encouraged to use your calculator, computer programs and the internet during this unit.

However you should make sure that you can also perform the basic processes such as solving equations, sketching graphs, differentiation, determining definite integrals, without the assistance of such technology when required to do so.

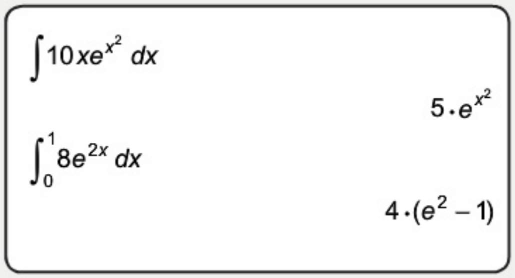


$\frac{d}{dt}(5t^2+6t)$

$10 \cdot t + 6$

$\frac{d}{da}(a^3-3a^2+5)|a=3$

9



$\int 10xe^{x^2}\,dx$

$5 \cdot e^{x^2}$

$\int_0^1 8e^{2x}\,dx$

$4 \cdot (e^2-1)$

Note

The illustrations of calculator displays shown in the book may not exactly match the display from your calculator. The illustrations are not meant to show you exactly what your calculator will necessarily display but are included more to inform you that at that moment the use of a calculator could well be appropriate.

1.

Logarithmic functions

- Logarithms
- Laws of logarithms
- Using logarithms to solve equations
- Natural logarithms
- Logarithmic functions
- Graphs of logarithmic functions
- Logarithmic scale
- Graphs with logarithmic scales
- Use of logarithmic scales
- Miscellaneous exercise one

Situation One

Various estimates could be made for how many years ago the population of the world was just one million. Now it exceeds seven billion, i.e. it now exceeds 7 000 000 000.

Discuss any difficulties a person would face if they were to try to display these world population figures, from one million to seven billion, graphically.

Situation Two

In 1982 an earthquake measuring 6.0 on the Richter scale (a scale for measuring the intensity of an earthquake) occurred in the Yemen, and is thought to have resulted in approximately 2800 deaths.

In 2010 an earthquake measuring 7.0 on the Richter scale occurred in Haiti and, according to some estimates, may have resulted in more than 300 000 deaths.

Now 7.0 is approximately 1.17×6.0 and yet the number of deaths in the Haiti earthquake far exceeds 1.17×2800!

An earthquake in Japan in 2011 measured 9.0 on the Richter scale. Despite this being higher than the Richter scale measurement for the Haiti earthquake the death toll was thought to be about 16 000. Whilst this is a tragically high number of fatalities it is well below the Haiti earthquake death toll, despite the higher rating on the Richter scale.

Discuss the above comments comparing earthquake death tolls and Richter scale readings. Do some research about the Richter scale.

Situation Three

Let us suppose that the number of cells in a colony of bacteria doubles every hour. When timing commences the colony consists of 50 cells. Assuming the doubling continues, how long will it take for the number of cells to reach ten million?

iStock.com/AlexRaths

Situation Four

An object with a temperature of 90°C is placed in an environment with temperature 20°C. The temperature of the object, t minutes later, is T°C, where T approximately follows the mathematical rule

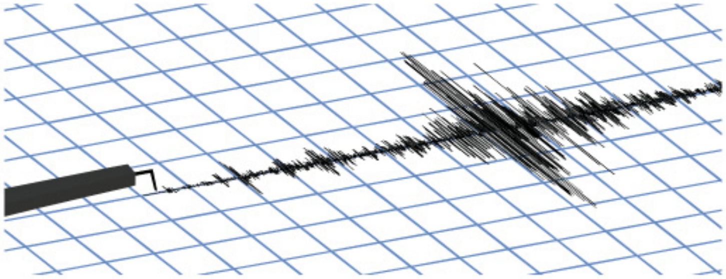
$$T = 20 + 70e^{-0.4t}.$$

After how many minutes will the temperature of the object be 35°C?

WS Logarithms

Logarithms

If you did some research about the Richter scale, as Situation Two on the previous page suggested, you probably discovered that its scale is *logarithmic* rather than linear. However this immediately raises the question: *What does it mean for a scale to be logarithmic?* We will investigate what a logarithm is in this chapter.



Shutterstock.com/Bro Studio

Situation Three, and indeed Situation Four, involved solving an equation in which the unknown value featured as an *index* in the equation.

Situation Three required you to solve $50 \times 2^t = 10\,000\,000.$

Situation Four required you to solve $35 = 20 + 70e^{-0.4t}.$

How did you go about solving these equations?

Let us consider equations of this type further, i.e. equations in which the unknown features as an index.

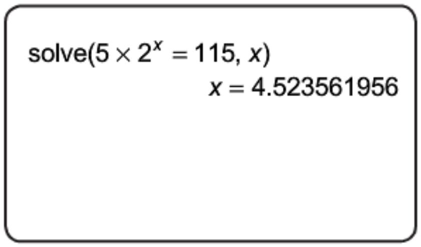
Suppose we have to solve: $5 \times 2^x = 80$
Divide each side of the equation by 5. $2^x = 16$
From our knowledge of the powers of 2: $x = 4$

Suppose instead we were asked to solve: $5 \times 2^x = 115$
Dividing each side of the equation by 5. $2^x = 23$

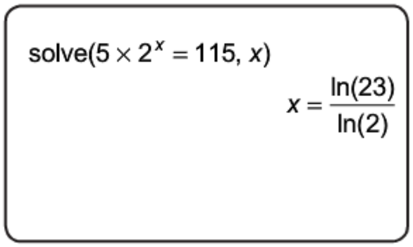
Now our knowledge of the powers of 2 tells us that x must lie between 4 ($2^4 = 16$) and 5 ($2^5 = 32$).

If we want to be more precise than this we could:

- look at the graph of $y = 2^x$ and see where it cuts the line $y = 23$.
- try some values between $x = 4$ and $x = 5$, evaluate 2^x, and adjust our trials accordingly (trial and adjust).
- use the solve facility of some calculators. Note that whilst the display below left gives the answer as 4.523 561 956 the display below right is typical of the response we would get from a calculator that has been asked to give the solution as an exact value. This exact solution is given using *logarithms*, 'ln' being the abbreviation for the 'natural logarithm' of a number.



solve($5 \times 2^x = 115, x$)
$x = 4.523561956$



solve($5 \times 2^x = 115, x$)
$x = \dfrac{\ln(23)}{\ln(2)}$

Let us now consider this idea of the *logarithm* of a number.

Note: The idea of a logarithm of a number proves useful in a number of applications other than simply being able to give the exact solution to some exponential equations, as we shall see as this course continues. Hence their introduction here.

- The number 10, raised to the power 2, is equal to 100, i.e. $10^2 = 100$.
  We say that the logarithm to the **base** 10, of 100 is 2. i.e. $\log_{10} 100 = 2$.

  The logarithm to the base 10, of 100, is the number to which 10 must be raised to get 100, i.e. 2.

- The number 2, raised to the power 3, is equal to 8, i.e. $2^3 = 8$.
  We say that the logarithm to the base 2, of 8, is 3. i.e. $\log_2 8 = 3$.

  The logarithm to the base 2, of 8, is the number to which 2 must be raised to get 8, i.e. 3.

- The number 5, raised to the power –1, is equal to 0.2, i.e. $5^{-1} = 0.2$.
  We say that the logarithm to the base 5, of 0.2, is –1. i.e. $\log_5 0.2 = -1$.

  The logarithm to the base 5, of 0.2, is the number to which 5 must be raised to get 0.2, i.e. –1.

To generalise: For some *positive* number *a*,

$$\boxed{\text{If } a^x = b \text{ then } \log_a b = x.}$$

i.e. The logarithm to the base *a*, of *b*, is the number to which *a* must be raised to give *b*, (*x* in this case).

For example:

$$\log_{10} 1000 = 3$$

(Because 1000 is ten to the power **3**.)

$$\log_{10} 10\,000 = 4$$

(Because 10 000 is ten to the power **4**.)

$$\log_2 32 = 5$$

(Because 32 is two to the power **5**.)

$$\log_2 0.5 = -1$$

(Because 0.5 is two to the power **–1**.)

$$\log_4 16 = 2$$

(Because 16 is four to the power **2**.)

$$\log_4 2 = 0.5$$

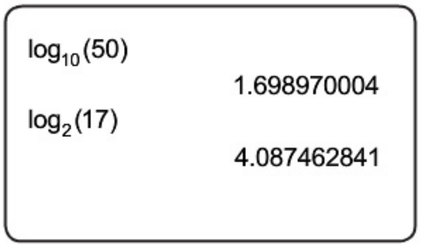
(Because 2 is four to the power **0.5**.)

$\log_{10}(1000)$	3
$\log_{10}(10000)$	4
$\log_2(32)$	5
$\log_2(0.5)$	–1
$\log_4(16)$	2
$\log_4(2)$	0.5

The statement $\log_a b = x$ is the logarithmic equivalent of the
exponential statement $a^x = b$ (which could be written as $a^{\log_a b} = b$).

Note:

- In the previous examples the numbers involved were such that the logarithms could be determined mentally. If this is not the case, for example $\log_{10} 50$ or perhaps $\log_2 17$, some calculators can evaluate such expressions directly.

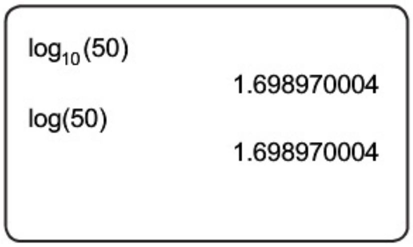


```
log10(50)
                    1.698970004
log2(17)
                    4.087462841
```

- To avoid having to write the base of the logarithm every time, we can omit the '10' for logarithms to base ten.

  Thus $\log 50$, with no specific base indicated, is taken as $\log_{10} 50$.

  Other bases need to be clearly indicated.



```
log10(50)
                    1.698970004
log(50)
                    1.698970004
```

- On the previous page it was stated that:

  For some *positive* number *a*,

$$\text{If } a^x = b \text{ then } \log_a b = x.$$

  If we require *a* to be positive (i.e. we cannot have the logarithm with a negative base) it follows that *b* must also be positive (because if *a* is positive it follows that *a* to some power must also be positive).

  I.e., we cannot determine the logarithm of a negative number.

  (If asked to determine the logarithm of a negative number, say $\log(-3)$, some calculators will indicate 'error' whilst others may give a 'complex' number. Whilst students of *Mathematics Specialist* will be familiar with the idea of complex numbers, they are beyond the scope of this *Mathematics Methods* unit. As far as this unit is concerned we cannot determine the logarithm of a negative number.)

EXAMPLE 1

Without the assistance of a calculator find

a $\log_2 16$ **b** $\log_4 8$ **c** $\log(0.1)$

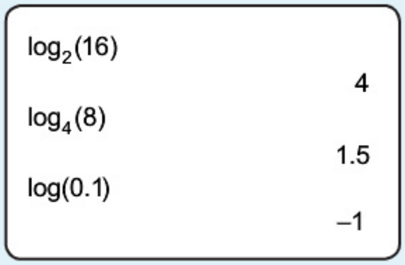
then use a calculator to confirm your answers.

Solution

a
$$\begin{aligned}
\text{Let} \quad \log_2 16 &= x \\
\text{then} \quad 2^x &= 16 \\
\text{i.e.} \quad 2^x &= 2^4 \\
\text{giving} \quad x &= 4 \\
\therefore \quad \log_2 16 &= 4
\end{aligned}$$

b
$$\begin{aligned}
\text{Let} \quad \log_4 8 &= x \\
\text{then} \quad 4^x &= 8 \\
\text{i.e.} \quad 2^{2x} &= 2^3 \\
\text{giving} \quad x &= 1.5 \\
\therefore \quad \log_4 8 &= 1.5
\end{aligned}$$

c
$$\begin{aligned}
\text{Let} \quad \log(0.1) &= x \\
\text{then} \quad 10^x &= 0.1 \\
\text{i.e.} \quad 10^x &= 10^{-1} \\
\text{giving} \quad x &= -1 \\
\therefore \quad \log(0.1) &= -1
\end{aligned}$$



```
log2(16)
                     4
log4(8)
                   1.5
log(0.1)
                    –1
```

Note: (For interest only.) Prior to the ready access of calculators, logarithms were commonly used as an aid to calculation.

Evaluating a product like 216.4 × 171.2 or a quotient like 136.5 ÷ 16.5 now takes only a few seconds but how would you cope with these calculations without a calculator?

In earlier centuries, mathematicians worked to produce tables of logarithms to the base ten. I.e. They produced conversion tables that allowed a number (say 2.884) to be expressed as a power of ten ($\sim 10^{0.46}$).

When two numbers had to be multiplied (or divided) these tables were used to convert each to powers of ten, these powers were then added (or subtracted) and other tables were then used to convert these powers of ten back to give the answer to the question. Though we no longer need to use logarithms in this way the $\boxed{\text{log}}$ button on our calculator does hold the conversion information.

From your calculator $\log_{10} 500 \approx 2.698\,97$ i.e. $10^{2.698\,97} \approx 500$,

$\log_{10} 0.24 \approx -0.6198$ i.e. $10^{-0.6198} \approx 0.24$.

Exercise 1A

Write each of the following as exponential statements.

1 $\log_2 8 = 3$

2 $\log_7 49 = 2$

3 $\log_{49} 7 = 0.5$

4 $\log_{10} 1000 = 3$

5 $\log_5 625 = 4$

6 $\log_4 32 = 2.5$

7 $\log_5 (0.04) = -2$

8 $\log_3 \left(\frac{1}{9}\right) = -2$

9 $\log_a x = y$

10 $\log_b y = c$

11 $\log_x a = p$

12 $\log_a x = 3$

13 $\log_3 5 = y$

14 $\log_2 3 = x$

15 $\log_x 5 = 4$

16 $\log_3 5 = p$

Write each of the following as logarithmic statements.

17 $2^6 = 64$

18 $3^4 = 81$

19 $81 = 9^2$

20 $9^{\frac{3}{2}} = 27$

21 $2^{-1} = 0.5$

22 $2^{-2} = 0.25$

23 $10^2 = 100$

24 $10^{-2} = 0.01$

25 $r = p^q$

26 $r^p = q$

27 $2^x = y$

28 $3^y = z$

29 $5^k = 4$

30 $7^y = 3$

31 $7 = 3^p$

32 $x = e^y$

Without the assistance of a calculator evaluate each of the following.

33 $\log_8 64$

34 $\log_2 128$

35 $\log 10\,000$

36 $\log_3 243$

37 $\log_2 \left(\frac{1}{2}\right)$

38 $\log_2 \left(\frac{1}{16}\right)$

39 $\log_6 \left(\frac{1}{216}\right)$

40 $\log_2 (0.125)$

41 $\log_9 243$

42 $\log (0.001)$

43 $\log_6 6$

44 $\log_7 1$

45 $\log_a 1$

46 $\log_4 32$

47 $\log_a a$

48 $\log_a (a^3)$

Use a calculator to evaluate each of the following, correct to three decimal places if rounding is necessary.

49 $\log 5$

50 $\log 25$

51 $\log 7$

52 $\log 49$

53 $\log 20$

54 $\log(7 + 3)$

55 $\log 7 + \log 3$

56 $\log 50 - \log 5$

57 If $\log_{10} b = c$

- **a** can c be negative?
- **b** can b be negative?

Laws of logarithms

Logarithm laws

Suppose that $\log_a b = x$ and $\log_a c = y$

It then follows that $b = a^x$ and $c = a^y$

But, we know that $a^x a^y = a^{x+y}$ and $a^x \div a^y = a^{x-y}$

i.e. $bc = a^{x+y}$ and $b \div c = a^{x-y}$

Writing these as logarithmic statements:

$$\log_a(bc) = x + y \quad \text{and} \quad \log_a\left(\frac{b}{c}\right) = x - y$$

Thus

$$\boxed{\log_a(bc) = \log_a b + \log_a c} \quad \text{and} \quad \boxed{\log_a\left(\frac{b}{c}\right) = \log_a b - \log_a c}$$

Again suppose $\log_a b = x$ from which it follows that $b = a^x$.

From the index laws we know that $(a^x)^n = a^{xn}$

i.e. $b^n = a^{xn}$

Writing this as a logarithmic statement:

$$\log_a(b^n) = nx$$

$$\boxed{\log_a(b^n) = n\log_a b}$$

Note also that from $a^1 = a$ it follows that $\boxed{\log_a a = 1}$

from $a^0 = 1$ it follows that $\boxed{\log_a 1 = 0}$

and from $\log_a(b^n) = n\log_a b$ it follows that $\boxed{\log_a\left(\frac{1}{b}\right) = -\log_a b}$

ISBN 9780170395205

EXAMPLE 2

Express each of the following as single logarithms.

a $\log x + \log y - 3\log z$

b $\log x + 1 - \log y$

Solution

a
$$\begin{aligned} \log x + \log y - 3\log z &= \log(xy) - \log(z^3) \\ &= \log\left(\frac{xy}{z^3}\right) \end{aligned}$$

b
$$\begin{aligned} \log x + 1 - \log y &= \log x + \log 10 - \log y \\ &= \log\left(\frac{10x}{y}\right) \end{aligned}$$

EXAMPLE 3

Without the assistance of a calculator, simplify $\log_2 12 + \log_2 36 - 3\log_2 3$.

Solution

$$\begin{aligned} \log_2 12 + \log_2 36 - 3\log_2 3 &= \log_2 12 + \log_2 36 - \log_2 3^3 \\ &= \log_2\left(\frac{12 \times 36}{27}\right) \\ &= \log_2 16 \\ &= \log_2(2^4) \\ &= 4\log_2 2 \\ &= 4 \end{aligned}$$

EXAMPLE 4

If $\log_a 4 = p$ and $\log_a 5 = q$, express each of the following in terms of p and q.

a $\log_a 20$

b $\log_a 0.8$

c $\log_a(100a^2)$

Solution

a
$$\begin{aligned} \log_a 20 &= \log_a(4 \times 5) \\ &= \log_a 4 + \log_a 5 \\ &= p + q \end{aligned}$$

b
$$\begin{aligned} \log_a 0.8 &= \log_a(4 \div 5) \\ &= \log_a 4 - \log_a 5 \\ &= p - q \end{aligned}$$

c
$$\begin{aligned} \log_a(100a^2) &= \log_a 100 + \log_a(a^2) \\ &= \log_a(4 \times 25) + 2\log_a a \\ &= \log_a 4 + \log_a 5^2 + 2 \\ &= \log_a 4 + 2\log_a 5 + 2 \\ &= p + 2q + 2 \end{aligned}$$

ISBN 9780170395205

Exercise 1B

Express each of the following as a single logarithm.

1 $\log x + \log z$

2 $2\log x + \log y$

3 $2\log x + 3\log y$

4 $2\log x - \log y$

5 $\log a + \log b - \log c$

6 $3\log a + 4\log b - 2\log c$

7 $5\log c - \log(c^3) + \log a$

8 $2 + \log x$

9 $3 - (\log x + \log y)$

10 $3 - \log x + \log y$

Evaluate each of the following (without the use of a calculator).

11 $\log_2 24 - \log_2 3$

12 $\log_2 20 + \log_2 8 - \log_2 10$

13 $\log(10^4) - \log 10$

14 $\log_a(a^3) + \log_b(b^2) - \log_c(c^4)$

15 $\log_3 45 + 2\log_3 2 - \log_3 20$

16 $\log_3 4 - 2\log_3 6 - 2$

17 $\log 5 - (\log 2 + 2\log 5)$

18 $\log_a b + 2\log_a(ab) - 3\log_a b$

19 $\dfrac{\log_a b^3}{2\log_a b}$

20 $\dfrac{\log_a 48 - \log_a 3}{\log_a 2}$

21 If $\log_a 2 = p$ and $\log_a 3 = q$, express each of the following in terms of p or q or both p and q.

a $\log_a 6$

b $\log_a 18$

c $\log_a 12$

d $\log_a 0.\bar{6}$

e $\log_a(9a^4)$

f $\log a\ 0.\bar{2}$

22 If $\log_5 7 = a$ and $\log_5 2 = b$, express each of the following in terms of a or b or both a and b.

a $\log_5 49$

b $\log_5 28$

c $\log_5 1.75$

d $\log_5 50$

e $\log_5 490$

f $\log_5 700$

Find y in terms of x (and a if necessary) for each of the following.

23 $\log_a y = x$

24 $\log_a y = \log_a(2x)$

25 $\log_a y = 3\log_a x$

26 $2\log_a y = 3\log_a x$

27 $\log_a y = \log_a a + \log_a x$

28 $\log_a y = 2 + \log_a x$

29 $\log_a y = -\log_a x$

30 $\log_a y + \log_a x = 2$

31 An experiment was conducted to test the rate at which students forget work. The students were tested on a particular period in History they had recently studied. They were then given repeat tests of a similar nature after that. In each test the group average was calculated. The results obtained seemed to fit fairly well with the rule:

Average score, t fortnights after the initial test $= 75 - 35\log(t + 1)$

- **a** What was the average score in the initial test?
- **b** What was the average score four weeks after the initial test?
- **c** What was the average score eight weeks after the initial test?
- **d** How many fortnights after the initial test did the average score fall to 40?

32 The Richter scale reading, R, of an earthquake of intensity I is given by

$$R = \log\left(\frac{I}{I_0}\right)$$

where I_0 is a minimum intensity level used for comparison.

- **a** Find R for an earthquake with intensity $1000\ I_0$.
- **b** An earthquake registers 5.4 on the Richter scale. Express its intensity in terms of I_0.
- **c** An earthquake measuring 6 on the Richter scale is how many times as intense as that of one measuring 5 on the Richter scale?
- **d** An earthquake measuring 7.7 on the Richter scale is how many times as intense as that of one measuring 5.9 on the Richter scale?

33 The acidity or alkalinity of a solution is measured by its pH. This is the negative of the logarithm to the base ten of the hydrogen ion concentration in moles per litre.

The letters pH stand for *potential of hydrogen*.

Thus pH $= -\log_{10}$(hydrogen ion concentration in moles per litre).

A pH below 7 indicates that a solution is acidic and a pH above 7 indicates alkaline. Values are usually between 0 and 14.

The pH of natural water is generally about 6 because dissolved carbon dioxide from the air forms carbonic acid.

Shutterstock.com/Deyan Georgiev

Find the pH of each of the following:

a	Grapes.	Hydrogen ion concentration	0.0001 moles/litre.
b	Beer.	Hydrogen ion concentration	0.000 031 6 moles/litre.
c	Urine.	Hydrogen ion concentration	0.000 000 25 moles/litre.
d	Eggs.	Hydrogen ion concentration	0.000 000 016 moles/litre.
e	Blood.	Hydrogen ion concentration	0.000 000 042 moles/litre.

If a solution has a pH of 5.25 what is its hydrogen ion concentration?

ISBN 9780170395205

34 Sound loudness is measured by comparing the intensity of the sound with the intensity of a sound that is just detectable by the human ear.

With L, the loudness of the sound in decibels (dB),
I, the intensity of the sound
and I_0, the intensity of sound just audible to the human ear, then

$$L = 10\log_{10}\left(\frac{I}{I_0}\right)$$

a If the noise level in a room was 40 dB it would be considered quiet.
Express the intensity of sound in this quiet room in terms of I_0.

b If the noise level in a room was 70 dB it would be considered noisy.
Express the intensity of sound in this noisy room in terms of I_0.

c How many times is the intensity of a 90 dB noise level that of the intensity of a 20 dB noise level?

Using logarithms to solve equations

Now that we know what logarithms are, and what rules they obey, can we use them to solve equations like

$$2^x = 23?$$

As mentioned earlier, we already have other methods for solving equations of this type but we include the logarithmic approach here because, as well as being useful functions in their own right, logarithms give us a method that can be quick to apply and that allows us to state an *exact* solution to the equation.

EXAMPLE 5

Use logarithms to solve the following equations, giving **exact** answers involving base ten logarithms.

a $2^x = 23$

b $2^{5x-1} = 3^x$

Solution

a $2^x = 23$

Taking logs of both sides

$$\begin{aligned} \log(2^x) &= \log 23 \\ \therefore\quad x\log 2 &= \log 23 \\ x &= \frac{\log 23}{\log 2} \end{aligned}$$

b $2^{5x-1} = 3^x$

Taking logs of both sides

$$\begin{aligned} \log(2^{5x-1}) &= \log(3^x) \\ (5x-1)\log 2 &= x\log 3 \\ 5x\log 2 - \log 2 &= x\log 3 \\ x(5\log 2 - \log 3) &= \log 2 \\ \therefore\qquad x &= \frac{\log 2}{5\log 2 - \log 3} \end{aligned}$$

Note:

- When 'taking logs of both sides' in the previous example we chose to use base ten logarithms because the question asked us to give our exact answer involving base ten logarithms. Without this requirement we could use any base of logarithm. Indeed a calculator, set the task of finding the exact solutions to the above equations, may well use *natural logarithms*, for which the abbreviation is 'ln'. We will consider the concept of natural logarithms later in this chapter.
- The answer for part **b** of the previous example could be written in a number of different ways, for example:

$$\frac{\log 2}{\log(2^5) - \log 3}, \quad \frac{\log 2}{\log 32 - \log 3}, \quad \frac{\log 2}{\log\left(\frac{32}{3}\right)}, \quad \frac{-\log 2}{\log\left(\frac{3}{32}\right)}, \quad \ldots$$

Hence do not be too quick to mark your answer wrong just because it appears different to the one given in the back of the book.

Exercise 1C

Solve each of the following, giving your answers in **exact** form involving logarithms to the base ten.

1 $3^x = 7$ **2** $7^x = 1000$ **3** $10^x = 27$

4 $2^x = 11$ **5** $3^x = 17$ **6** $7^x = 80$

7 $5^x = 21$ **8** $10^x = 15$ **9** $2^x = 70$

10 $6^{x+2} = 17$ **11** $3^{x+1} = 51$ **12** $8^{x-1} = 7$

13 $5^{x-1} = 3^{2x}$ **14** $2^{x+1} = 3^x$ **15** $4^{3x} = 5^{x+2}$

16 $3^{2x+1} = 2^{3x-1}$ **17** $5(2^x) = 3 - 2^{x+2}$ **18** $5^x + 4\,(5^{x+1}) = 63$

Use the substitution $y = 2^x$ to solve each of the following equations, giving your answers in exact form involving logarithms to the base ten.

19 $(2^x)^2 + 3(2^x) - 18 = 0$ **20** $2^{2x} - 2^{x+3} + 15 = 0$

21 If $x = \log_2 7$ find an exact expression for x involving base ten logarithms.
Hint: Write the equation in exponential form and then take the logarithm of both sides.

22 Without the assistance of your calculator, find exact expressions involving base ten logarithms for each of the following. (Hint: Do question **21** first.)

a $\log_3 5$ **b** $\log_2 12$ **c** $\log_9 15$

d $\log_9 4$ **e** $\log_{2.5}(6.8)$ **f** $\log_{5.4}(9)$

Applied questions

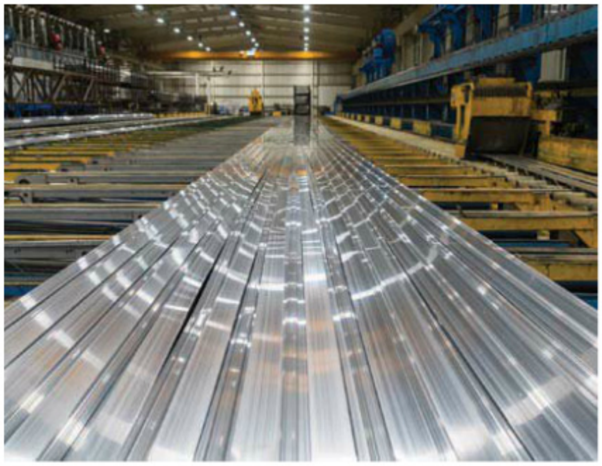
Solve the following questions **without** using the ability of your calculator to solve equations.

23 A strip of metal is to be made into a thin sheet by repeatedly passing it through a pair of compression rollers. The thickness, T, of the metal after it has passed through the rollers n times is given by

$$T = T_0\,(0.92)^n,$$

where T_0 is the initial thickness of the metal before any rolling has been done.

How many times should the metal be passed through the rollers if we require the final thickness to be as close as possible to, but *thinner* than, 20% of the initial thickness?



iStock.com/ozgurdonmaz

24 A group of 200 insects were monitored in a laboratory experiment and the population was found to grow such that the number, N, present t days after the experiment commenced, approximately fitted the model

$$N = 200\,(2.7)^{0.1t}.$$

Find **a** the number of insects in the group 3 days after the experiment commenced,

b the number of insects in the group 5 days after the experiment commenced,

c on which day the population first exceeded 1000.

25 A driver with a high blood alcohol level is more likely to have an accident than is a driver with a low, or zero, blood alcohol level. If R% is the likelihood or risk of an accident, and a% is the blood alcohol level then let us suppose that the rule

$$R = (2.8)^{20a},$$

for $a \geq 0$, is a reasonable mathematical description of what seems to be the case.

For what value of a, the percentage blood alcohol level, is the risk of an accident 51%?

26 A company expects the weekly sales of a particular chocolate bar to increase from the usual 100 000 bars to 250 000 bars whilst their new advertising campaign is running. However, market research indicates that t weeks after the campaign finishes the weekly figures will have fallen to N bars, where

$$N = 100\,000 + 150\,000\,(1.1)^{-0.8t}.$$

If this predicted model proves to be correct, what will be the weekly sales figures

a 4 weeks after the campaign ceases?

b 8 weeks after the campaign ceases?

The company plans to repeat the campaign when sales fall to 135 000 bars per week. Approximately how many weeks after the first campaign ceases will this happen?

iStock.com/harmpeti

27 If \$10 000 is invested at an interest rate of 8% per annum, compounded annually, the investment will grow to \$$P$ after x years where

$$P = 10\,000(1.08)^x\ .$$

a Find P after 3 years.

b Find P after 7 years.

c How long will it take (to the nearest year) for the investment to grow to \$50 000?

d How long will it take (to the nearest year) for the investment to grow to \$50 000 if

i the interest throughout is 10% rather than 8%?

ii the interest is 14% for the first 8 years and 10% thereafter?

e Find the annual interest rate necessary for the \$10 000 to double in value in 5 years. (Give your answer as a percentage, correct to one decimal place.)

Natural logarithms

Natural logarithms, a term mentioned a few pages earlier, are logarithms to the base 'e'. With e being a naturally occurring base in exponential equations describing growth and decay situations, it follows that it could well be a useful base to use with logarithms.

$$\text{If} \qquad b = e^x \qquad \text{then} \qquad x = \log_e b.$$

Note:
- We call logarithms to the base e *natural logarithms*.
- $\log_e x$ can be written as $\ln x$. (In this text we will use both forms.)
- Some calculators have two built-in logarithm options: $\boxed{\log}$ for $\log_{10} x$ and $\boxed{\ln}$ for $\log_e x$.

  Use your calculator to confirm that:

  $$\ln 1 = 0, \qquad \ln 2 \approx 0.693, \qquad \ln 10.5 \approx 2.351, \qquad \ln 2.718\,28 \approx 1.$$

- Whilst e and 10 are the more common bases for logarithms (indeed base ten logarithms are sometimes referred to as *common logarithms*), we have already seen that other bases are possible. In some cases logarithms to other bases may be readily evaluated, for example,

  $$\log_2 8 = 3 \qquad \text{and} \qquad \log_5 25 = 2.$$

  In other cases, for example,

  $$\log_2 7 \qquad \text{or} \qquad \log_5 0.6,$$

  we might use the ability of some calculators to evaluate the logarithm.

  $$\log_2 7 \approx 2.807, \qquad \log_5 0.6 \approx -0.317.$$

ISBN 9780170395205

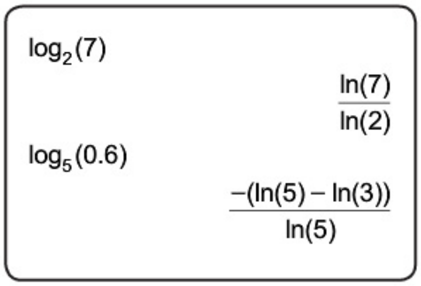
- It is possible to express a logarithm in one base as an expression involving logarithms to another base. Indeed question 21 of the previous exercise asked you to do this when it asked you to express $\log_2 7$ in terms of base ten logarithms.

  Applying the technique suggested in that question to the general case gives us a *change of base formula*, as shown below.

  $$\begin{array}{llrcl} \text{If} & x = \log_a b \quad \text{then} & a^x & = & b \\ \text{Taking logarithms} & & \log_c(a^x) & = & \log_c b \\ \therefore & & x\log_c a & = & \log_c b \\ \text{Giving} & & x & = & \dfrac{\log_c b}{\log_c a} \end{array}$$

  Hence $\boxed{\log_a b = \dfrac{\log_c b}{\log_c a}}$ the **change of base formula**.

  Some calculators, if working in exact mode, will, when given the logarithm in a base other than 10 or e, display the answer in terms of natural logarithms.



$\log_2(7)$

$\dfrac{\ln(7)}{\ln(2)}$

$\log_5(0.6)$

$\dfrac{-(\ln(5)-\ln(3))}{\ln(5)}$

EXAMPLE 6

Use logarithms to solve the equation $e^{x+1} = 5$ giving your answer

a exactly,

and **b** correct to four decimal places.

Solution

a With the equation involving e it makes sense to use natural logarithms rather than logarithms to base ten.

$$\begin{array}{lrcll} & e^{x+1} & = & 5 & \\ \therefore & (x+1)\log_e e & = & \log_e 5 & \\ \text{Thus} & x+1 & = & \log_e 5 & \\ & x & = & \ln 5 - 1 & \text{is the exact solution.} \end{array}$$

b $\therefore$ $x = 0.6094$ is the solution correct to 4 decimal places.

The reader should confirm that 0.6094 is also obtained if base ten logarithms are used instead of natural logarithms.

Exercise 1D

Evaluate each of the following without the use of a calculator.

1 $\log_e e$

2 $\log_e\left(\frac{1}{e}\right)$

3 $\log_e(e^3)$

4 $\log_e \sqrt{e}$

5 $\ln \sqrt[3]{e}$

6 $\ln\left(\frac{1}{\sqrt{e}}\right)$

7 $\ln(e^{-3})$

8 $\ln\left(\frac{1}{\sqrt[3]{e}}\right)$

Clearly showing your use of natural logarithms, solve each of the following equations giving your answers as exact values.

9 $e^{x+1} = 7$

10 $e^{x+3} = 50$

11 $e^{x-3} = 100$

12 $e^{2x+1} = 15$

13 $5e^{3x-1} = 3000$

14 $4e^{x+2} + 3e^{x+2} = 7000$

15 $e^{2x} - 30e^x = -200$ (Hint: let $y = e^x$.)

Express each of the following in terms of natural logarithms and prime numbers (without the assistance of a calculator).

16 $\log_7 2$

17 $\log_2 21$

18 $\log_3 200$

19 $\log_5 50$

20 $\log_6 9$

21 $\log_9 6$

22 $\log_4 300$

23 $\log_8 220$

24 If $A = 2000e^{-t}$ find an exact expression for t in terms of A and evaluate it, correct to three decimal places, for **a** $A = 1500$,

b $A = 500$,

c $A = 50$.

25 The population of a particular country was thought to be 22 300 000 in 2010.

Figures suggest that the population is growing such that, t years after 2010, it will be approximately P, where $P = 22\,300\,000e^{0.02t}$.

If the population growth continues as suggested, in which year would the population of this country reach **a** 32 000 000?

b 45 000 000?

26 A certain culture of bacteria grows in such a way that t days after observation commences the number of bacteria present, N, is given by:

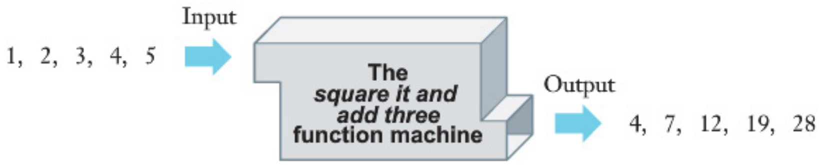
$$N \approx 5000e^{0.55t}.$$

According to this rule how many days after observation commences, to the nearest day, would the number of bacteria be **a** 80 thousand?

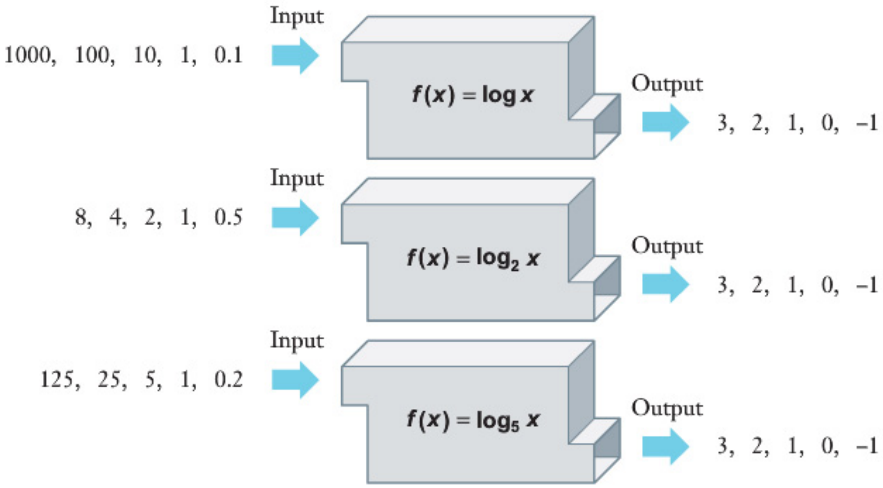
b 750 thousand?

Logarithmic functions

The *Preliminary work* mentioned that it can be helpful to view a function as a machine with a specific output for each given input:



Applying this idea to the concept of logarithms:



Graphs of logarithmic functions

What do the graphs of $y = \log x$, $y = \ln x$, $y = \log_5 x$, $y = \log_2 x$, etc., look like?

As the *Preliminary work* mentioned, it is anticipated that you are familiar with how the graph of $y = af[b(x - c)] + d$, differs from that of $y = f(x)$.

Use your calculator to view the graphs of $y = \log_a x$ for various values of $a > 0$.

WS Plotting log functions

In particular, starting with $y = f(x)$:

- Multiplying the right hand side of the equation by 'a' stretches (dilates) the graph parallel to the y-axis with scale factor 'a'. If 'a' is negative the graph is also reflected in the x-axis.
- Replacing x by bx dilates the graph parallel to the x-axis with a scale factor of $\frac{1}{b}$.
- Replacing x by $x - c$ translates the graph c units to the right. (If c is negative the translation is to the left).
- Adding 'd' to the right hand side of the equation translates the graph d units vertically upwards. (If d is negative the translation is vertically downwards.)

INVESTIGATE

Are these same effects evident when we consider the graphs of logarithmic functions?

ISBN 9780170395205

Exercise 1E

1 Determine the coordinates of the point where the graph of $y = \log_2(x + 8)$ cuts

a the x-axis, **b** the y-axis.

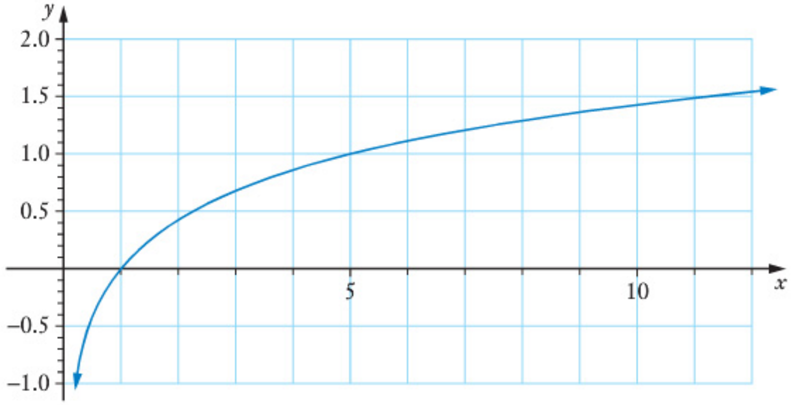
2 What are the coordinates of the point that is common to *all* graphs of the form $y = \log_p x$?

3 Find the coordinates of the point where the graph of $y = \log_a x$ cuts the line $y = 1$.

4 What is the equation of the vertical asymptote of the graph of

a $y = \log_p x$? **b** $y = \log_p(x - 3)$? **c** $y = \log_p x - 3$?

5 The graph below shows $y = \log_5 x$.



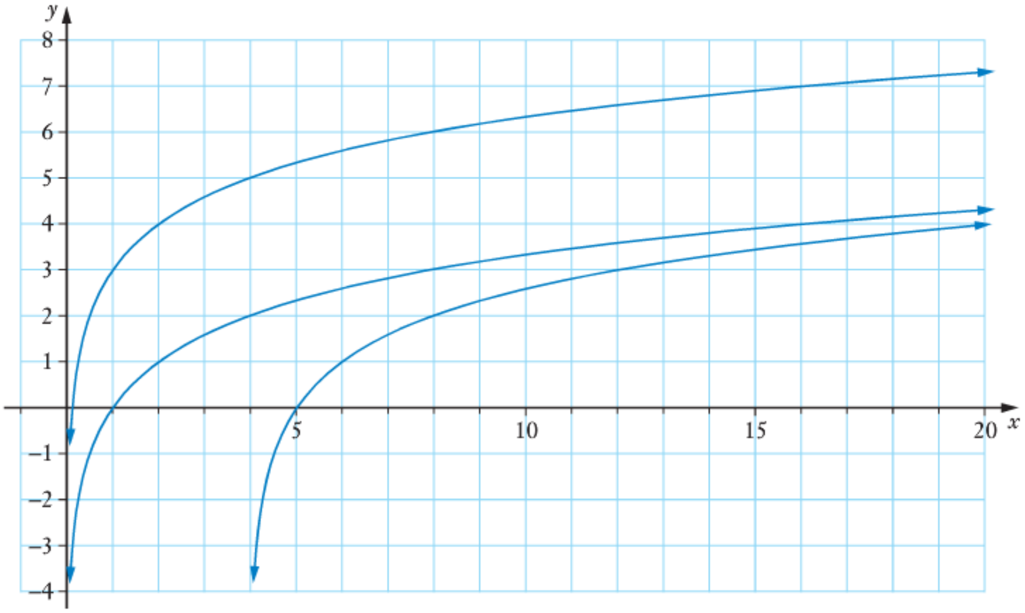
Use the graph to determine approximate solutions to each of the following.

a $\log_5 x = 0.5$, **b** $\log_5 x = 1.5$, **c** $x - 5^{0.8} = 0$, **d** $\log_5(x - 1) = 1.3$.

e Now solve each of the equations algebraically, with the assistance of your calculator to evaluate powers, giving answers rounded to 3 decimal places.

6 The graph below shows $y = \log_a x$, $y = \log_a(x - b)$, and $y = \log_a x + c$.

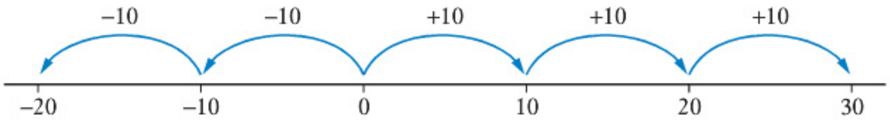
Determine the values of a, b and c given that they are all positive integers.



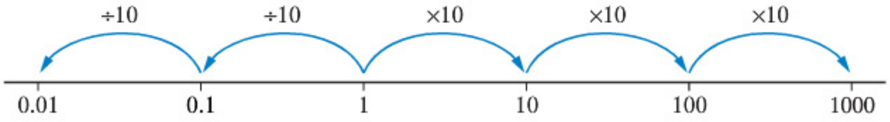
Logarithmic scale

The number line below shows a linear scale.

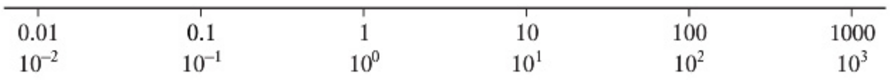
Suppose we start at a particular number location on this line. If moving a particular distance to the right (or left), increases (or decreases) the number we are located at by, say, 10, then on this linear scale all such movements of this size will increase (or decrease) the number we are located at by 10.



However, on a logarithmic scale, if moving a particular distance to the right (or left) multiplies (or divides) the number we are located at by, say, 10, then all such movements to the right (or left) will multiply (or divide) by 10.



In this way, in a logarithmic scale, the distance between consecutive powers of ten is constant.

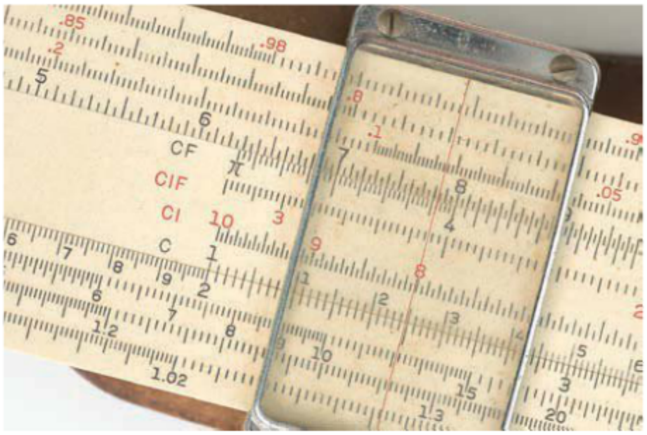


Notice that the logarithmic scale shown above displays the numbers 1 to 1000 in the space that the linear scale at the top of the page displayed just zero to 30. This ability to display a greater range in the same space is one feature that makes logarithmic scales useful. Consider again Situation One encountered at the beginning of this chapter, for example. It would be difficult on a linear scale to show both the comparatively small world population of one million and the much larger current population of more than seven billion. Use of a logarithmic scale may solve this problem.

Before the ready availability of electronic calculators a device called a slide rule was a helpful aid when performing calculations. The slide rule was marked using a logarithmic scale rather than a linear scale. By placing two such scales with the same base together, and adding length a to length b, the fact that

$$\log_k a + \log_k b = \log_k (ab)$$

means that the combined length gives the product $a \times b$.



Shutterstock.com/nattul

Graphs with logarithmic scales

Some graph paper have a logarithmic scale on one axis (log-linear graph paper) or on both axes (log-log paper).

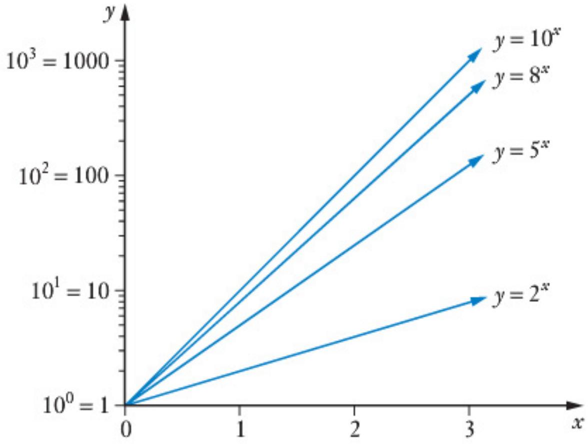
The log-linear graph on the right has a logarithmic scale on the y-axis.

On this graph, functions with equations of the form

$$y = a^x$$

will appear as straight lines

(as indeed will all functions of the form $y = ka^{bx}$, for k, a and b constants).



Use of logarithmic scales

As mentioned on the previous page, if we wish to display data that has a large range, a logarithmic scale can be useful.

We have also already seen in this chapter that this 'multiplication effect' of a logarithmic scale is used in the Richter scale of earthquake intensity, in the modelling of memory activity, in measuring the acidity or alkalinity of solutions (the pH value) and in sound measurement (decibels). Three of these applications are mentioned again below and also mention is made of the use of a logarithmic scale in musical scales.

The Richter scale

A seismograph is an instrument that measures vibrations from an earthquake graphically. The base ten logarithm of the amplitude of these measurements (corrected for the distance the seismograph is from the earthquake epicentre) gives the strength of the earthquake on the Richter scale.

The use of base ten logarithms means that for each unit increase on the Richter scale, the amplitude of the vibrations is multiplied by ten.

pH scale

The pH scale (potential of Hydrogen) is a measure of the acidity or alkalinity of a solution. This is the negative of the logarithm to the base ten of the hydrogen ion concentration in moles per litre.

A pH of 7 is regarded as neutral. Pure water is neutral, it is neither acidic nor alkaline. The pH of pure water is a reference point for acidity and alkalinity. A pH above 7 indicates a solution is alkaline, below 7 indicates the solution is acidic.

A solution with a pH of 3 is ten times as acidic as a solution with a pH of 4.

A solution with a pH of 10 is one hundred times as alkaline as a solution with a pH of 8.

Scale of loudness

The decibel (dB) scale measures loudness and is based on multiples of ten. Hence this too is a logarithmic scale using base ten logarithms.

Shutterstock.com/Chilika77

Music scale

If one musical note has frequency f and another has frequency $2f$ the frequency ratio is said to be one octave. Thus, each time the frequency doubles we go up one octave. This use of powers again means that a logarithmic scale is used.

To determine how many doublings are involved in a change from a frequency f_1 to f_2 we solve

$$\frac{f_2}{f_1} = 2^x$$

Hence $$\log\left(\frac{f_2}{f_1}\right) = x\log 2$$ giving $$x \approx 3.32\log\left(\frac{f_2}{f_1}\right).$$

Exercise 1F

1 A particular scale measures N as a function of L according to the rule

$$N = -\log_{10}(2L).$$

Find **a** N when $L = 3.2 \times 10^{-8}$,

b L when $N = 9.5$.

2 If x octaves are involved between a note of frequency f_1 hertz (Hz) and one of f_2 Hz then

$$x = \frac{1}{\log 2} \times \log\left(\frac{f_2}{f_1}\right).$$

a How many octaves are there between a frequency of 20 Hz to one of 50 Hz?

b If something has a frequency range of 3 octaves, and the lower frequency is f_1, express the higher frequency in terms of f_1.

3 The pH of a solution is defined as

$$\text{pH} = -\log(\text{H}^+),$$

where H^+ is the hydrogen ion concentration in moles per litre.

a Find H^+ for pure water, pH = 7.

b Find the pH for lemon juice, $\text{H}^+ = 0.01$ moles per litre.

4 The 'logit' function (pronounced *lowjit*) is used in some branches of probability and statistics. If p is the probability of an event occurring then

$$\operatorname{logit}(p) = \ln\left(\frac{p}{1-p}\right).$$

a If $p = 0.2$ find $\operatorname{logit}(p)$ giving your answer correct to two decimal places.

b If $\operatorname{logit}(p) = 4$ find p giving your answer correct to two decimal places.

c If an event has a probability of occurring of p what is the significance of $\operatorname{logit}(p)$ being negative?

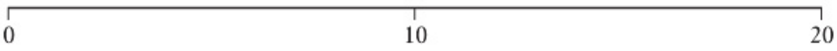
d If, for real x and real k, $\ln\left(\frac{x}{1-x}\right) = k$, show that x will always be between zero and one, whatever the value of k.

5 Comment on the following statement:

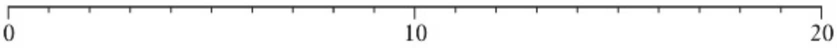
The cost of the damage caused
by an earthquake of Richter scale 7
is ten times that of one with Richter scale 6.

Shutterstock.com/Darrenp

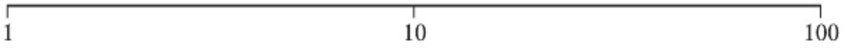
6 With the linear scale shown below indicating 0, 10 and 20,



we know where to mark 1, 2, 3, etc.:



However, with the logarithmic scale below, indicating 1, 10 and 100,



where would we mark 2, 4, 5, 20, 30, 50?

Try to draw such a logarithmic scale yourself with these numbers appropriately placed. (Use 5 cm for the distance from 1 to 10, and the distance from 10 to 100, i.e. use 5 cm to represent 'multiplication by 10'.)

(As mentioned on page 21, before the ready availability of electronic calculators a device called a slide rule used such a scale and was helpful in performing calculations.)

Miscellaneous exercise one

WS Logarithms review

This miscellaneous exercise may include questions involving the work of this chapter and the ideas mentioned in the Preliminary work section at the beginning of the book.

Differentiate the following with respect to x.

1 $5x^3$

2 $x(x^2 + 1)$

3 $\dfrac{x-3}{2x+5}$

4 $(x^3 + 1)^4$

5 e^x

6 $2e^x$

7 $10e^x$

8 $e^x + 3x^2 + x^3$

9 e^{5x}

10 $3e^{4x}$

11 $3e^{2x}$

12 $2e^{3x} + 3e^{2x}$

Write each of the following as exponential statements.

13 $\log_3 81 = 4$

14 $\log_6 216 = 3$

15 $\log_2 (0.25) = -2$

16 $\log_a b = c$

17 $\log_a c = b$

18 $\log_b a = c$

19 $\log_c a = b$

20 $\log_x 2 = 5$

Write each of the following as logarithmic statements.

21 $2^3 = 8$

22 $25 = 5^2$

23 $4^{-1} = 0.25$

24 $2^{-3} = 0.125$

25 $7^x = y$

26 $a^2 = p$

27 $10^y = z$

28 $x = e^y$

Evaluate each of the following (without the assistance of a calculator).

29 $\log_2 32$

30 $\log_5 125$

31 $\log_{10} 10$

32 $\log 1000$

33 $5 + \ln e$

34 $4 - \ln e^2$

35 $6\ln\sqrt{e}$

36 $\log_2 8+\ln\left(\frac{1}{e}\right)$

37 $\log_a 1$

38 $\log_a a$

39 $\log_a(a^3)$

40 $\log_a\sqrt{a}$

Use natural logarithms to solve each of the following equations, giving exact answers.

41 $e^{x+1}=12$

42 $e^{x+2}=25$

43 $e^{x-1}=150$

44 $e^{2x+1}=34$

45 $5e^{x+1}+3e^{x+1}=200$

46 $e^{2x}-12e^{x}=-35$

Express each of the following as a single logarithm.

47 $3\log x+\log y$

48 $2\log x-3\log y$

49 $2\log a+\log b-3\log c$

50 $3+\log x$

51 $2+\ln x$

52 $3-\ln x+2\ln y$

53 A particular company required P tonnes of fossil fuel in 2010. Figures suggest that this annual requirement is increasing in such a way that t years after 2010 the company will require $Pe^{0.1t}$ tonnes. If this suggested rule is correct, by what year will the requirement for fossil fuel for this company be approximately five times its 2010 requirement?

Shutterstock.com/cornfield

54 A body is initially at rest at an origin, O. It then moves in a straight line such that its acceleration, t seconds later, is $0.1e^{0.1t}$ m/s^2.

a Find the velocity of the body when $t=10$.

b Find the displacement of the body from O when $t=10$.

c Find a formula involving T for the distance the body travels from $t=T$ to $t=T+1$.

Use your formula from **c** to determine, correct to 3 decimal places, the distance the body moves in

d the third second,

e the tenth second.

ISBN 9780170395205

2.

Calculus involving logarithmic functions

- Differentiating $y = \ln x$
- Integration to give logarithmic functions
- Miscellaneous exercise two

The *Preliminary work* section at the beginning of this book reminded us:

> Whenever we are faced with the task of finding the gradient formula, gradient function, or derivative of some 'new' function, for which we do not already have a rule, we can simply go back to the basic principle:

$$\text{Gradient at } \mathrm{P}(x, f(x)) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$$

Thus if $y = \ln x$ then $\dfrac{dy}{dx} = \lim_{h\to 0}\dfrac{\ln(x+h)-\ln(x)}{h}$.

Let us explore this limit for some values of x.

Suppose $x = 2$.

h	$\dfrac{\ln(2+h)-\ln(2)}{h}$
1	0.405 465 (6 decimal places)
0.1	0.487 902 (6 decimal places)
0.01	0.498 754 (6 decimal places)
0.001	0.499 875 (6 decimal places)
0.000 1	0.499 988 (6 decimal places)
0.000 01	0.499 999 (6 decimal places)

The table suggests that at $x = 2$, $y = \ln x$ has a gradient of 0.5, i.e. $\frac{1}{2}$.

Complete similar tables for $y = \ln x$ at $x = 5$, and at $x = 10$, and use your tables to suggest the gradient of $y = \ln x$ for these values of x.

iStock.com/f8grapher

WS Differentiating exponential and logarithmic functions

Differentiating $y = \ln x$

Did your results for the gradient of $y = \ln x$ at $x = 2$, 5 and 10 suggest that:

$$\text{If} \qquad y = \ln x$$

$$\text{then} \qquad \frac{dy}{dx} = \frac{1}{x}?$$

Rather than considering $\lim\limits_{h \to 0} \dfrac{\ln(x+h) - \ln(x)}{h}$ for other values of x we can confirm the result suggested above using the fact that $\dfrac{dy}{dx} = \dfrac{1}{\left(\dfrac{dx}{dy}\right)}$. (A statement that is justified at the bottom of this page.)

If $\quad y = \log_e x \quad$ then $\quad x = e^y$.

From this it follows that $\quad \dfrac{dx}{dy} = e^y$.

Thus $\quad \dfrac{dy}{dx} = \dfrac{1}{e^y} = \dfrac{1}{x}$.

$$\text{If } y = \log_e x \text{ then } \frac{dy}{dx} = \frac{1}{x}.$$

Justification of the fact that: $\dfrac{dy}{dx} = \dfrac{1}{\left(\dfrac{dx}{dy}\right)}$.

We cannot simply assume this result to be true by the rules of fractions because $\dfrac{dy}{dx}$ is not a fraction (it is the limit of a fraction). Instead the result can be justified as follows:

Using the chain rule: $\quad \dfrac{dz}{dy}\dfrac{dy}{dx} = \dfrac{dz}{dx} \qquad$ ← equation [1]

Now suppose that $z = x$. Differentiation gives $\dfrac{dz}{dx} = 1$.

Thus equation [1] becomes $\quad \dfrac{dx}{dy}\dfrac{dy}{dx} = 1 \quad$ and so $\quad \dfrac{dy}{dx} = \dfrac{1}{\left(\dfrac{dx}{dy}\right)} \quad$ as required.

ISBN 9780170395205

EXAMPLE 1

Differentiate **a** $3x^4 + \log_e x$ **b** $3\ln x$

Solution

a If $y = 3x^4 + \log_e x$

$$\frac{dy}{dx} = 12x^3 + \frac{1}{x}$$

b If $y = 3\ln x$

$$\begin{aligned}\frac{dy}{dx} &= 3\left(\frac{1}{x}\right)\\ &= \frac{3}{x}\end{aligned}$$

EXAMPLE 2

Differentiate $\log_e(3x^2 + 5x)$.

Solution

If $y = \log_e(3x^2 + 5x)$

Let $u = 3x^2 + 5x$ then $y = \log_e u$.

$\dfrac{du}{dx} = 6x + 5$ and $\dfrac{dy}{du} = \dfrac{1}{u}$.

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du}\frac{du}{dx} \qquad \text{(Chain rule)}\\ &= \frac{1}{u} \times (6x + 5)\\ &= \frac{6x + 5}{3x^2 + 5x}\end{aligned}$$

The general statement of the above example is:

If $y = \log_e f(x)$ then, by the chain rule, $\dfrac{dy}{dx} = \dfrac{f'(x)}{f(x)}$

EXAMPLE 3

Differentiate **a** $\log_e(3x + 5)$ **b** $\log_e(x^2 + 5)$

Solution

a If $y = \log_e(3x + 5)$

$$\frac{dy}{dx} = \frac{3}{3x + 5}$$

b If $y = \log_e(x^2 + 5)$

$$\frac{dy}{dx} = \frac{2x}{x^2 + 5}$$

EXAMPLE 4

Differentiate $\log_e [(x+3)(x+4)]$.

Solution

If
$$\begin{aligned} y &= \log_e [(x+3)(x+4)] \\ &= \log_e (x^2+7x+12) \\ \frac{dy}{dx} &= \frac{2x+7}{(x+3)(x+4)} \end{aligned}$$

The reader should confirm that if the above question was first written as

$$\log_e (x+3) + \log_e (x+4)$$

and then differentiated the same answer would result.

Exercise 2A

Differentiate each of the following with respect to x. **For some it may be advisable to use the laws of logarithms *before* differentiating.**

1 $\log_e x$

2 $\log_e 2x$

3 $5x^2 + \log_e x$

4 $x + e^x + \log_e x$

5 $\log_e (3x+2)$

6 $\ln (2x+3)$

7 $\ln (2x-3)$

8 $\ln (x^2+1)$

9 $\ln (\cos x)$

10 $\log_e (x^2)$

11 $\log_e (\sqrt[3]{x})$

12 $\log_e (3\sqrt{x})$

13 $\log_e \left(\frac{x}{5}\right)$

14 $\log_e [x(x+3)]$

15 $\log_e [(x+4)(x-3)]$

16 $x \log_e x$

17 $(\log_e x)^3$

18 $\log_e \left(\frac{1}{x}\right)$

19 $\frac{1}{\log_e x}$

20 $e^x \log_e x$

21 $\frac{\log_e x}{x}$

22 $(1 + \log_e x)^3$

23 $\log_e [x(x+5)(x+3)]$

24 $\log_e \left[\frac{x+1}{x+3}\right]$

25 $\log_e [(x^2+5)^4]$

26 $\log_e \left[\frac{x}{x^2-1}\right]$

27 $\log_e \left[\frac{(x+2)^3}{x-2}\right]$

Find the gradient of each of the following curves at the given point on the curve.

28 $y = 7\log_e x$ at $(1, 0)$.

29 $y = x\log_e x$ at $(e^2, 2e^2)$.

30 $y = 3x^2 + \log_e x$ at $(1, 3)$.

31 $y = \dfrac{-2\log_e x}{x}$ at $(1, 0)$.

Find the exact coordinates of the point(s) on the following curves where the gradient is as stated.

32 $y = \ln x$ gradient, 0.25.

33 $y = \ln(x^2)$, gradient 4.

34 $y = \ln(6x - 5)$, gradient 0.24.

35 $y = \ln[x(x + 3)]$, gradient 0.5.

Find the equation of the tangent to the given curve at the indicated point.

36 $y = \log_e x$ at the point $(1, 0)$.

37 $y = \log_e x$ at the point $(e, 1)$.

Remembering from chapter one that $\log_a b = \dfrac{\log_c b}{\log_c a}$ differentiate each of the following.

38 $y = \log_4 x$.

39 $y = \log_6 x$.

40 If $y = 50\ln x$, use calculus to determine the approximate small change in y when x changes from 10 to 10.1.

Check your answer by evaluating $(50\ln 10.1 - 50\ln 10)$, correct to four decimal places.

41 An object moves along a straight line such that its displacement, x metres, from an origin O, at time t seconds, is given by

$$x = t + \ln t$$

Find the velocity and acceleration of the object when $t = 2$.

42 Use calculus to determine the nature and coordinates of any turning points on the graph of

$$y = x^2 - 50\ln 2x,\ x > 0.$$

Integration of reciprocals

Integration to give logarithmic functions

If $y = \ln x$ then $\dfrac{dy}{dx} = \dfrac{1}{x}$.

Thus $\displaystyle\int \frac{1}{x}\,dx = \ln x + c.$

If $y = \ln f(x)$ then $\dfrac{dy}{dx} = \dfrac{f'(x)}{f(x)}$.

Thus $\displaystyle\int \frac{f'(x)}{f(x)}\,dx = \ln f(x) + c.$

Any algebraic fraction for which the numerator is the derivative of the denominator will integrate to give a natural logarithmic function.

Note:

- With $\ln x$ only defined for $x > 0$ this unit will only consider:

  $\displaystyle\int \frac{1}{x}\,dx$ for $x > 0$, and $\displaystyle\int \frac{f'(x)}{f(x)}\,dx$ for $f(x) > 0$.

  Thus, for this unit:

  $\displaystyle\int \frac{1}{x}\,dx = \ln x + c,$ for $x > 0.$

  $\displaystyle\int \frac{f'(x)}{f(x)}\,dx = \ln f(x) + c,$ for $f(x) > 0.$

- Although it is beyond the requirements of this unit, suppose we were asked to determine $\displaystyle\int \frac{1}{x}\,dx$ for $x < 0$.

  Writing the answer as $\ln x + c$ would present a problem because we would then be faced with the logarithm of a negative number.

  However, this situation is avoidable if, for $x < 0$, we were to write $\displaystyle\int \frac{1}{x}\,dx$ as $\displaystyle\int \frac{-1}{-x}\,dx$, for which the answer is $\ln(-x) + c$.

  Thus we could say that for $x > 0$, $\displaystyle\int \frac{1}{x}\,dx = \ln x + c,$

  and for $x < 0$, $\displaystyle\int \frac{1}{x}\,dx = \int \frac{-1}{-x}\,dx = \ln(-x) + c.$

  Combining these two statements using the absolute value gives

  $$\int \frac{1}{x}\,dx = \ln|x| + c. \qquad x \neq 0.$$

  This is mentioned here to explain why your calculator may, when asked to determine $\displaystyle\int \frac{1}{x}\,dx$, display an answer that includes the absolute value.

ISBN 9780170395205

In the next two examples, two methods of solution are shown.
In one method the approach is to make an intelligent first attempt at the antiderivative, differentiate it, and then use the result to adjust the first attempt appropriately. If we are attempting to antidifferentiate an expression that is of the form

$$\frac{f'(x)}{f(x)}$$

or some scalar multiple thereof, our initial attempt should be of the form

$$\ln f(x).$$

In 'method two' the given expression is first manipulated so that the task becomes that of determining

$$a\int \frac{f'(x)}{f(x)}\,dx$$

from which the answer, $a\ln f(x) + c$, follows.

The reader should be able to follow both methods but is advised to adopt whichever one they prefer.

EXAMPLE 5

Find $\int \frac{5}{2x}\,dx$ (for $x > 0$).

Solution

Intelligent guess

Try $\quad y = \ln 2x$

Then
$$\begin{aligned} \frac{dy}{dx} &= \frac{2}{2x} \\ &= \frac{1}{x} \end{aligned}$$

Our initial trial needs to be multiplied by $\frac{5}{2}$.

$$\therefore \qquad \int \frac{5}{2x}\,dx = \frac{5}{2}\ln 2x + c.$$

Rearrange

$$\begin{aligned} \int \frac{5}{2x}\,dx &= \frac{5}{2}\int \frac{1}{x}\,dx \\ &= \frac{5}{2}\ln x + c. \end{aligned}$$

The answers for the two methods used above may appear different but in fact they are different ways of writing the same thing:

$$\begin{aligned} \frac{5}{2}\ln 2x + c &= \frac{5}{2}[\ln 2 + \ln x] + c \\ &= \text{a constant} + \frac{5}{2}\ln x + c \\ &= \frac{5}{2}\ln x + \text{a constant}. \end{aligned}$$

EXAMPLE 6

Find $\int \frac{10x}{x^2+1}\,dx$.

Solution

Noticing that the numerator is a multiple of the derivative of the denominator we either make an intelligent guess and then adjust, or rearrange.

Intelligent guess

$$y = \ln(x^2+1).$$

Then $$\frac{dy}{dx} = \frac{2x}{x^2+1}$$

Our initial trial needs to be multiplied by 5.

$$\therefore \quad \int \frac{10x}{x^2+1}\,dx = 5\ln(x^2+1)+c.$$

Rearrange

$$\begin{aligned}\int \frac{10x}{x^2+1}\,dx &= 5 \times \int \frac{2x}{x^2+1}\,dx \\ &= 5\ln(x^2+1)+c.\end{aligned}$$

With practice, the integrals can be written directly.

EXAMPLE 7

Find **a** $\int \frac{16}{2x+5}\,dx \quad (2x+5>0)$ **b** $\int \frac{15x^2}{x^3+1}\,dx \quad (x^3+1>0)$.

Solution

a $\int \frac{16}{2x+5}\,dx = 8\ln(2x+5)+c$

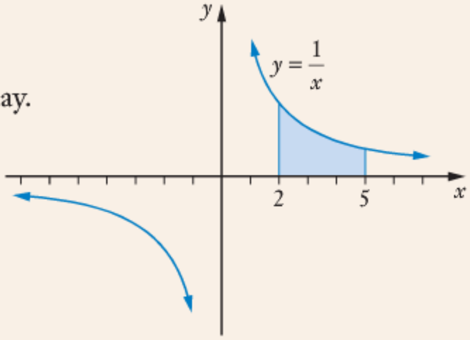
b $\int \frac{15x^2}{x^3+1}\,dx = 5\ln(x^3+1)+c$

EXAMPLE 8

Find the area between the x-axis and $y = \frac{1}{x}$ from $x = 2$ to $x = 5$.

Solution

First make a sketch or view the situation on a calculator display.



$$\begin{aligned}\text{Required area} &= \int_2^5 \frac{1}{x}\,dx \\ &= [\ln x]_2^5 \\ &= \ln 5 - \ln 2 \\ &= \ln 2.5\end{aligned}$$

Exercise 2B

Find the following indefinite integrals. (Assume denominators are greater than zero.)

1 $\displaystyle\int \frac{5}{x}\,dx$

2 $\displaystyle\int \frac{4}{x}\,dx$

3 $\displaystyle\int \left(x+\frac{2}{x}\right) dx$

4 $\displaystyle\int \frac{1}{2x}\,dx$

5 $\displaystyle\int \frac{2x}{x^2+1}\,dx$

6 $\displaystyle\int \left(x^2+\frac{5}{x}\right) dx$

7 $\displaystyle\int \left(4x+e^x+\frac{2}{x}\right) dx$

8 $\displaystyle\int \frac{2}{x+1}\,dx$

9 $\displaystyle\int \frac{8x}{x^2-3}\,dx$

10 $\displaystyle\int \frac{5}{5x-3}\,dx$

11 $\displaystyle\int \frac{10}{2x+1}\,dx$

12 $\displaystyle\int \frac{6x}{x^2+1}\,dx$

13 $\displaystyle\int \frac{2x+1}{x^2+x+3}\,dx$

14 $\displaystyle\int \frac{6x+15}{x^2+5x}\,dx$

15 $\displaystyle\int \frac{20x}{x^2+4}\,dx$

16 $\displaystyle\int \frac{\sin x}{\cos x}\,dx$

17 $\displaystyle\int \frac{\cos x}{\sin x}\,dx$

18 $\displaystyle\int \frac{\sin 2x}{\cos 2x}\,dx$

19 $\displaystyle\int \tan x\,dx$

20 $\displaystyle\int \tan 5x\,dx$

21 $\displaystyle\int 6\tan 2x\,dx$

22 $\displaystyle\int \frac{\sin x-\cos x}{\sin x+\cos x}\,dx$

23 $\displaystyle\int \frac{2+\cos 2x}{4x+\sin 2x}\,dx$

24 $\displaystyle\int \frac{e^x+1}{e^x+x}\,dx$

Evaluate the following definite integrals, giving **exact** answers.

25 $\displaystyle\int_1^3 \frac{1}{x}\,dx$

26 $\displaystyle\int_2^3 \frac{3}{x}\,dx$

27 $\displaystyle\int_1^2 \left(e^x+\frac{1}{x}\right) dx$

28 At time t seconds, $t \geq 0$, a body moving in a straight line has a displacement from an origin O of x metres and a velocity of v metres/second where

$$v=\frac{1}{t+2}.$$

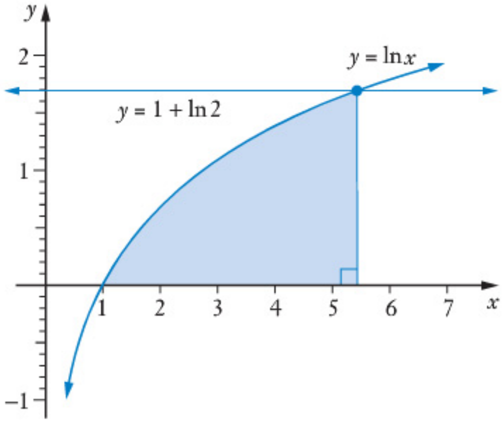
If, when $t=0$, $x=0$, determine an expression for x in terms of t.

29 Find exactly, the area between $y = \dfrac{2x+1}{x}$ and the x-axis from $x = 1$ to $x = 3$.

30 Find exactly, the area enclosed by $y = \dfrac{1}{x+2} - 1$ and the axes.

31 Determine $\dfrac{dy}{dx}$ for $y = x\ln x - x$.

Hence, without the assistance of your calculator, determine the shaded area in the diagram shown on the right, giving your answer as an exact value.



32 Find the area between $y = \tan x$ and the x-axis from $x = 0$ to $x = \dfrac{\pi}{6}$, giving your answer in exact form.

33 Find the constants a and b given that for $\{x \in \mathrm{R}: x \neq -4, x \neq -2\}$

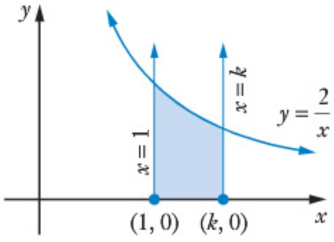
$$\frac{a}{x+4} + \frac{b}{x+2} = \frac{2(4x+13)}{(x+4)(x+2)}.$$

Hence find an expression for $\displaystyle\int \frac{2(4x+13)}{(x+4)(x+2)}\,dx \ (x > -2)$.

34 **a** Find k exactly ($k > 1$), given that the region shown shaded in the diagram has an area of 1 square unit.

b If the line $x = b$ divides the shaded region in the diagram into two regions each of area 0.5 square units, find b.

c If $(c, 0)$ is midway between $(1, 0)$ and $(k, 0)$ find the exact area between $y = \dfrac{2}{x}$ and the x–axis from $x = 1$ to $x = c$.



Miscellaneous exercise two

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

Find $\dfrac{dy}{dx}$ for each of the following.

1 $y = \sin 2x$

2 $y = \cos 3x$

3 $y = e^{4x}$

4 $y = 5e^{4x}$

5 $y = \dfrac{2x-3}{x+1}$

6 $y = (3x-1)^4$

7 $y = 1 + 2\log_e x$

8 $y = x^2 \ln x$

9 $y = \dfrac{1}{x} + 3e^{2x}$

10 $y = \log_e(1 + x + x^2)$

11 Solve $2^x = 11$ giving the *exact* answer using base ten logarithms.

12 If $\log_a 5 = p$ and $\log_a 4 = q$, express each of the following in terms of p or q or both p and q:

a $\log_a 25$

b $\log_a 500$

c $\log_a 80$

d $\log_a 10$

e $\log_a(20a^3)$

f $\log_5 4$

13 Without the assistance of a calculator, determine the value of x ($x > 0$) in each of the following statements.

a $\log_x 64 = 3$

b $\log_x 64 = 2$

c $\log_x 64 = 6$

d $\log_{10} 100 = x$

e $\log 17 - \log 2 = \log x$

f $\log 17 + \log 2 = \log x$

g $\log\sqrt{2} = x\log 2$

h $3\log 2 = \log x$

14 Find an expression for p in each of the following:

a $\log_a x + \log_a y = \log_a p$

b $\log_x p = y$

c $3\log_a x - \log_a y = \log_a p$

d $2 + 0.5\log_{10} y = \log_{10} p$

15 Find the equation of the tangent to $y = \ln x$ at the point $(e^2, 2)$.

16 A pump is used to extract air from a steel container. Each minute, the pump reduces the amount of air in the container by 12% of what it was at the beginning of that minute.

Write down an expression for Q, the quantity of air in the container after t minutes pumping, in terms of Q_0, the initial quantity present.

For how long must the pump work if we require just 5% of the original amount to remain?

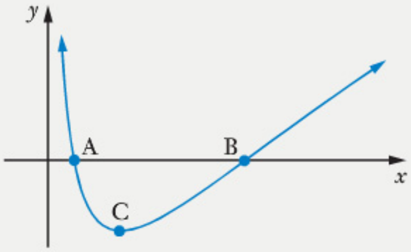
17 Given that $f'(x) = 3x^2 \ln(3x + 2)$ determine

a $f''(x)$,

b $f''(1)$.

18 The graph on the right shows part of the curve

$$y = (\log_e x)^2 - 1.$$



Answer the following without the use of a graphic calculator.

a Determine the exact coordinates of points A and B, the x-axis intercepts, and prove that there are no other places where this function cuts either axis.

b Determine the exact coordinates of point C, the local minimum, and prove that this function has no other stationary points.

c Determine whether or not $y = (\log_e x)^2 - 1$ has any points of inflection and if so determine their location.

19 Use the first principles definition:

$$\frac{d}{dx} f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

to prove that $\quad \dfrac{d}{dx} \sin x = \cos x \quad$ and $\quad \dfrac{d}{dx} \cos x = -\sin x.$

You may assume the following:

- $\displaystyle \lim_{h \to 0} \frac{\sin h}{h} = 1$
- $\displaystyle \lim_{h \to 0} \frac{1 - \cos h}{h} = 0$
- $\displaystyle \lim_{h \to 0} (f(h) \pm g(h)) = \lim_{h \to 0} f(h) \pm \lim_{h \to 0} g(h).$

ISBN 9780170395205

3.

Continuous random variables

- Distribution as a histogram
- Continuous random variables
- Probability density function (pdf)
- Uniform (or rectangular) distributions
- Non-uniform distributions
- Expected value, variance and standard deviation
- Change of scale and origin
- Cumulative distribution function
- Miscellaneous exercise three

Situation

Suppose that at a particular time the people living in Australia, and aged less than 100, had the following age distribution:

$0 \le \text{age} < 50$:	15 613 000 people.
$50 \le \text{age} < 100$:	7 517 000 people.
Total:	23 130 000 people.

1 If one of these people is selected at random what is the probability of the person being

a under 50? **b** from 50 to not yet 100?

Suppose we have more detailed information about the ages of these people:

$0 \le \text{age} < 25$:	7 481 000 people.
$25 \le \text{age} < 50$:	8 132 000 people.
$50 \le \text{age} < 75$:	6 046 000 people.
$75 \le \text{age} < 100$:	1 471 000 people.
Total:	23 130 000 people.

2 If one of these people is selected at random what is the probability of the person being

a under 25? **b** from 75 to not yet 100?

If one of the people who is such that $50 \le \text{age} < 100$ is chosen at random what is the probability that their age is such that

c $50 \le \text{age} < 75$? **d** $75 \le \text{age} < 100$?

Even more detailed information is given below:

$0 \le \text{age} < 10$:	2 973 000 people.
$10 \le \text{age} < 20$:	2 866 000 people.
$20 \le \text{age} < 30$:	3 370 000 people.
$30 \le \text{age} < 40$:	3 210 000 people.
$40 \le \text{age} < 50$:	3 194 000 people.
$50 \le \text{age} < 60$:	2 942 000 people.
$60 \le \text{age} < 70$:	2 322 000 people.
$70 \le \text{age} < 80$:	1 371 000 people.
$80 \le \text{age} < 90$:	734 000 people.
$90 \le \text{age} < 100$:	148 000 people.
Total:	23 130 000 people.

3 If one of these people is selected at random what is the probability of the person being

a under 30? **b** from 40 to not yet 100?

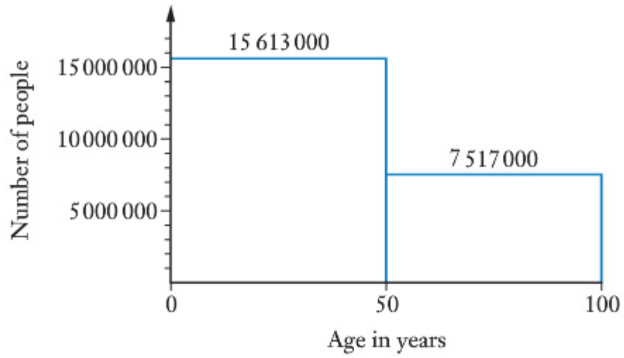
If one of the people who is less than 80 years of age is chosen at random what is the probability of the person being

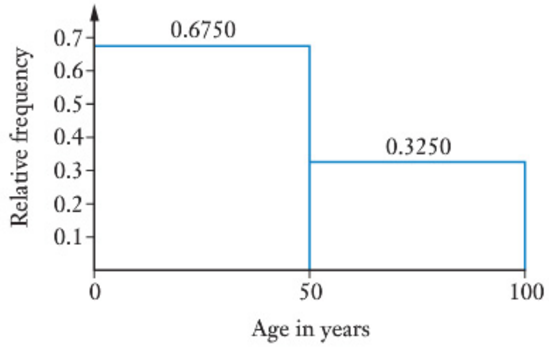
c 60 or more?

(Based on information from the Australian Bureau of Statistics.)

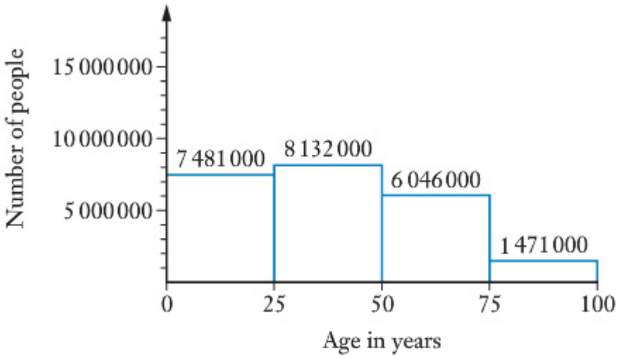
Distribution as a histogram

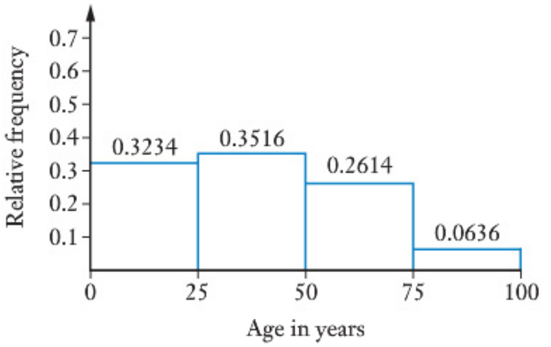
The initial information given in the situation on the previous page could be presented as a histogram showing the real data, below left, or showing the relative frequencies, below right.

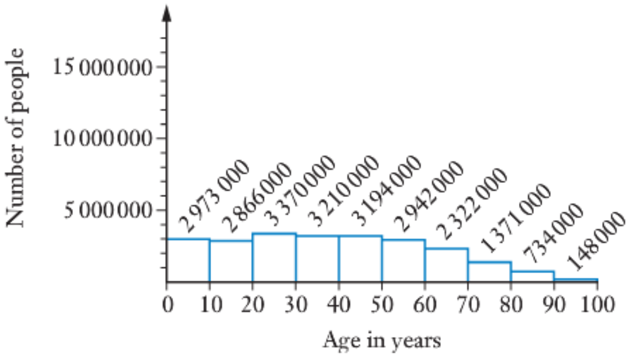


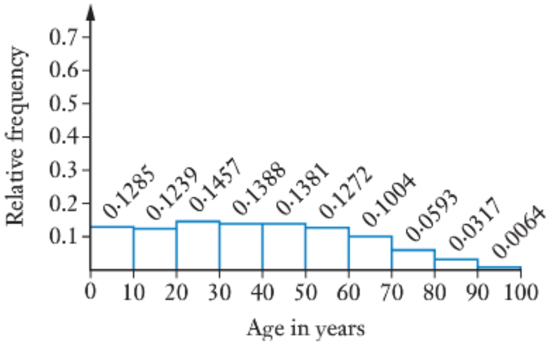


The more detailed information can also be displayed in this way, we simply have more columns:









The data forms columns on the histogram because the data was presented in 'blocks' or 'bins', e.g. under 25, 25 to 49, 50 to 74. The histogram can become smoother the more blocks it is divided into (i.e. as we reduce the 'bin width').

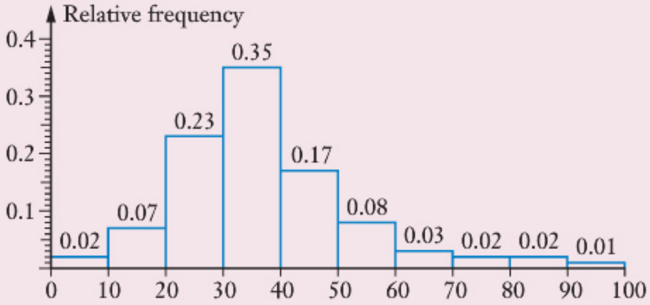
For each of the questions on the previous page you needed to determine the probability of something occurring by using relative frequency. The relative frequency histograms shown above give us the probability of a randomly selected individual being in a particular age range. Showing how the total probability of 1 is distributed across the possible outcomes is something you are familiar with from your work on *discrete random variables*, a concept the *Preliminary work* reminded you of. The difference here is that we have a *continuous* random variable – a person's age. Age does not have to take distinct, separate values. We may talk of someone being 18 years old but in reality this means that if their age is x years then $18 \le x < 19$.

EXAMPLE 1

Scientific research into a particular species of animal collects 100 of the animals, from newborn to fully grown, records various information about them, and then releases them back into the wild. One piece of recorded information, the weight of each animal, gave rise to the table shown below.

Weight (kg)	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Number of animals	2	7	23	35	17	8	3	2	2	1

This table gave the following frequency histogram showing relative frequencies.



If X is the weight, in kg, of a randomly selected animal, from the 100 collected, determine:

a $P(50 \le X < 60)$

b $P(50 \le X < 90)$

c $P(X = 45)$

d $P(X \ge 90 | X > 50)$

Note: Because X is a *continuous* variable we are not restricted to particular values, as we would be with a discrete variable. Between zero and ten there are, in theory, an infinite number of values that X can take. Hence the probability of any particular value is negligible. I.e. $P(X = k) = 0$. Thus there is no difference between $P(50 \le X \le 60)$, $P(50 \le X < 60)$, $P(50 < X \le 60)$ and $P(50 < X < 60)$.

Solution

a From the table, $P(50 \le X < 60) = \dfrac{8}{100}$ i.e. 0.08.

Or, from the graph, $P(50 \le X < 60) = 0.08$.

b From the table, $P(50 \le X < 90) = \dfrac{8+3+2+2}{100}$ i.e. 0.15.

Or, from the graph, $P(50 \le X < 90) = 0.08 + 0.03 + 0.02 + 0.02 = 0.15$.

c From the note, $P(X = 45) = 0$

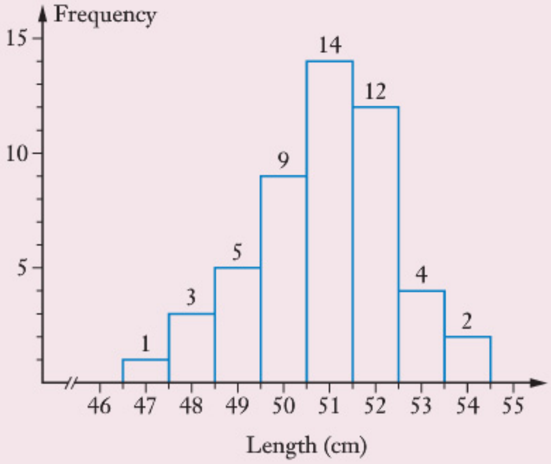
d From the table, $P(X \ge 90 | X > 50) = \dfrac{1}{8+3+2+2+1} = 0.0625$

Or, from the graph, $P(X \ge 90 | X > 50) = \dfrac{0.01}{0.08+0.03+0.02+0.02+0.01} = 0.0625$

The initial situation, involving the age distribution of Australians aged between 0 and 100, involved data based on the entire population – i.e. the ages were given for *all* people aged under 100 living in Australia (with some rounding of figures involved). Asked to determine probabilities for one item selected from the population, when we know the relevant data for the whole population, means that the probabilities will be accurate, not estimates. However, in some cases the continuous variable may be for a sample drawn from a larger population, as in example 1. We could then use the relative frequencies of our sample to give an estimate of the probabilities for the population.

EXAMPLE 2

The lengths of 50 non-premature newborn babies born in an Australian hospital gave rise to the frequency histogram on the right:



Use the above information to suggest values for each of the following probabilities where X cm is the length of a randomly selected non-premature newborn baby born in an Australian hospital.

a $P(X \leq 50.5)$

b $P\ (49.5 < X < 53.5)$

c $P(X > 50.5 \mid X < 53.5)$

Solution

a Of the 50 measurements that contributed to the histogram, 18 (= 1 + 3 + 5 + 9) of them involve a length less than or equal to 50.5 cm.

The given data suggests that
$$\begin{aligned} P(X \leq 50.5) &= \frac{18}{50} \\ &= 0.36 \end{aligned}$$

b Of the 50 measurements that contributed to the histogram, 39 (= 9 + 14 + 12 + 4) of them involve a length between 49.5 cm and 53.5 cm.

The given data suggests that $\quad P(49.5 < X < 53.5) = 0.78$

c Given that $X < 53.5$ we need only consider the 48 lengths for which this is true.

$$\begin{aligned} P(X > 50.5 \mid X < 53.5) &= \frac{P(50.5 < X < 53.5)}{P(X < 53.5)} \\ &= \frac{30}{48} \qquad \text{i.e. } 0.625. \end{aligned}$$

ISBN 9780170395205

Exercise 3A

1 The road accident statistics for a country for one year showed that for motorcyclists (drivers not passengers) in the age range fifteen to fifty-nine, 186 had died in road accidents with the distribution of the ages of these riders as shown on the right.

Age (x yrs)	Drivers killed
$15 \le x < 20$	40
$20 \le x < 25$	59
$25 \le x < 30$	29
$30 \le x < 35$	19
$35 \le x < 40$	16
$40 \le x < 45$	11
$45 \le x < 50$	8
$50 \le x < 55$	2
$55 \le x < 60$	2

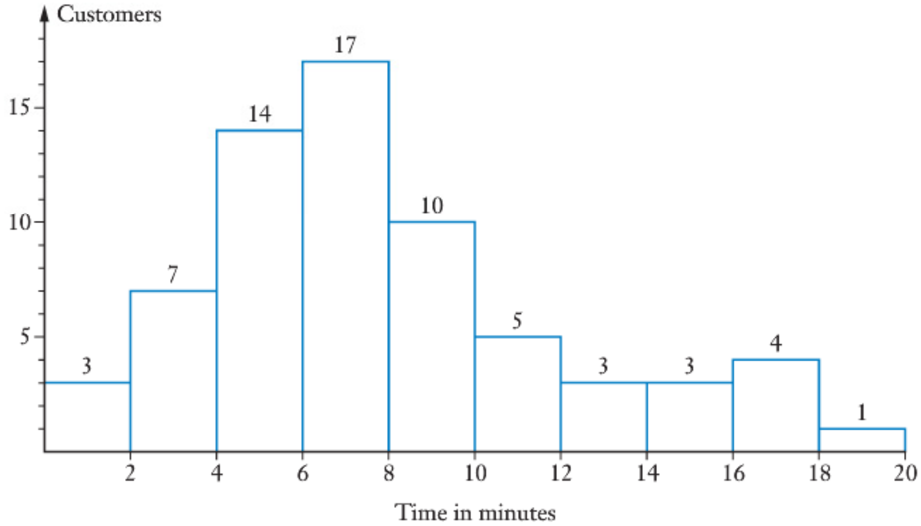
If one of these fatalities is selected at random, determine the probability that it will be for a motorcyclist aged

a less than 40 years old,

b less than 30 years old,

c at least 50 years old.

2 A pharmacy monitored the time that each customer had to wait for their prescription, from handing the prescription to the assistant, to the prescription being ready for collection. The information collected for the 67 customers requiring a prescription during one day gave rise to the following histogram.



If X minutes is the time a customer has to wait at this pharmacy for a prescription to be ready, use the information given above to suggest values for each of the following probabilities, giving your estimates as percentages to the nearest 1%.

a $P(X < 6)$

b $P(X < 4)$

c $P(X > 10)$

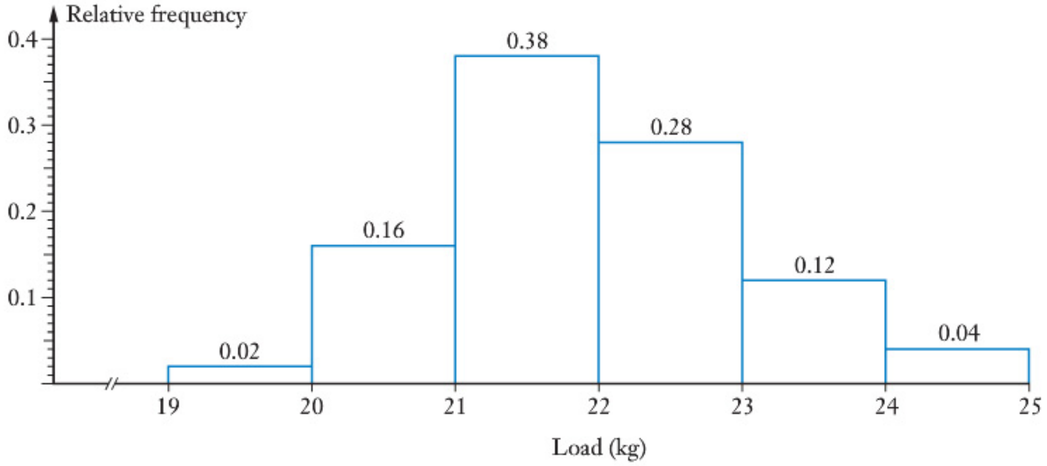
d $P(6 < X < 12)$

e $P(10 < X < 20)$

f $P(X > 6 | X < 10)$

iStock.com/anek_s

3 To test the strength of a particular type of wire under load, samples of the wire have increasing loads attached to one end of the wire, whilst the other end is fixed in a clamp. The load that causes the wire to break is noted. The results for 50 such samples gave rise to the following relative frequency histogram.



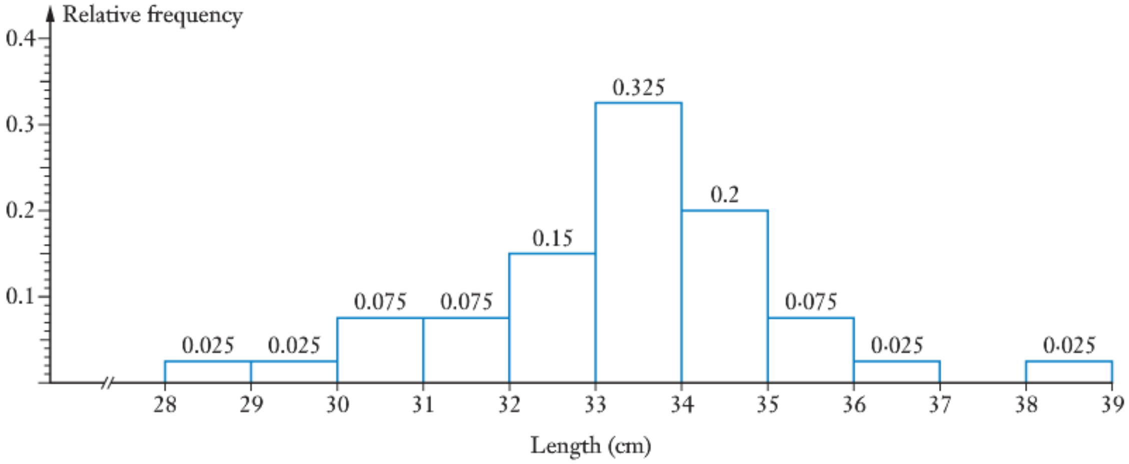
a How many of the 50 wires broke when a load between 22 kg and 24 kg was attached?

b If a random sample of wire of this type breaks when the load attached is X kg, use the above results to suggest values for each of the following probabilities.

i $P(X > 21)$ **ii** $P(X < 21)$ **iii** $P(X < 21 | 20 < X < 23)$

4 During research into a particular species of animal, a number of the adult male animals are caught, measured, tagged and released back into the wild.

The lengths of these animals gave rise to the following histogram of relative frequencies.



If an adult male animal of this species is captured and measured, use the above data to suggest the probability the length of the animal, L cm, is such that:

a $L \geq 33$

b $L > 33$

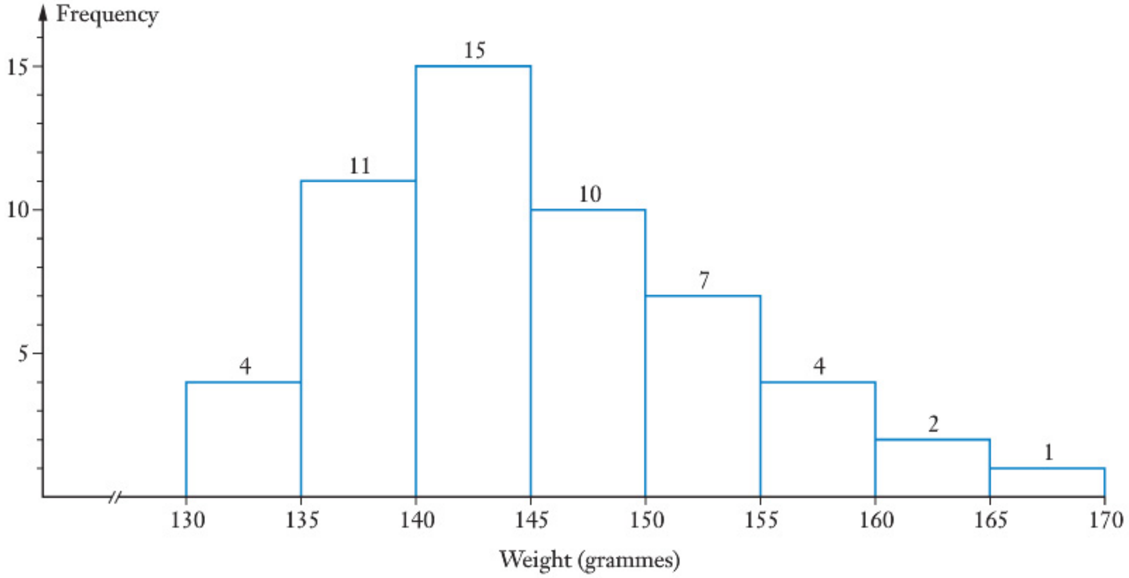
c $30 < L < 35$

d $L > 35$ given that $L > 33$

e $L > 33$ given that $32 < L < 36$

ISBN 9780170395205

5 A number of apples of a particular species were purchased from a supermarket. The weight of each apple was measured and noted. The distribution of weights gave rise to the following frequency histogram.



An apple of this same species is purchased from a supermarket. If the weight of this apple is W grams use the above data to suggest values for each of the following probabilities, giving your estimates as decimals, rounded to two decimal places.

a $P(W < 140)$ **b** $P(W \geq 140)$

c $P(140 < W < 160)$ **d** $P(W < 150 | W > 140)$

Continuous random variables

WS
Discrete and continuous random variables

Each question of the previous exercise involved a *continuous* variable. The age of a motorcyclist, the length of time it takes to get a prescription ready, the load that causes a wire to break, the length of an animal and the weight of an apple are all things which can take any value, between reasonable limits appropriate to the situation. In practice the accuracy of a measurement will be limited by the measuring device we use, but if infinite accuracy were possible *any* value, between reasonable limits, would be possible.

- Discrete random variables commonly occur when we are *counting* events, for example the number of successes in a number of trials as with the binomial distribution.
- Continuous random variables commonly occur when we are *measuring* something, for example, heights, weights, times.

Suppose X is a continuous random variable that can take any value, x, in an interval. It makes no sense to talk of $P(X = x)$ because, with an infinite number of possible values, the probability of X taking any one particular value is negligible. Instead we talk of the probability of the value of X lying in some range of values, as in the previous exercise. Rather than having a probability distribution in which each of the possible values has a particular probability, as in a discrete random variable, with continuous random variables we instead talk of a **probability density function**, or **pdf**.

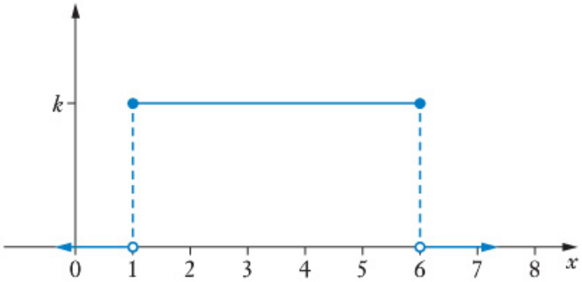
Probability density functions

Probability density function (pdf)

If $f(x)$ is the probability density function for a continuous random variable, X, then the area under $f(x)$ from $x = a$ to $x = b$ gives $P(a < X < b)$.

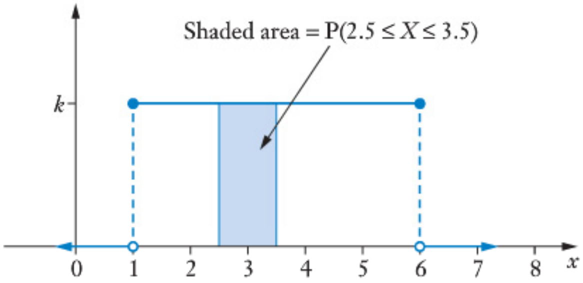
Suppose we wanted to randomly generate a number in the range 1 to 6.

In theory any number in the range is possible and we are as likely to get one in the range 1 to 2 as we are one in the range 2 to 3 or 3 to 4 etc. The distribution of probabilities is uniform across the range 1 to 6 and zero elsewhere.

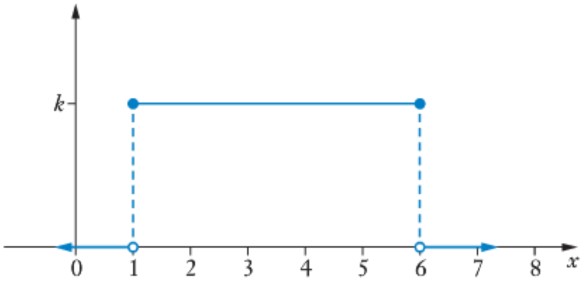


This is an example of a **uniform (or rectangular) distribution**.

For the generation of a random number in the range 1 to 6 we might be interested in $P(2.5 \leq X \leq 3.5)$. This can be shown as an area under the graph, as shown below.



The **probability density function**, or **pdf** of the random variable X will be the function that defines the following graph.



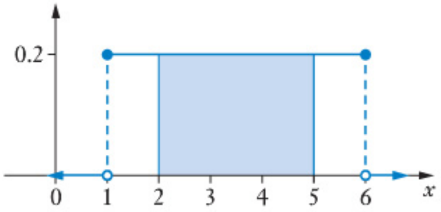
From our understanding of probability the total area under the graph must be 1. Hence for the graph above, k must be 0.2.

Hence the probability density function is given by:

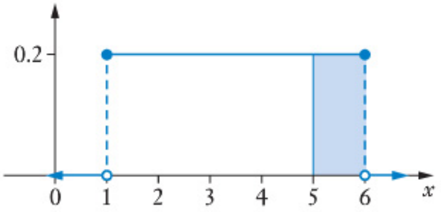
$$f(x) = \begin{cases} 0.2 & \text{for} \quad 1 \leq x \leq 6 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

ISBN 9780170395205

The diagrams below show $P(2 \le X \le 5)$ and $P(5 \le X \le 6)$ as areas under $f(x)$.



Shaded area $= P(2 \le X \le 5)$
$= 0.6$



Shaded area $= P(5 \le X \le 6)$
$= 0.2$

Note:

- With continuous random variables we do not determine $P(X = a)$ by evaluating $f(a)$. Instead $f(x)$ allows $P(a < X < b)$ to be determined:

$$P(a < X < b) \quad = \quad \text{area under } y = f(x) \text{ from } x = a \text{ to } x = b.$$

- With $P(X = a)$ being negligible it follows that $P(X \ge a) = P(X > a)$. This is consistent with the idea of the area showing probability. Whether we include a boundary line or not does not alter the area being considered.

- The graphs of $f(x)$ shown above correctly use filled and open circles to indicate where the function is and is not respectively. Thus $f(1) = 0.2$, not 0 and similarly $f(6) = 0.2$, not 0. However, whether a particular value is included or not will not alter the determination of probabilities because $P(X \ge a) = P(X > a)$. Thus when a question presents a probability density function graphically, as in the next example, the open and filled circles are often omitted.

- The previous page stated the probability density function as

$$f(x) = \begin{cases} 0.2 & \text{for} \quad 1 \le x \le 6 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

  To avoid writing '= 0 for all other values of x' we could instead say $f(x) = 0.2$ is a probability density function defined for the interval $1 \le x \le 6$.

- The probability density function must not dip below the x-axis because that would suggest a negative probability, which is meaningless.

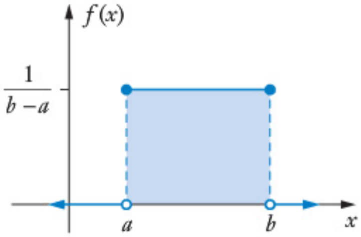
  Indeed for $f(x)$ to be a pdf on the interval $a < x < b$ we must have

  $f(x) \ge 0$ for all x in $a < x < b$,

  and the area under $f(x)$ for $a < x < b$ must equal 1, i.e. $\int_a^b f(x)\ dx = 1$.

Uniform (or rectangular) distributions

The graph below shows the probability distribution for a **uniformly distributed** continuous random variable.



For the **uniform distribution** shown the probability density function, $f(x)$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for} \quad a \le x \le b \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

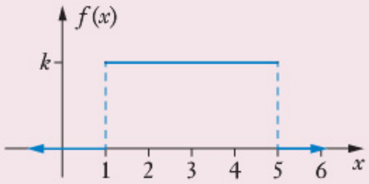
We say that the continuous random variable is **uniformly distributed on the interval $a \le x \le b$.**

Note: The interval $a \le x \le b$ is sometimes written $[a, b]$ and $a < x < b$ is sometimes written (a, b).

By symmetry, **the mean, the expected value, or the long-term average**, of the distribution will be halfway between a and b, i.e. it will equal $\frac{a+b}{2}$.

EXAMPLE 3

The continuous random variable X has the probability density function shown on the right.



Determine **a** k

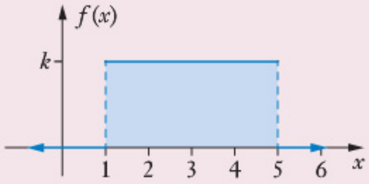
b $P(X < 3)$

c $P(2 < X < 5)$

d $P(X < 3 \mid 2 < X < 5)$

Solution

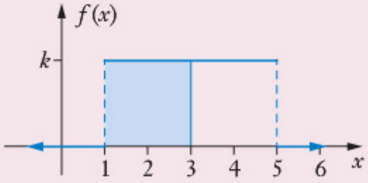
a The area of the region shaded blue in the diagram on the right must equal 1.



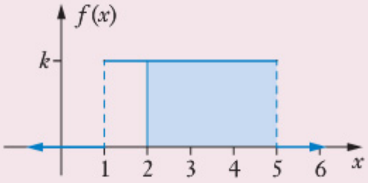
Thus $4k = 1$

$k = 0.25$

b $$\begin{aligned} P(X<3) &= 2k \\ &= 0.5 \end{aligned}$$



c $$\begin{aligned} P(2<X<5) &= 3k \\ &= 0.75 \end{aligned}$$



d $$\begin{aligned} P(X<3\,|\,2<X<5) &= \frac{P(2<X<3)}{P(2<X<5)} \\ &= \frac{0.25}{0.75} \\ &= \frac{1}{3} \end{aligned}$$

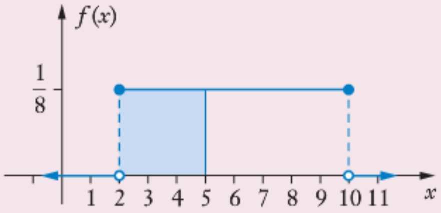
EXAMPLE 4

A continuous random variable X is uniformly distributed in the interval $2 \le X \le 10$.

Determine **a** $P(X \le 5)$ **b** $P(3 \le X \le 8)$ **c** $P(X \le 5\,|\,3 \le X \le 8)$

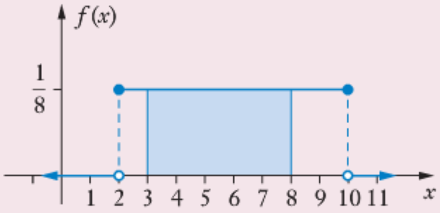
Solution

a



$$P(X \le 5) = \frac{3}{8}$$

b



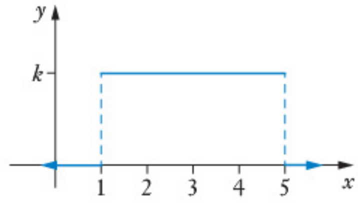
$$P(3 \le X \le 8) = \frac{5}{8}$$

c $$\begin{aligned} P(X \le 5\,|\,3 \le X \le 8) &= \frac{P(3 \le X \le 5)}{P(3 \le X \le 8)} \\ &= \frac{2}{5} \end{aligned}$$

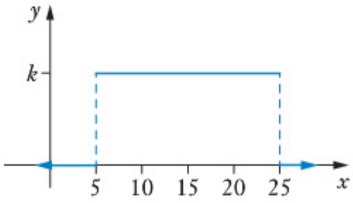
Exercise 3B

Given that each of the graphs in questions **1** to **8** show uniform probability density functions, $y = f(x)$, determine k in each case.

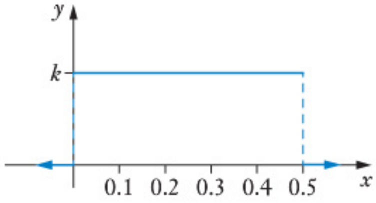
1



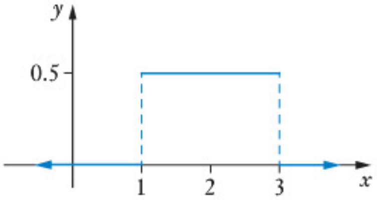
2



3

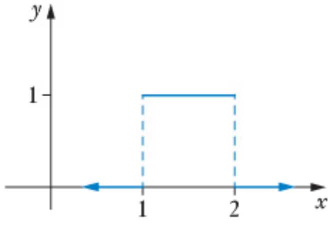


4



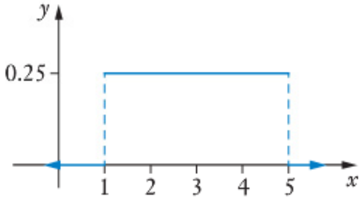
$P(X \leq k) = 0.75$

5



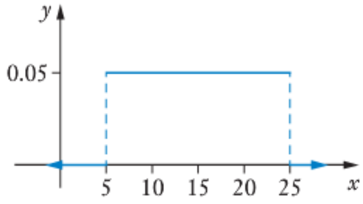
$P(X \geq k) = 0.2$

6



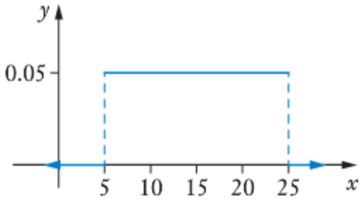
$P(X > k \mid X > 3) = 0.25$

7



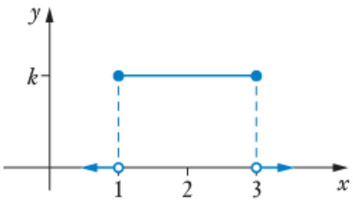
$P(X > 10 \mid X < k) = 0.5$

8



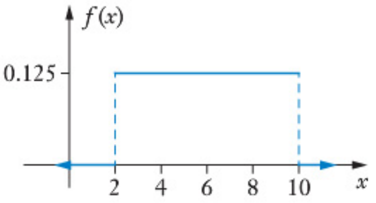
$P(X > k \mid X > 10) = 0.5$

9 Express algebraically, with the value of k determined, the uniform probability function, $f(x)$, shown graphed on the right.



ISBN 9780170395205

10 The continuous random variable X has the probability density function shown on the right.



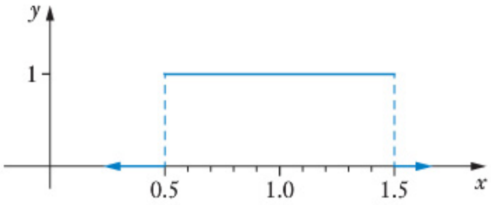
Determine **a** $P(X < 4)$

b $P(X = 4)$

c $P(X < 8)$

d $P(X > 4 \mid X < 8)$

11 The continuous random variable X has the probability density function shown on the right.



Determine

a $E(X)$, the long term average of X.

b $P(X > 1.2)$

c $P(X > 2)$

d $P(X < 2)$

e $P(X < 1 \mid X < 1.3)$

12 The continuous random variable X is uniformly distributed on the interval

$$0 \le x \le 50.$$

Find

a $E(X)$, the long term average of X.

b $P(X = 20)$

c $P(X < 20)$

d $P(X \le 20)$

e $P(X < 20 \mid X < 25)$

f $P(X < 25 \mid X < 20)$

g $P(X > 20 \mid X < 25)$

h $P(X > 25 \mid X < 20)$

13 Guided tours of a particular historic building commence every forty minutes. If we use a uniformly distributed continuous random variable X to model the time, in minutes, that a person randomly arriving at the building has to wait for the next tour to commence, show the probability density function of X graphically and find

a $P(X \le 20)$

b $P(X \ge 15)$

c $P(X \le 20 \mid X \ge 15)$

iStock.com/zensu

Non-uniform distributions

Not all probability distributions for continuous random variables are uniform. However, whatever the shape of the distribution, if $f(x)$ is a probability density function on the interval $a < x < b$ we must have

- $f(x) \geq 0$ for all x in $a < x < b$,

and

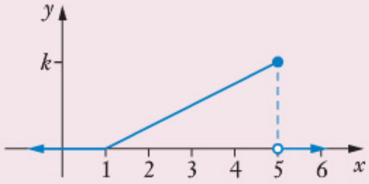
- the area under $f(x)$ for $a < x < b$ must equal 1, i.e. $\int_a^b f(x)\ dx = 1$.

EXAMPLE 5

The continuous random variable X has the probability density function shown on the right.

Determine **a** k

b $P(X \geq 3)$

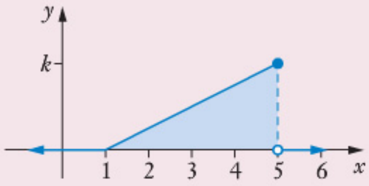


Solution

a The area of the region shaded blue in the diagram on the right must equal 1.

Thus $0.5 \times 4 \times k = 1$

$k = 0.5$



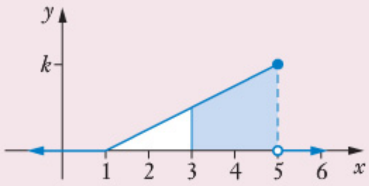
b In the diagram on the right the unshaded part of the bigger triangle is a triangle of base 2 units and height $0.5k$ units.

Unshaded area $= 0.5 \times 2 \times 0.25$ units^2

$= 0.25$ units^2

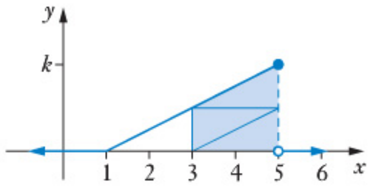
$\therefore$ Shaded area $= 0.75$ units^2

$P(X \geq 3) = 0.75$



Alternatively the answer for part **b** could be determined by:

- finding the area of the shaded trapezium directly. Shaded area $= \dfrac{0.5k + k}{2} \times 2$.
- considering the area under the curve as being made up of four triangles of equal area, three of which are shaded, as shown on the right.
- evaluating $\int_3^5 \left(\frac{1}{8}x - \frac{1}{8}\right) dx$.



ISBN 9780170395205

EXAMPLE 6

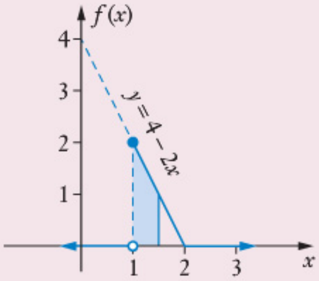
The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} 4-2x & \text{for} \quad 1 \le x \le 2 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Determine $P(X < 1.5)$.

Solution

The probability density function is shown on the right, with the blue shaded area representing $P(X < 1.5)$.



$$\therefore \quad P(X < 1.5) = \frac{2+1}{2} \times 0.5$$
$$= 0.75$$

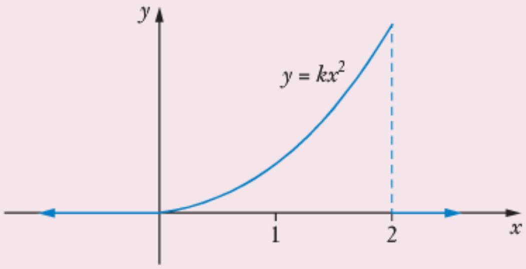
Alternatively, using calculus: $\quad P(X < 1.5) = \int_1^{1.5} (4-2x)\, dx$

$$= \left[4x - x^2\right]_1^{1.5}$$
$$= (6 - 2.25) - (4 - 1)$$
$$= 0.75$$

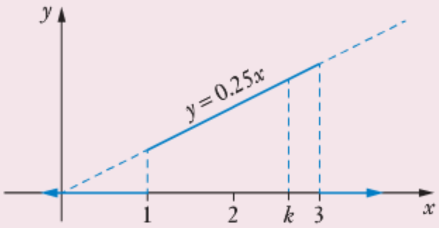
EXAMPLE 7

Each of the diagrams below show probability density functions. Use calculus to determine the value of k in each case.

a



b



$P(1 \le X \le k) = 0.75$

Solution

a

$$\int_0^2 kx^2\, dx = 1$$
$$\therefore \quad \left[\frac{kx^3}{3}\right]_0^2 = 1$$
$$\frac{8k}{3} = 1$$
$$k = \frac{3}{8}$$

b

$$\int_1^k 0.25x\, dx = 0.75$$
$$\therefore \quad \left[\frac{x^2}{8}\right]_1^k = 0.75$$
$$\frac{k^2}{8} - \frac{1}{8} = 0.75$$
$$k^2 = 7$$
$$k = \sqrt{7}$$

(Negative solution not applicable.)

Repeat part **b** of this example without using calculus.

EXAMPLE 8

Let us suppose that the continuous random variable X is the time in minutes between an event occurring and it next occurring, and that X has the probability density function:

$$f(x) = \begin{cases} 0.5e^{-0.5x} & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x \leq 0. \end{cases}$$

Determine **a** $\mathrm{P}(X \leq 2)$,

b $\mathrm{P}(5 \leq X \leq 10)$,

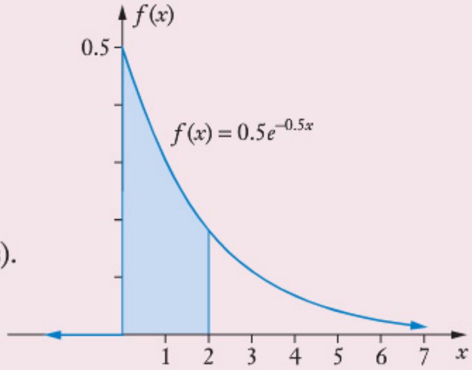
c the value of d for which $\mathrm{P}(X \leq d) = 0.5$.

Solution

a $\mathrm{P}(X \leq 2) = \int_0^2 0.5e^{-0.5x}\ dx$

Either by calculator, or algebraically, as shown below,

$$\begin{aligned} \int_0^2 0.5e^{-0.5x}\ dx &= \left[-e^{-0.5x}\right]_0^2 \\ &= -e^{-1} + e^0 \\ &= 0.6321 \text{ (correct to 4 decimal places).} \end{aligned}$$



b

$$\begin{aligned} \mathrm{P}(5 \leq X \leq 10) &= \int_5^{10} 0.5e^{-0.5x}\ dx \\ &= 0.0753 \text{ (correct to 4 decimal places)} \end{aligned}$$

c If $\mathrm{P}(X \leq d) = 0.5$ then $\int_0^d 0.5e^{-0.5x}\ dx = 0.5$

Solve algebraically, as shown below, or by calculator.

$$\begin{aligned} \therefore \qquad \left[-e^{-0.5x}\right]_0^d &= 0.5 \\ -e^{-0.5d} + 1 &= 0.5 \\ 0.5 &= e^{-0.5d} \end{aligned}$$

Taking natural logs of both sides $\quad \ln 0.5 = -0.5d$

Thus $\quad d = 1.386$ (correct to 3 decimal places)

ISBN 9780170395205

Note: Probability density functions of the form

$$f(x) = \begin{cases} ke^{-kx} & \text{for} \quad x > 0 \\ 0 & \quad\quad\;\; \text{elsewhere,} \end{cases}$$

as in the previous example, are typically involved when the random variable is the time between an event occurring and it next occurring (the interarrival time).

For functions of this type it follows that for $0 < x < \infty$ and $k > 0$, then $f(x) > 0$ and, as the reader should confirm by performing the integration,

$$\int_0^\infty f(x)\,dx = 1.$$

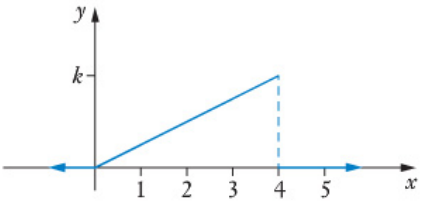
Exercise 3C

Given that each of the graphs in questions **1** to **16** show probability density functions,

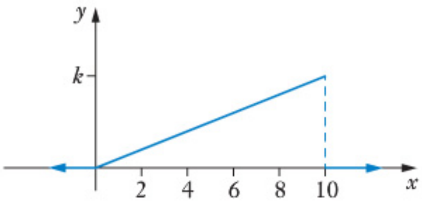
$$y = f(x),$$

determine k in each case.

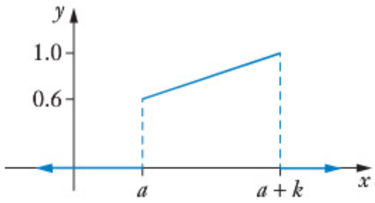
1



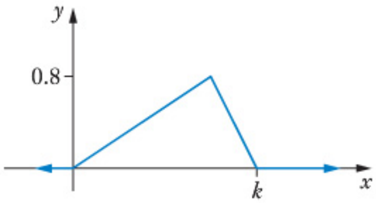
2



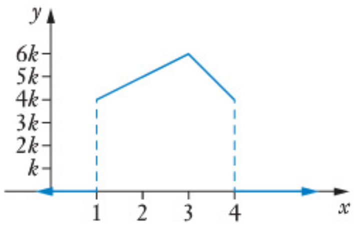
3



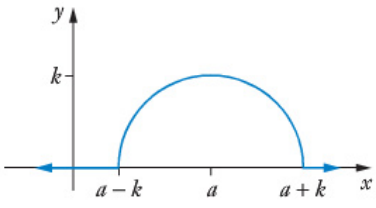
4



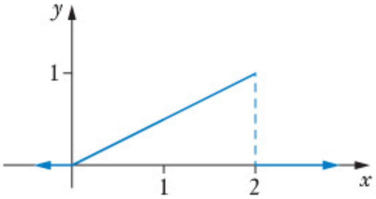
5



6

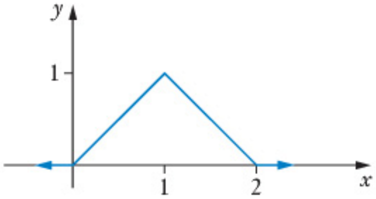


7



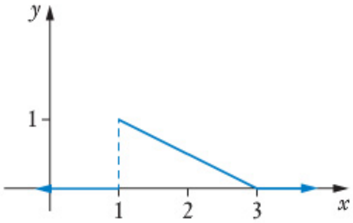
$P(X \leq k) = 0.0625$

8



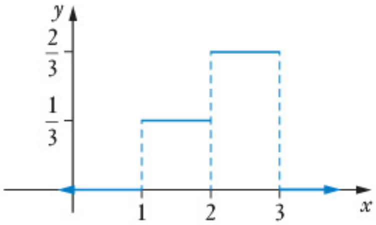
$P(X > k) = 0.02$

9



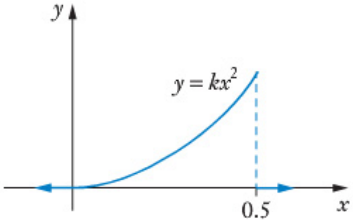
$P(X \leq k) = 0.51$

10

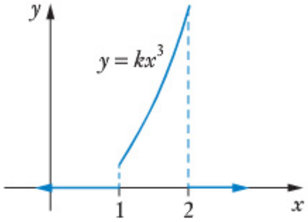


$P(X \leq k) = \frac{2}{3}$

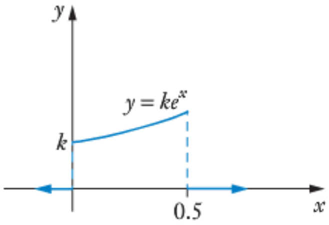
11



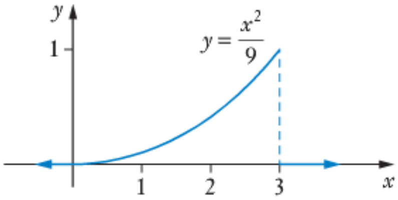
12



13

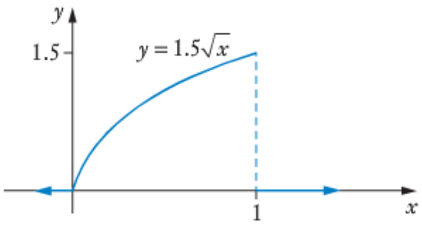


14



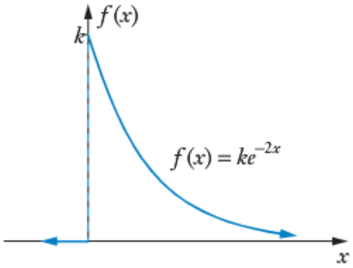
$P(X < k) = 0.512$

15



$P(X \leq k) = 0.125$

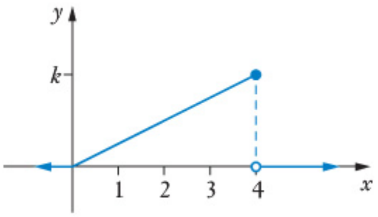
16



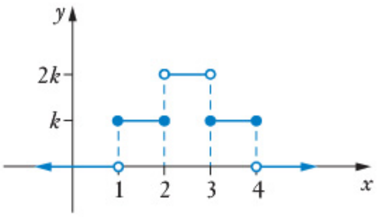
ISBN 9780170395205

17 For each of the probability density functions shown graphed below, write the probability density function, $f(x)$. (In each case k should be determined.)

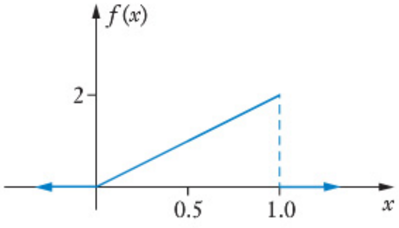
a



b



18 The continuous random variable X has the probability density function shown on the right.



Determine

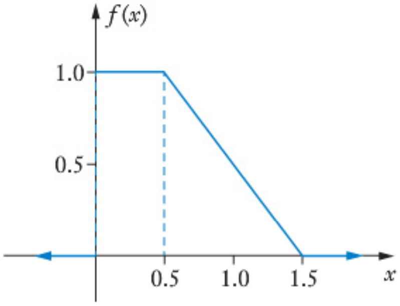
a $P(X < 0.5)$

b $P(X > 0.75)$

c $P(X < 2)$

d $P(X > 0.5 | X < 0.75)$

19 The continuous random variable X has the probability density function shown on the right.



Determine

a $P(X < 0.5)$

b $P(X > 1)$

c $P(X < 1)$

d $P(X > 0.5 | X < 1)$

20 For $0 \le x \le 5$, only one of the following functions could be a probability density function. Decide which function it is and explain why the other two cannot be probability density functions for $0 \le x \le 5$.

$$f(x) = 0.5 - 0.08x \qquad g(x) = 0.5 - 0.12x \qquad h(x) = 0.4 - 0.08x$$

21 A continuous random variable, X, has probability density function

$$f(x) = \begin{cases} 0.08(x+2) & \text{for } -2 \le x \le 3 \\ 0 & \text{for all other values of } x. \end{cases}$$

Determine **a** $P(X \ge 0)$ **b** $P(1 \le X \le 2)$ **c** $P(X \le 2 | X \ge 1)$

22 A continuous random variable, X, has probability density function

$$f(x) = \begin{cases} kx & \text{for } 0 \le x \le 5 \\ 0 & \text{for all other values of } x. \end{cases}$$

Determine **a** k **b** $P(X \le 4)$

c $P(2 \le X \le 4)$ **d** $P(X \ge 2 | X \le 4)$

23 A continuous random variable, X, has pdf: $f(x) = \dfrac{3}{2x^2}$ defined for $[1, 3]$.

a Confirm that $\displaystyle\int_1^3 \frac{3}{2x^2}\,dx = 1$.

Determine **b** $P(X \geq 2)$ **c** $P(2 \leq X \leq 2.5)$ **d** $P(X \leq 2.5 \,|\, X \geq 2)$

24 a If
$$f(x) = \begin{cases} x^2 + kx & \text{for} \quad 1 \leq x \leq 4 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

and $\displaystyle\int_1^4 f(x)\;dx = 1$, determine k.

b For the value of k determined in part **a** could $f(x)$ represent a probability density function? (Explain your answer.)

25 A continuous random variable, X, has pdf:
$$f(x) = \begin{cases} k(1-x)(x-3) & \text{for} \quad 1 \leq x \leq 3 \\ 0 & \text{for} \quad \text{for all other values of } x \end{cases}$$

Determine **a** k,

b $P(X \leq 2)$,

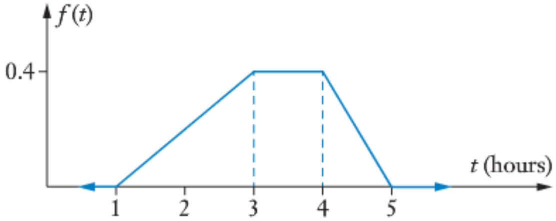
c $P(X \leq 2.5)$, giving your answer correct to 2 decimal places,

d q, correct to two decimal places, given that $P(X \geq q) = 0.6$.

26 A continuous random variable, X, has pdf:
$$f(x) = \begin{cases} \dfrac{a + bx - x^2}{9} & \text{for} \quad 1 \leq x \leq 4 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

If $P(X \leq 2) = \dfrac{5}{27}$ determine a and b.

27 The time a motor mechanic may take to carry out a particular repair can depend upon a number of things. The skill and experience of the mechanic, the age, make and condition of the vehicle are just some of the factors that could influence the time taken.

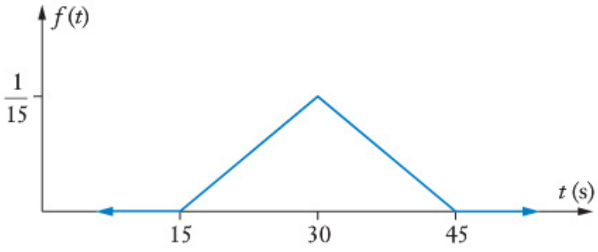


Suppose the time taken to complete a certain repair is between 1 and 5 hours with probability density function, $f(t)$, shown on the right.

a Find the probability that the repair takes less than 3 hours to complete.

b Given that the repair took more than 3 hours to complete what is the probability that it took less than 4 hours to complete?

28 Suppose that the time, in seconds, that a person actually records when asked to estimate a time of 30 seconds can be represented by a random variable with the probability density function, $f(t)$, shown on the right.



Find the probability that a person set this task will

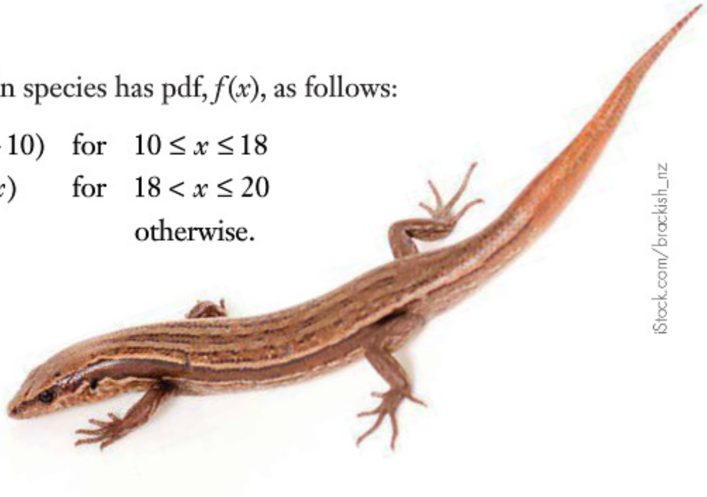
a record a time that is less than 25 seconds,

b record a time that is within 5 seconds of 30 seconds,

c record a time that is less than 40 seconds given that they record a time that is greater than 30 seconds.

29 The length, X cm, of an adult lizard of a certain species has pdf, $f(x)$, as follows:

$$f(x) = \begin{cases} 0.025(x-10) & \text{for} \quad 10 \le x \le 18 \\ 0.1(20-x) & \text{for} \quad 18 < x \le 20 \\ 0 & \quad\quad\;\, \text{otherwise.} \end{cases}$$



iStock.com/brackish_nz

Find the probability that an adult lizard of this species has a length that is

a less than 18 cm,

b greater than 14 cm,

c less than 19 cm.

30 Suppose that the number of years a particular make of washing machine lasts before its first breakdown is a continuous random variable, X, with pdf:

$$f(x) = 0.2e^{-0.2x} \qquad \text{for } (0, \infty).$$

a Determine the probability that a washing machine of this make will last more than eight years before its first breakdown. (Correct to four decimal places.)

b (Hint: Binomial distribution.) If we had six washing machines of this make, determine the probability that exactly two will last more than eight years before they experience their first breakdown, the other four experiencing their first breakdown within eight years.

Shutterstock.com/wavebreakmedia

ISBN 9780170395205

Expected value, variance and standard deviation

The *Preliminary work* reminded us of the fact that if the discrete random variable, X, has possible values x_i, with $\mathrm{P}(X = x_i) = \mathrm{p}_i$ then $\mathrm{E}(X)$, the expected or long term mean value is given by:

$$\mathrm{E}(X) \quad = \quad \Sigma(x_i\ \mathrm{p}_i)$$

the summation being carried out over all of the possible values x_i.

Further, if we use the Greek letter, μ, to represent $\mathrm{E}(X)$ then the **variance**, $\mathrm{Var}(X)$ is given by:

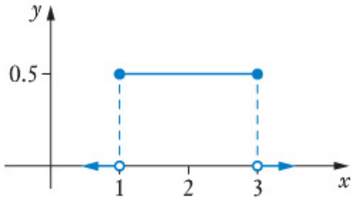
$$\mathrm{Var}(X) \quad = \quad \Sigma[\mathrm{p}_i(x_i - \mu)^2]$$

The **standard deviation** is then the square root of the variance.

For a continuous random variable X, with probability density function $f(x)$, the corresponding statements are:

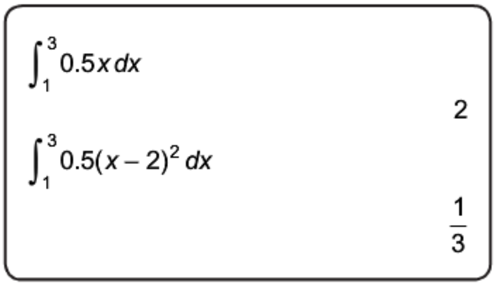
$$\begin{aligned} \mathrm{E}(X) &= \int_{-\infty}^{\infty} x\,f(x)\,dx \\ \mathrm{Var}(X) &= \int_{-\infty}^{\infty}\left[f(x)(x-\mu)^2\right]dx \end{aligned}$$

Thus, for the pdf shown on the right,



$$\begin{aligned} \mathrm{E}(X) &= \int_{-\infty}^{\infty} x\,f(x)\,dx \\ &= \int_{1}^{3}(x \times 0.5)\,dx \\ &= \left[\frac{x^2}{4}\right]_1^3 \\ &= \frac{9}{4} - \frac{1}{4} \\ &= 2 \quad \text{(As we would expect.)} \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= \int_{-\infty}^{\infty}\left[f(x)(x-\mu)^2\right]dx \\ &= \int_{1}^{3}\left[0.5(x-2)^2\right]dx \\ &= \frac{1}{3} \end{aligned}$$



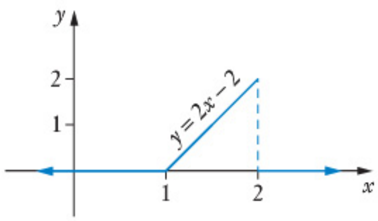
$$\int_1^3 0.5x\,dx$$

$$2$$

$$\int_1^3 0.5(x-2)^2\,dx$$

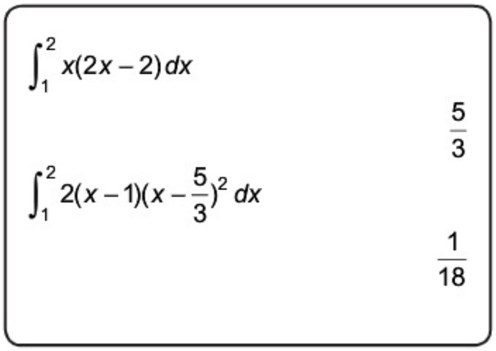
$$\frac{1}{3}$$

Similarly, for the pdf shown on the right:



$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x\,f(x)\,dx \\
&= \int_{1}^{2} x(2x-2)\,dx \\
&= \left[\frac{2x^3}{3} - x^2\right]_1^2 \\
&= \left(\frac{16}{3} - 4\right) - \left(\frac{2}{3} - 1\right) \\
&= \frac{5}{3}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} \left[f(x)(x-\mu)^2\right] dx \\
&= \int_{1}^{2} \left[2(x-1)\left(x-\frac{5}{3}\right)^2\right] dx \\
&= \frac{1}{18}
\end{aligned}$$



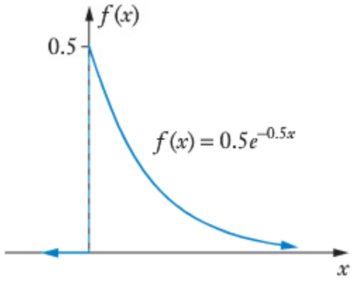
$\int_1^2 x(2x-2)dx$

$\frac{5}{3}$

$\int_1^2 2(x-1)(x-\frac{5}{3})^2\,dx$

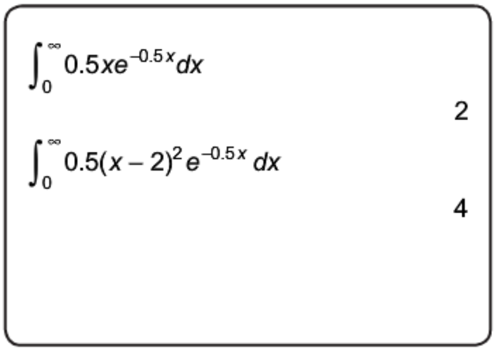
$\frac{1}{18}$

For the pdf shown,



carrying out appropriate integrations on a calculator,

$$\begin{aligned}
\mathrm{E}(X) &= 2 \\
\mathrm{Var}(X) &= 4 \\
\mathrm{SD}(X) &= 2
\end{aligned}$$



$\int_0^\infty 0.5xe^{-0.5x}dx$

2

$\int_0^\infty 0.5(x-2)^2 e^{-0.5x}\,dx$

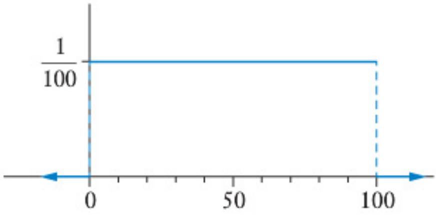
4

Change of scale and origin

Suppose that the temperature of something, recorded in °C, is a continuous random variable X.

Further suppose that X is uniformly distributed between 0 and 100.

The probability density function for X will be as shown on the right.



By symmetry, or by calculus (see right), $\mathrm{E}(X) = 50$.

$$\int_0^{100} \left(\frac{1}{100} x \right) dx = 50.$$

By calculus (see right), $\mathrm{Var}(X) = \frac{2500}{3}$.

$$\int_0^{100} \left(\frac{1}{100} (x-50)^2 \right) dx = \frac{2500}{3}.$$

Hence
$$\begin{aligned} \mathrm{SD}(X) &= \frac{50}{\sqrt{3}} \\ &= \frac{50\sqrt{3}}{3}. \end{aligned}$$

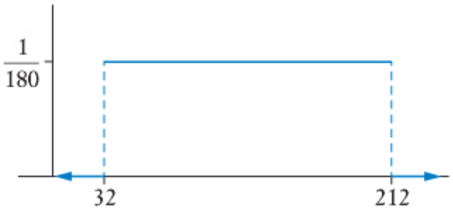
Suppose instead that the temperatures were measured in °F (degrees Fahrenheit).

To change a temperature in °C to the equivalent temperature in °F we multiply by 1.8 and add 32. I.e. $°\mathrm{F} = 1.8 \times °\mathrm{C} + 32$

Hence, $0°\mathrm{C} = 32°\mathrm{F}$

and $100°\mathrm{C} = 212°\mathrm{F}$.

We would now have a uniform distribution (Y) involving temperatures from 32°F to 212 °F.



By symmetry, or by calculus (see right), $\mathrm{E}(Y) = 122$.

$$\int_{32}^{212} \left(\frac{1}{180} x \right) dx = 122.$$

By calculus (see right), $\mathrm{Var}(Y) = 2700$.

$$\int_{32}^{212} \left(\frac{1}{180} (x-122)^2 \right) dx = 2700.$$

Hence $\mathrm{SD}(Y) = 30\sqrt{3}$

Notice that $122 = 1.8 \times 50 + 32$

I.e. $\mathrm{E}(Y) = 1.8 \times \mathrm{E}(X) + 32$

And $30\sqrt{3} = 1.8 \times \frac{50\sqrt{3}}{3}$

I.e. $\mathrm{SD}(Y) = 1.8 \times \mathrm{SD}(X)$

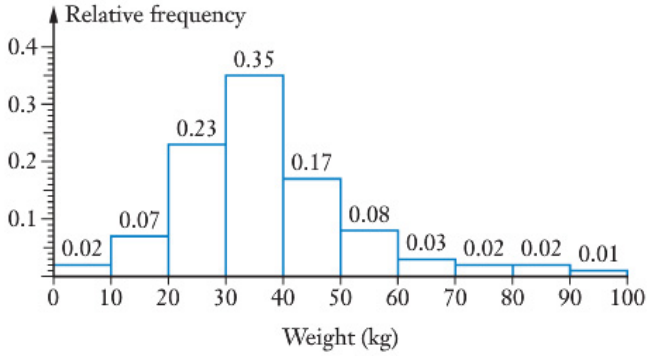
This should not really be any surprise because the *Preliminary work* section reminded us of the effect of *changes of scale and origin*.

If the random variable X has mean μ and standard deviation σ (and variance σ^2) then the random variable $aX + b$ will have mean $a\mu + b$ and standard deviation $|a|\,\sigma$ (and variance $a^2\sigma^2$).

Cumulative distribution function

Let us again consider the weights of the 100 animals of a particular species that we considered in Example 1 on page 45:

Weight (kg)	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Number of animals	2	7	23	35	17	8	3	2	2	1
Relative frequency	0.02	0.07	0.23	0.35	0.17	0.08	0.03	0.02	0.02	0.01



We could present the information in terms of the 'running totals' or *cumulative* frequencies:

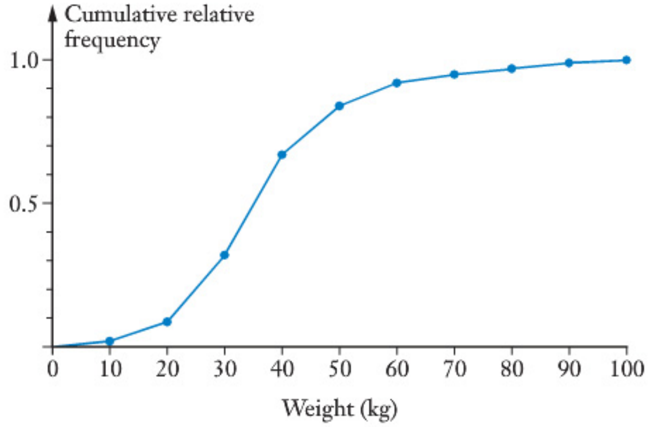
Weight (kg)	≤ 10	≤ 20	≤ 30	≤ 40	≤ 50	≤ 60	≤ 70	≤ 80	≤ 90	≤ 100
Number of animals	2	9	32	67	84	92	95	97	99	100
Relative frequency	0.02	0.09	0.32	0.67	0.84	0.92	0.95	0.97	0.99	1

As we would expect for cumulative frequencies:

- the numbers of animals increase (or could stay the same) as we move from one cell to the next cell on the right,
- the final frequency figure is 100, the total number of animals involved,

and
- the final relative frequency figure is 1.

Showing the cumulative relative frequencies as a line graph:



(A graph showing cumulative frequencies in this way is also known as an **ogive**.)

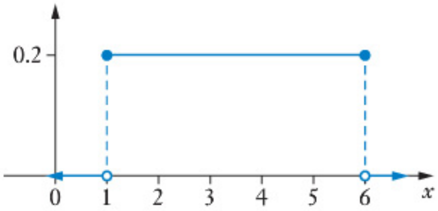
ISBN 9780170395205

Given a probability density function, pdf, for a random variable, X, we can similarly create a cumulative distribution function, **cdf**.

The cumulative distribution function, $f(x)$, will be such that $f(k)$ gives the probability that X will take a value less than or equal to k.

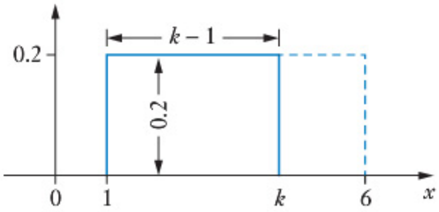
Consider the following uniform pdf:

$$f(x) = \begin{cases} 0.2 & \text{for} \quad 1 \le x \le 6 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$



$$\text{For } 1 \le k \le 6, \qquad \begin{aligned} P(X \le k) &= \int_1^k 0.2\,dx \\ &= [0.2x]_1^k \\ &= 0.2(k-1) \end{aligned}$$

Note that the same formula could be obtained without integration, by considering the area of a rectangle with a base of $(k - 1)$ and height 0.2.



Hence the cumulative distribution function will be

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ 0.2(x-1) & \text{for} \quad 1 \le x \le 6 \\ 1 & \text{for} \quad x > 6 \end{cases}$$

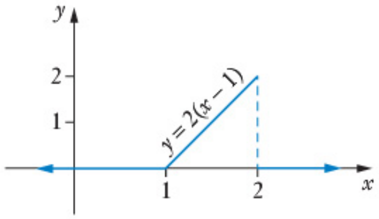
If we want to determine $P(3 \le X \le 5)$, we could then proceed as follows.

$$\begin{aligned} P(3 \le X \le 5) &= P(X \le 5) - P(X < 3) \\ &= 0.2(5-1) - 0.2(3-1) \\ &= 0.8 - 0.4 \\ &= 0.4 \end{aligned}$$

Of course this answer is exactly the same as would be obtained by considering the appropriate rectangular area, or by evaluating $\int_3^5 0.2\,dx$. The cumulative distribution function has simply 'done the integration for us'.

ISBN 9780170395205

Consider the triangular pdf shown on the right.



For $1 \le k \le 2$,

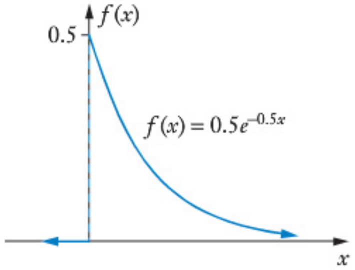
$$\begin{aligned} P(X \le k) &= \int_1^k 2(x-1)\,dx \\ &= \left[(x-1)^2\right]_1^k \\ &= (k-1)^2 \end{aligned}$$

Hence

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x \le 1 \\ (x-1)^2 & \text{for} \quad 1 < x \le 2 \\ 1 & \text{for} \quad x > 2 \end{cases}$$

(The placement of the '= part of the inequality' could differ from that shown above.)

Consider the exponential pdf shown on the right.



For $k \ge 0$,

$$\begin{aligned} P(X \le k) &= \int_0^k 0.5e^{-0.5x}\,dx \\ &= \left[-e^{-0.5x}\right]_0^k \\ &= 1 - e^{-0.5k} \end{aligned}$$

Hence the cumulative distribution function will be:

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-0.5x} & \text{for} \quad x \ge 0 \end{cases}$$

Using this function to determine $P(5 \le X \le 10)$ as an example,

$$\begin{aligned} P(5 \le X \le 10) &= P(X \le 10) \quad - \quad P(X < 5) \\ &= (1 - e^{-5}) \quad - \quad (1 - e^{-2.5}) \\ &= 0.0753 \quad \text{(correct to 4 decimal places)} \end{aligned}$$

(As obtained by determining $\int_5^{10} 0.5e^{-0.5x}\,dx$ in example 8 earlier in this chapter on page 58.)

Thus whilst a pdf allows probabilities to be determine using integration, a cdf allows probabilities to be determined by direct substitution into a formula.

If the continuous random variable X has probability density function $f(x)$ we can define the cumulative distribution function of X more formally as

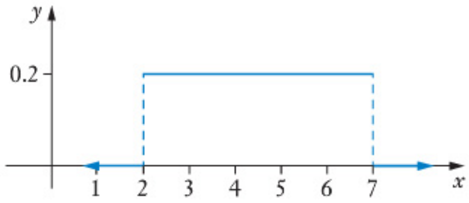
$$\int_{-\infty}^{x} f(t)\,dt$$

Exercise 3D

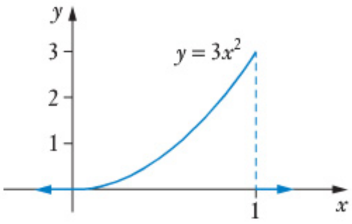
Expected value, variance and standard deviation.
(Evaluate appropriate definite integrals using your calculator if you wish.)

Use calculus to determine the mean (expected value) and variance of each of the pdfs shown in questions **1** to **4**.

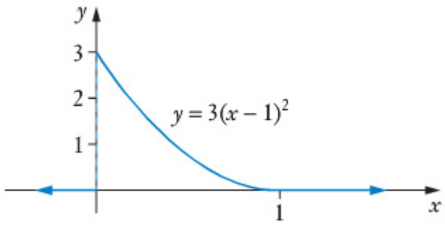
1



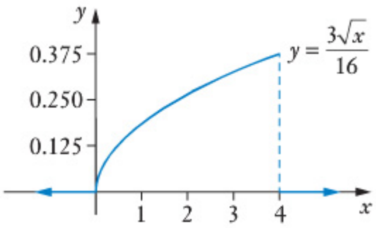
2



3



4



5 Jennifer spends X minutes in the shower, with X having the probability density function $f(x)$ defined as follows,

$$f(x) \quad = \quad \begin{cases} \dfrac{12-x}{50} & \text{for} \quad 2 \le x \le 12 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Find **a** E(X), the expected value of X,

b the standard deviation of X.

6 The distance, in metres, between consecutive defects in a piece of wire is a continuous random variable X, with probability density function:

$$f(x) \quad = \quad \begin{cases} 0.01e^{-0.01x} & \text{for} \quad x > 0 \\ 0 & \qquad\quad \text{elsewhere.} \end{cases}$$

Performing integrations using your calculator, determine the mean distance between defects.

7 Let us suppose that the number of hours of homework Sandy does each day is a continuous random variable X, with probability density function, $f(x)$, given by

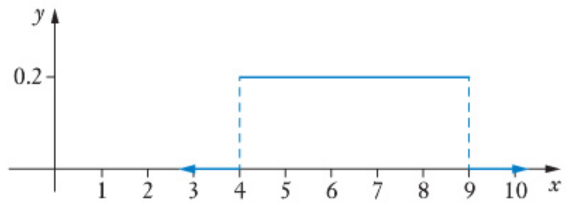
$$f(x) \quad = \quad \begin{cases} \dfrac{3(6x - x^2 - 5)}{32} & \text{for} \quad 1 \le x \le 5 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Find the mean, variance and standard deviation of X.

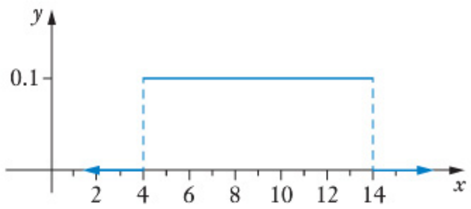
Change of scale and origin

Use your answers to question **1** to determine the mean and variance of the pdfs shown in questions **8** and **9**.

8



9



10 If a continuous random variable X has an expected value of 12 and a standard deviation of 3, find the expected value and the standard deviation of the continuous random variable Y in each of the following situations

a $Y = 3X$ **b** $Y = X + 3$ **c** $Y = 2X + 5$

11 If a continuous random variable X has an expected value of 20 and a standard deviation of 4, find the expected value and the standard deviation of the continuous random variable Z in each of the following situations

a $Z = 5X + 2$ **b** $Z = 2X + 5$ **c** $Z = 3X + 4$

12 The random variable X involves temperatures measured in degrees Celsius.

X has mean 48 and variance 16.

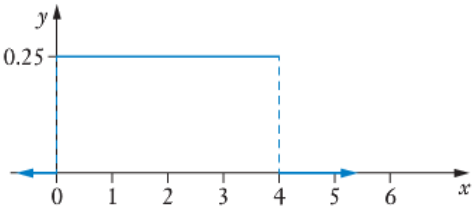
If the random variable Y involves the same temperature distribution as X but with the temperatures changed to degrees Fahrenheit, what will be the mean, variance and standard deviation of Y? (Note, $°\text{F} = 1.8 \times °\text{C} + 32$.)

13 How would changing a random variable involving lengths measured in centimetres to the same lengths measured in metres alter the mean and standard deviation?

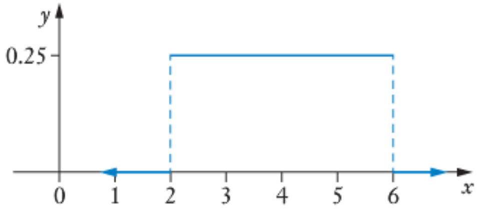
Cumulative distribution function

Express each of the following probability density functions as cumulative distribution functions.

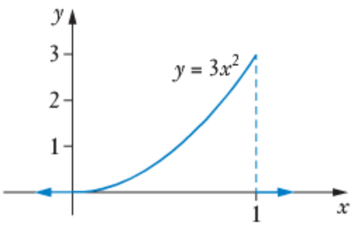
14



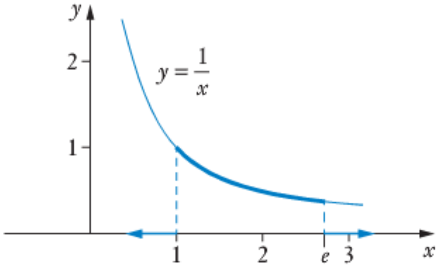
15



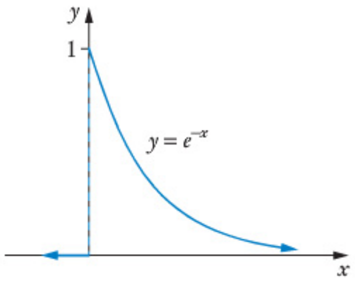
16



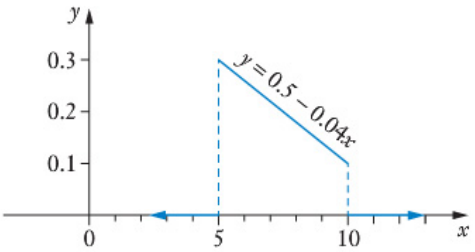
17



18



19



20 The continuous random variable X has the cumulative distribution function

$$\mathrm{P}(X \le x) \;=\; \begin{cases} 0 & \text{for} \quad x < 5 \\ 0.1(x-5) & \text{for} \quad 5 \le x \le 15 \\ 1 & \text{for} \quad x > 15. \end{cases}$$

Determine: **a** $\mathrm{P}(X \le 12)$ **b** $\mathrm{P}(X \le 8)$

c $\mathrm{P}(8 \le X \le 12)$ **d** $\mathrm{P}(X > 8)$

21 Let us suppose that in World War I the number of wartime flying hours a pilot would total before being shot down is a continuous random variable X with cumulative distribution function:

$$\mathrm{P}(X \le x) \;=\; \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-\frac{x}{15}} & \text{for} \quad x \ge 0. \end{cases}$$

Determine, correct to four decimal places, the probability that a World War I pilot:

a is shot down before totalling 20 wartime flying hours,

b totals at least 20 wartime flying hours before being shot down,

c is shot down before reaching a total of 5 wartime flying hours.

d If a pilot manages to total 15 wartime flying hours without being shot down, what is the probability that this pilot makes it to at least 20 hours?

e (Hint: Binomial distribution.)
If we consider 5 World War I pilots, what is the probability that at least 3 of them each totalled at least 20 wartime flying hours before being shot down?

Alamy Stock Photo/Antony Nettle

Miscellaneous exercise three

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

1 Solve $3^x - 1 = 5$ giving the *exact* answer using base ten logarithms.

2 A continuous random variable X is uniformly distributed on the interval

$$-2 \le X \le 3.$$

Determine **a** $P(X \ge 0)$ **b** $P(1 \le X \le 2)$ **c** $P(X \le 2 \mid X \ge 1)$

3 If $\log_c 2 = p$ and $\log_c 10 = q$, express each of the following in terms of p or q or both p and q:

a $\log_c 5$ **b** $\log_c 40$ **c** $\log_c 200$

d $\log_c (8c)$ **e** $\log_2 10$ **f** $\log 2$

4 Clearly showing your use of natural logarithms, solve each of the following equations giving your answers as exact values.

a $e^x + e^{x+1} = 17$ **b** $e^{2x+1} = 50^{x-7}$

5 (Revision of the binomial distribution.)

If a normal fair six-sided die is rolled five times what is the probability of obtaining a six on

a exactly three of the rolls?

b all of the first three rolls?

c more than three of the rolls?

d Given that a six occurs more than once in the five rolls, what is the probability that a six occurs more than twice?

Differentiate each of the following with respect to x. For some it may be advisable to use the laws of logarithms *before* differentiating.

6 $y = \ln 5x$ **7** $y = 3x + \ln 3x$ **8** $y = 2\ln x$

9 $y = 2\ln(x^3)$ **10** $y = \ln(2\sqrt{x})$ **11** $y = \ln\left(\frac{2}{x}\right)$

12 Through geological surveys and test drilling, a company discovers a new oil field off the coast of Australia. The experts predict that profitable extraction of the oil can be carried out and that in any one year this extraction will reduce the quantity of oil remaining in the field by 5% of what it was at the beginning of that year. Extraction will become unprofitable when just 20% of the original quantity remains.

For how many years can the company expect the field to remain profitable?

ISBN 9780170395205

13 Find the exact coordinates of the point(s) on the following curves where the gradient is as stated.

a $y = x + \ln 2x$ gradient 1.5.

b $y = \ln[x(x+3)]$ gradient 0.5.

14 The total cost, \$$C$, for producing x units of a certain product is given by:

$$C \approx 600 + 200\ln(1+x).$$

Find **a** an expression for $\frac{dC}{dx}$, the rate of change of C with respect to x,

b the average cost per unit when $\frac{dC}{dx} = 2$.

15 Find the equation of the tangent to $y = \ln(2\sin x)$ at the point $\left(\frac{\pi}{6}, 0\right)$.

16 Let us suppose that for any randomly chosen undecayed atom of a radioactive element, the atom will decay X hours later where X has the cumulative distribution function:

$$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\frac{x}{8}} & \text{for } x \ge 0. \end{cases}$$

Determine the probability that if we randomly choose an undecayed atom of this radioactive element the atom will decay

a within the next 8 hours,

b within the next twenty-four hours,

c more than twenty-four hours from now.

17 Use calculus to determine the exact coordinates of any stationary points on the graph of $y = 2x^2 - \log_e x$, $(x > 0)$, and use the 2nd derivative test and/or the sign test to determine whether they are maximum, minimum or inflection points.

18 A random variable X has probability density function:

$$f(x) = \begin{cases} ae^{-bx} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Prove that $a = b$.

If $a = 0.25$, find the mean and variance of X, using your calculator to perform appropriate integrations.

4.

The normal distribution

- Standard scores
- Normal distribution
- Using a calculator
- In the old days – using a book of tables
- Notation
- Quantiles
- The normal distribution pdf
- Using the normal distribution to model data
- Can we use the normal distribution to model discrete data?
- Limitations of probability models for predicting real behaviour
- Miscellaneous exercise four

Note

Students taking both *Mathematics Methods and Mathematics Applications*, and who studied *Mathematics Applications Unit Two* earlier, will be familiar with much of the work of this chapter. However there are some ideas mentioned here that were not included in the treatment of the concept in *Mathematics Applications*, and some of the questions in this chapter were not in the corresponding chapter of *Mathematics Applications Unit Two*. Hence any students in this category are encouraged to work through this chapter anyway, as the repeat work will serve as useful revision, and the new work will extend their current understanding appropriately.

Situation

Test 1 27	Kym sits a Mathematics test and achieves a mark of 27. In the next test she scores 30. Has she improved?	Test 2 30
	Before answering this question we might first ask: What was each test out of?	
$\frac{27}{40}$	Suppose that test 1 was out of 40 and test 2 was out of 50. Can we now decide whether she has improved?	$\frac{30}{50}$
	Before answering we may want to know if the tests were of similar difficulty. What was the mean mark in each test?	
Mean 23	Suppose the mean in test 1 was 23 and in test 2 was 25. Now can we judge whether her test 2 mark shows an improvement?	Mean 25
	What if we also knew the standard deviation for each test as well?	
St dev 5	Suppose the standard deviation in test 1 was 5 marks and in test 2 was 10 marks.	St dev 10

Now can you suggest whether or not Kym's mark in test 2 was an improvement on her mark in test 1?

iStock.com/Thomas_EyeDesign

Standard scores

In the *situation* on the previous page did you consider expressing Kym's test scores in terms of the number of standard deviations each was from the mean?

Expressing a score as a number of standard deviations above or below the mean is called **standardising** the score. We obtain the **standard score**.

$$\text{Standardised score} = \frac{\text{Raw score} - \text{mean}}{\text{Standard deviation}}$$

EXAMPLE 1

Jennifer scores 23, 35 and 17 in tests A, B and C respectively. If the mean and standard deviation in each of these tests are as given below express each of Jennifer's test scores as standardised scores.

Test A:	mean	30,	standard deviation	5
Test B:	mean	32,	standard deviation	6
Test C:	mean	15,	standard deviation	2.5

Solution

In Test A Jennifer's standardised score is $\frac{23-30}{5}$ i.e. -1.4.

In Test B Jennifer's standardised score is $\frac{35-32}{6}$ i.e. 0.5.

In Test C Jennifer's standardised score is $\frac{17-15}{2.5}$ i.e. 0.8.

Exercise 4A

1 Express each of the following as a standard score.

a A score of 65 in a test that had a mean of 60 and a standard deviation of 5.

b A score of 72 in a test that had a mean of 55 and a standard deviation of 10.

c A score of 50 in a test that had a mean of 58 and a standard deviation of 4.

d A score of 60 in a test that had a mean of 58 and a standard deviation of 4.

e A score of 58 in a test that had a mean of 64 and a standard deviation of 8.

2 SuMin scores 30, 50, 7 and 26 in tests A, B, C and D respectively. If the mean and standard deviation in each of these tests are as given below, express each of SuMin's test scores as standardised scores.

Test A:	mean	20,	standard deviation	4
Test B:	mean	60,	standard deviation	10
Test C:	mean	6,	standard deviation	0.8
Test D:	mean	25,	standard deviation	5

3 All of the first-year students on a particular technology course sat exams in the core subjects of Mathematics, Chemistry, Electronics and Computing. The exam results produced the following summary statistics:

Mathematics exam:	mean mark	60,	standard deviation	10.4
Chemistry exam:	mean mark	72,	standard deviation	7.2
Electronics exam:	mean mark	48,	standard deviation	14.6
Computing exam:	mean mark	63,	standard deviation	7.4

One student scored 56 in Mathematics, 74 in Chemistry, 39 in Electronics and 72 in Computing. Standardise each of these scores and rank the subjects for this student, listing them from best to worst on the basis of these standard scores.

4 All year ten students in a particular region sat exams in Mathematics, English, Science and Social Studies. The exam results in these subjects produced the following means and standard deviations.

Mathematics:	Mean	63,	Standard deviation	14
English:	Mean	64,	Standard deviation	10
Science:	Mean	72,	Standard deviation	8
Social Studies:	Mean	106,	Standard deviation	22

One student achieved the following scores:

76 in Mathematics, 75 in English,
78 in Science, 104 in Social Studies.

Rank the four subjects in order for this student, highest standardised score first.

5 Jill and her boyfriend Jack sit the same maths exam, along with the 156 other candidates studying the course for which the exam formed a part of the assessment.

- The exam was marked out of 120.
- The mean mark for the entire 158 students was 65.2 and the standard deviation of the marks was 8.8.
- Jill scored 74 out of 120 and Jack scored 63 out of 120.

Complete the three incomplete responses from Jill shown below in the following conversation between her and her mother:

Jill (arriving home from school): *'Hi Mum. How's your day been?'*

Jill's mum: *'Pretty good, dear. How was yours?*
Did you get any marks back from the exams you did?'

Jill: *'Yeah, I got my maths mark.'*

Jill's mum: *'What did you get?'*

Jill, quoting her exam mark as a standard score replied:
'Well, I got _____.'

Jill's mum: *'What! That sounds awful! What was the average?'*

Jill, again quoting standard scores: *'The mean was _____.'*

Jill's mum: *'What! What did Jack get?'*

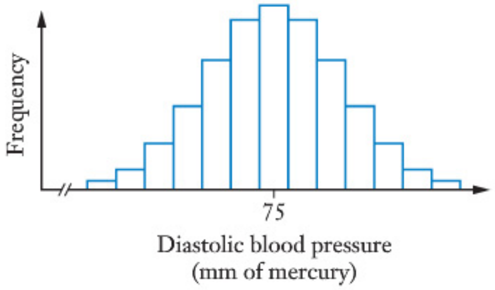
Jill: *'Oh, he got _____.'*

Jill's mum (who knew something about mathematics):
'Wait a minute. Are we talking standard scores here?'

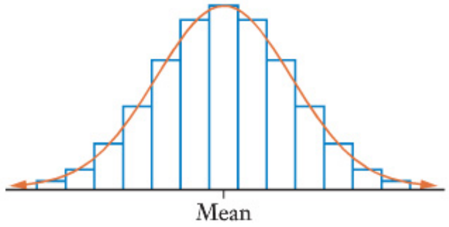
WS The standard normal curve

Normal distribution

Suppose the diastolic blood pressure of a large number of adults was measured and the mean value was found to be 75 mm of mercury (mm of mercury being the units blood pressure is measured in). The data collected, if presented as a histogram, could well have a shape similar to the diagram shown on the right, i.e. a symmetrical distribution with many values close to the mean and the number of values decreasing as we move further from the mean.



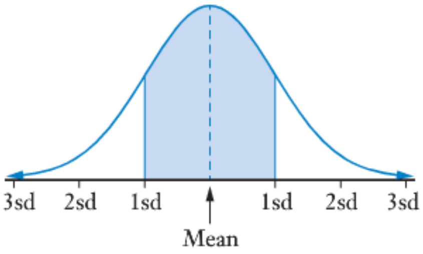
Fitting a smooth curve to the midpoints of the columns we obtain a '**bell-shaped curve**' as shown on the right.



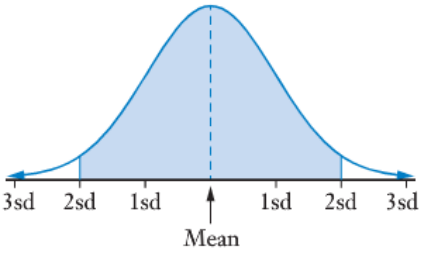
If we make many measurements of something that occurs naturally, for example, the heights of many adult females, the weights of many domestic cats, the foot lengths of many adult males, the histogram of the data often follows this sort of shape.

Data of this kind is said to be **normally distributed**. In **normal distributions** approximately two thirds of the population lie within one standard deviation of the mean, 95% would lie within two standard deviations of the mean and almost all would lie within three standard deviation of the mean.

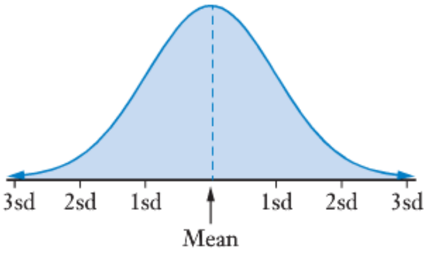
This is the 68%, 95%, 99.7% rule.



68.3% of area under curve shaded



95.4% of area under curve shaded



99.7% of area under curve shaded

In terms of probabilities, we could say that the probability of a randomly selected individual from a normally distributed population being within

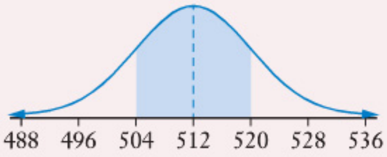
- one standard deviation of the mean is 0.683,
- two standard deviations of the mean is 0.954,
- three standard deviations of the mean is 0.997.

Note: The normal distribution is also referred to as the Gaussian distribution, after the German mathematician Carl Gauss.

EXAMPLE 2

A box of breakfast cereal has 'contains 500 grams of breakfast cereal' printed on it. Suppose that in fact the weight of breakfast cereal contained in these boxes is normally distributed with a mean of 512 grams and a standard deviation of 8 grams. Determine the probability that a randomly chosen box of this cereal contains between 504 grams and 520 grams.

Solution



With a mean of 512 grams and a standard deviation of 8 grams:

504 grams is one standard deviation below the mean

and 520 grams is one standard deviation above the mean.

For normally distributed data the probability that a randomly chosen data point is within 1 standard deviation of the mean is, from the previous page, 0.683.

Thus the probability that a randomly chosen box of this cereal contains between 504 grams and 520 grams is 0.68.

The above example could be worked out using the '68' in the *68%, 95% 99.7% rule* because the question involved numbers of standard deviations that this rule relates to. What would we have done if instead the question had asked for the probability of a randomly chosen box of the cereal containing less than 500 grams? In this case 500 grams is 1.5 standard deviations below the mean, a situation not covered by the 68%, 95% 99.7% rule. In this case we can use the ability of various calculators to determine such probabilities, as shown by the next example (also based on the breakfast cereal situation of the above example).

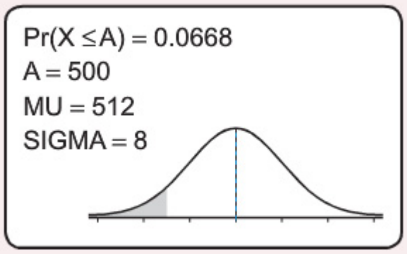
EXAMPLE 3

A box of breakfast cereal has 'contains 500 grams of breakfast cereal' printed on it. Suppose that in fact the weight of breakfast cereal contained in these boxes is normally distributed with a mean of 512 grams and a standard deviation of 8 grams.

a Determine the probability that a randomly chosen box of this cereal contains less than 500 grams.

b In a random sample of 100 boxes of this cereal approximately how many boxes should we expect to contain less than 500 g?

Solution



Pr(X ≤A) = 0.0668
A = 500
MU = 512
SIGMA = 8

a For a randomly distributed set of values, with mean 512 and standard deviation 8, we require P(Randomly chosen value < 500).

Many calculators can display such information for normally distributed data.

The required probability is 0.0668.

The probability that a randomly chosen box of this cereal contains less than 500 grams is 0.0668.

b In any batch of boxes of this cereal we should expect that the proportion of them that contain less than 500 grams is about 0.07. Thus in a random sample of 100 boxes of this cereal we would expect approximately 7 boxes to contain less than 500 g.

Using a calculator

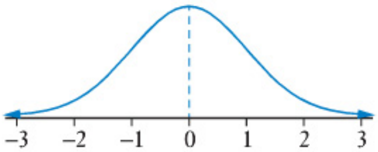
The various calculators have different capabilities and routines with regard to displaying probabilities for normally distributed sets of data.

You will gain familiarity with the ability of *your* calculator in this regard in the next exercise.

In the old days – using a book of tables

Prior to the ready availability of calculators with built-in statistical routines for determining probabilities associated with normal distributions, these probabilities were determined using books of statistical tables.

These books give probabilities for just one normal distribution, the **standard normal distribution**. For this the random variable has a mean of 0 and a standard deviation of 1, as shown on the right.



Normal distributions having means and standard deviation not equal to these standard values needed to be standardised. We encountered this idea of standardising data by expressing it as a number of standard deviations above or below the mean at the beginning of this chapter. Calling the original score an '*x* score' and the standardised score a '*z* score' we have:

$$z \text{ score} \quad = \quad \frac{x \text{ score} - \text{mean of } x \text{ scores}}{\text{standard deviation of } x \text{ scores}}$$

Thus before the ready availability of sophisticated calculators, to answer the previous example which required us to determine the probability that from a normally distributed set of data, *X*, with mean 512 and standard deviation 8, a randomly selected item would have a value less than 500 we would have changed the 500 to a standard score:

$$\begin{aligned}\text{standard score} &= \frac{500 - 512}{8}\\ &= -1.5\end{aligned}$$

(i.e. a score of 500 is 1.5 standard deviations below the mean)

and then used the table of probabilities for the standard normal distribution to determine the required probability.

$$\begin{aligned}P(X < 500) &= P(Z < 1.5)\\ &= 0.0668\end{aligned}$$

z	**0.00**	**0.01**	**0.02**	**0.03**
–1.9	0.0287	0.0281	0.0274	0.0268
–1.8	0.0359	0.0351	0.0344	0.0336
–1.7	0.0446	0.0436	0.0427	0.0418
–1.6	0.0548	0.0537	0.0526	0.0516
–1.5	0.0668	0.0655	0.0643	0.0630

Thus, as before, the probability that a randomly chosen box of the cereal contains less than 500 grams is 0.0668.

ISBN 9780170395205

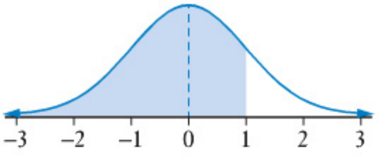
Exercise 4B

Get to know the capabilities of your calculator with regard to normal probability distributions.

The questions of this exercise refer to data sets involving normally distributed scores, X.

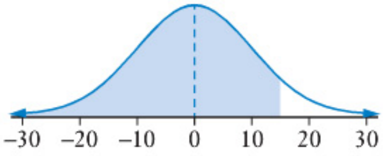
Using your calculator make sure that you can obtain each of the probabilities given in questions **1** to **8** below (correct to 4 decimal places), and each value of k in questions **9** to **17**.

1 mean = 0
standard deviation = 1



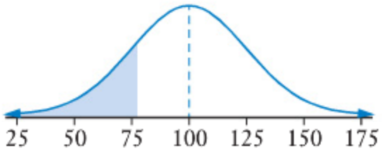
$P(X < 1) = 0.8413$
Can you also get 0.84 using the 68%, 95%, 99.7% rule?

2 mean = 0
standard deviation = 10



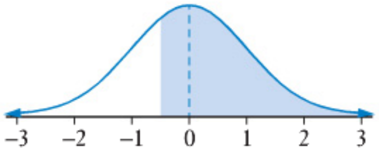
$P(X < 15) = 0.9332$

3 mean = 100
standard deviation = 25



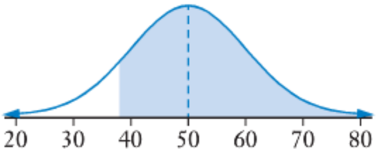
$P(X < 78) = 0.1894$

4 mean = 0
standard deviation = 1



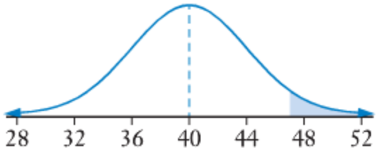
$P(X > -0.5) = 0.6915$

5 mean = 50
standard deviation = 10



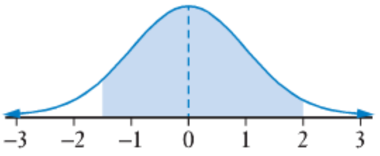
$P(X > 38) = 0.8849$

6 mean = 40
standard deviation = 4



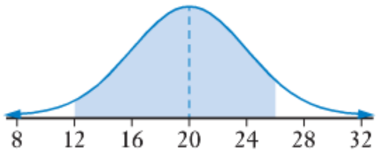
$P(X > 47) = 0.0401$

7 mean = 0
standard deviation = 1



$P(-1.5 < X < 2) = 0.9104$

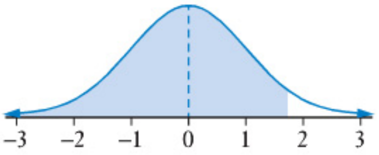
8 mean = 20
standard deviation = 4



$P(12 < X < 26) = 0.9104$

9

mean = 0
standard deviation = 1

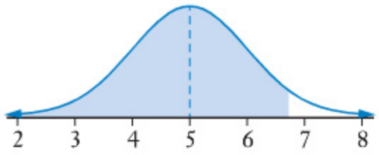


$P(X < k) = 0.9573$

$\therefore \quad k = 1.72$ (2 decimal places)

10

mean = 5
standard deviation = 1

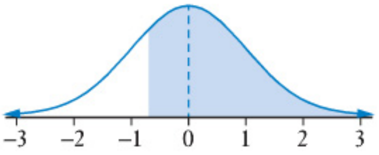


$P(X < k) = 0.9671$

$\therefore \quad k = 6.84$ (2 decimal places)

11

mean = 0
standard deviation = 1

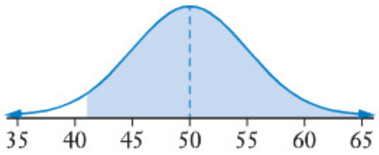


$P(X > k) = 0.7517$

$\therefore \quad k = -0.68$ (2 decimal places)

12

mean = 50
standard deviation = 5

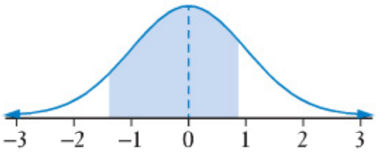


$P(X > k) = 0.9656$

$\therefore \quad k = 40.9$ (1 decimal place)

13

mean = 0
standard deviation = 1

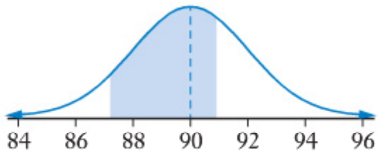


$P(-1.4 < X < k) = 0.7215$

$\therefore \quad k = 0.85$ (2 decimal places)

14

mean = 90
standard deviation = 2



$P(87.2 < X < k) = 0.5964$

$\therefore \quad k = 90.92$
(2 decimal places)

15

mean = 40
standard deviation = 15
$P(X < k) = 0.9850$
$\therefore k = 72.55$ (2 decimal places)

16

mean = 10
standard deviation = 0.5
$P(X > k) = 0.0721$
$\therefore k = 10.73$
(2 decimal places)

17

mean = 0.1
standard deviation = 0.01
$P(0.08 < X < k) = 0.3036$
$\therefore k = 0.0955$ (4 decimal places)

Notation

If we use X to represent the possible values of a normally distributed set of measurements having a mean μ and standard deviation σ (and hence variance σ^2) this is sometimes written:

$$X \sim \mathrm{N}(\mu, \sigma^2).$$

With μ (mu) pronounced 'myew', this is read as:

X is normally distributed with mean myew and standard deviation sigma.

EXAMPLE 4

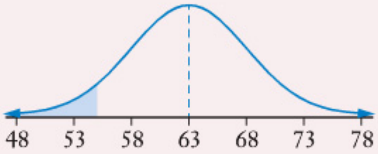
If $X \sim N(63, 25)$ determine $P(X < 55)$.

Solution

X is normally distributed with a mean of 63 and a standard deviation of 5.

Using a calculator:

$P(X < 55) \;=\; 0.0548$



Using a tables book:

$$\begin{aligned} P(X < 55) &= P(Z < -1.6) \\ &= 0.0548 \end{aligned}$$

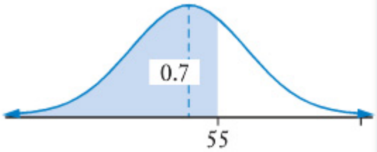
(Shown for interest only.)

Quantiles

Quantiles are the values which a particular proportion of the distribution falls below.

Thus if 0.7 (70%) of the distribution is below 55 then 55 is the 0.7 quantile.

Alternatively we can refer to 55 as being the 70th **percentile**.



WS Normal distribution – Worded problems

WS Normal distribution – Worded problems 2

Note:
- We are already accustomed to referring to the 0.25 quantile as the first, or lower, **quartile** and the 0.75 quantile as the third, or upper, quartile.
- If the quartiles divide a distribution in to four equal parts and the percentiles divide the distribution into 100 equal parts, what might deciles and quintiles do?

EXAMPLE 5

If $X \sim N(20, 3^2)$ determine

a the 0.5 quantile,

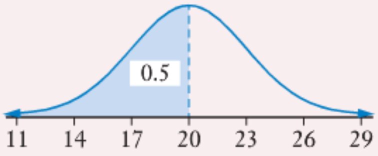
b the 0.82 quantile,

c the 24th percentile,

d the 62nd percentile.

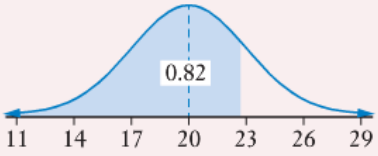
Solution

a



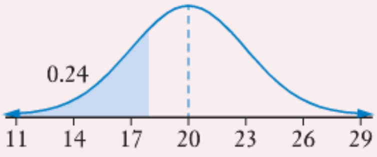
By inspection:
The 0.5 quantile is 20.

b



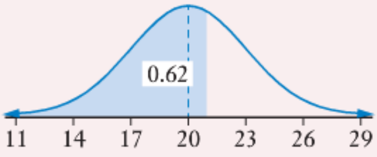
Using a calculator:
The 0.82 quantile is 22.7.

c



Using a calculator:
The 24th percentile is 17.9.

d



Using a calculator:
The 62nd percentile is 20.9.

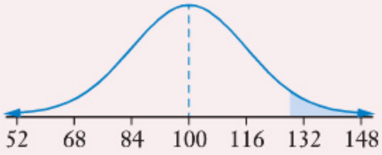
EXAMPLE 6

Eight thousand, two hundred and forty students were given an IQ test. The scores were normally distributed with a mean of 100 and a standard deviation of 16.

a Determine how many of the students, to the nearest ten, achieved a score in excess of 128.

b What were the minimum and maximum scores of the middle 60% of students on this test?

Solution



a Let X be the scores obtained in the test.

Thus $X \sim \mathrm{N}(100, 16^2)$.

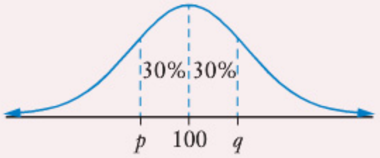
We require $\mathrm{P}(X > 128)$.

Using a calculator, $\mathrm{P}(X > 128) = 0.0401$.

Number scoring more than 128:

$$0.0401 \times 8240 \approx 330$$

Approximately 330 students achieved a score in excess of 128.



b If p is the lowest score achieved by the middle 60% then $\mathrm{P}(X < p) = 0.2$ i.e. $p = 86.53$

and if q is the highest score achieved by the middle 60% then $\mathrm{P}(X < q) = 0.8$ i.e. $q = 113.47$

(Some calculators can determine p and q more directly for this symmetrical situation.)

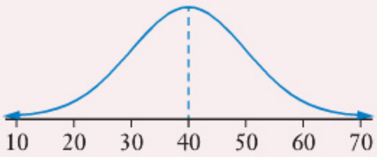
The lowest and highest scores achieved by the middle 60% are 86.5 and 113.5 respectively (to the nearest half-mark).

EXAMPLE 7

If $X \sim \mathrm{N}(40, 10^2)$ determine each of the following probabilities using the 68%, 95%, 99.7% rule, and *not* the statistical capability of your calculator.

a $\mathrm{P}(30 < X < 50)$ **b** $\mathrm{P}(20 < X < 60)$ **c** $\mathrm{P}(40 < X < 60)$ **d** $\mathrm{P}(X \le 50)$

Solution



a 30 is one standard deviation below the mean and 50 is one standard deviation above the mean.

Thus $\mathrm{P}(30 < X < 50) = 0.68$

b $\mathrm{P}(20 < X < 60) = 0.95$

c $$\mathrm{P}(40 < X < 60) = \frac{0.95}{2}$$
$$= 0.48 \text{ (correct to 2 decimal places)}$$

d $$\mathrm{P}(X \le 50) = 0.5 + \frac{0.68}{2}$$
$$= 0.84$$

As the previous chapter explained, we make no distinction between $\mathrm{P}(X \le 50)$ and $\mathrm{P}(X < 50)$. Including the line or not makes no difference to the area of the region.

EXAMPLE 8

Let us suppose that the time from Simon getting out of bed until his arrival at school is normally distributed with a mean of 55 minutes and a standard deviation of 5 minutes. Simon's arrival at school is classified as being late if it occurs after 9:10 a.m.

a One day Simon gets out of bed at 8:08 a.m. What is the probability of him arriving late?

b For a period of time Simon always gets out of bed at the same time but finds that he arrives late approximately 85% of the time! What time is he getting out of bed (to the nearest minute)?

Solution

a Let T minutes be the time from getting out of bed until arrival at school.

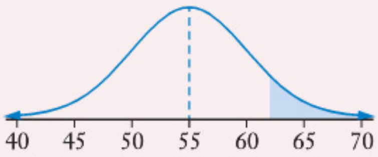
Thus $T \sim \mathrm{N}(55, 5^2)$.

Simon has 62 minutes to get to school before he is late.

We require: $\mathrm{P}(T > 62)$

Calculator gives: $\mathrm{P}(T > 62) = 0.0808$.

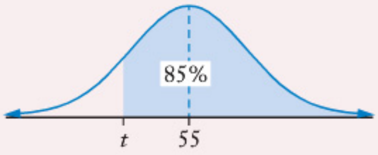
If Simon gets out of bed at 8:08 a.m. the probability of him arriving late is 0.0808.



b The time that Simon is allowing himself to get to school is causing him to be late approximately 85% of the time.

We require t for which $\mathrm{P}(T > t) = 0.85$.

Calculator gives: $t \approx 49.8$



Thus Simon is allowing approximately 50 minutes to get to school and for 85% of the days the journey takes longer than this, causing him to be late 85% of the time.

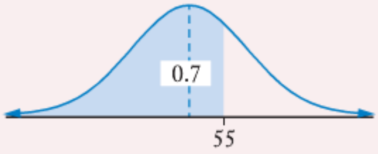
Simon is getting out of bed at 8:20 a.m.

EXAMPLE 9

The continuous random variable, X, is normally distributed with $\mathrm{P}(X < 55) = 0.7$.

a How many standard deviations from the mean is a score of 55?

b If the standard deviation of X is 4 find the mean of the distribution, giving your answer correct to one decimal place.

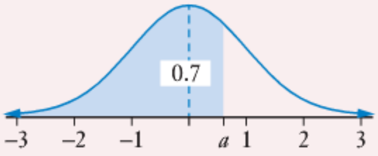


Solution

a The standard normal distribution shows standard deviations from the mean (see diagram).

Thus $Z \sim \mathrm{N}(0, 1)$ and $\mathrm{P}(Z \leq a) = 0.7$

From calculator $a = 0.5244$



55 is 0.524 standard deviations from the mean.

b $\text{Mean} + 0.524 \times 4 = 55$

$\therefore \quad \text{Mean} = 52.9$ correct to one decimal place.

EXAMPLE 10

Pre-bagged packs of bananas are marked as 'contains approximately 2 kg'. Let us suppose that in fact that the weight of the contents of such bags are normally distributed with mean 2.015 kg and standard deviation 0.01 kg.

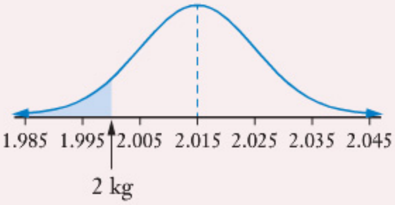
Shutterstock.com/Chubykin Arkady

a What is the probability that the contents of a randomly chosen bag of these bananas has a weight of less that 2 kg?

b If five such bags are randomly selected what is the probability that three or more will have contents weighing less than 2 kg?

Solution

a Let the weight of the bags be represented by the continuous random variable X.



Thus $X \sim \mathrm{N}(2.015, 0.01^2)$.

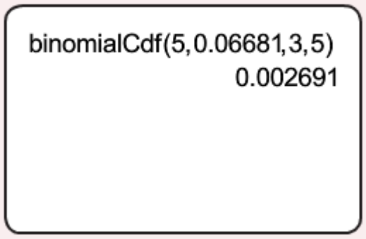
Calculator gives $\quad \mathrm{P}(X < 2) = 0.066\,807\,2$

The probability that the contents of a randomly chosen bag of the bananas having a weight of less than 2 kg is 0.0668 (correct to 4 decimal places).

b We now consider a binomial distribution Y with $Y \sim \mathrm{Bin}(5, 0.066\,81)$.

We require $\mathrm{P}(Y \geq 3)$.

By calculator, as shown below:



```
binomialCdf(5,0.06681,3,5)
                  0.002691
```

Or as shown in the working below,

$$\begin{aligned}\mathrm{P}(Y \geq 3) &= {}^5C_3\,(0.066\,81)^3\,(1 - 0.066\,81)^2 + {}^5C_4\,(0.066\,81)^4\,(1 - 0.066\,81)^1 + (0.066\,81)^5 \\ &= 0.0027 \text{ (correct to 4 decimal places).}\end{aligned}$$

Thus if five of the bags are randomly selected the probability that three or more will have contents weighing less that 2 kg is 0.0027.

ISBN 9780170395205

The normal distribution pdf

Though not something you necessarily need to know, but included here for interest, is the fact that the probability density function, $f(x)$, for a normal distribution with a mean of μ and standard deviation σ, i.e. $X \sim \mathrm{N}(\mu, \sigma^2)$, is

$$f(x) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}}e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for } -\infty < x < \infty.$$

Thus for the standard normal distribution, $X \sim \mathrm{N}(0, 1^2)$

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi}}e^{-0.5x^2} \qquad \text{for } -\infty < x < \infty.$$

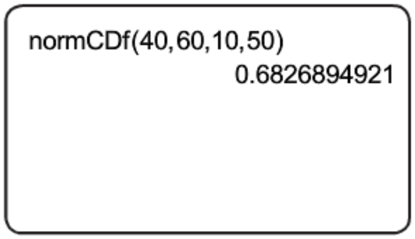
We would then expect $\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-0.5x^2}\,dx \quad = \quad 1.$

Confirm this result on your calculator but note that $f(x)$ is not readily integrated by any methods we have encountered. Instead, and beyond the scope of this unit, your calculator may use a numerical method, and may give an answer very close to 1, due to the numerical approximation involved in the method.

Taking, as an example, a normally distributed random variable X, with mean 50 and a standard deviation of 10, i.e. $X \sim \mathrm{N}(50, 10^2)$, we can (with the assistance of a calculator) check that the formula given at the top of this page does give an answer consistent with our understanding that approximately two-thirds of the population lie within one standard deviation of the mean:

$$\begin{aligned} \mathrm{P}(40 \le X < 60) \quad &= \quad \int_{40}^{60} \frac{1}{10\sqrt{2\pi}}e^{-0.5\left(\frac{x-50}{10}\right)^2}\,dx \\ &\approx \quad 0.683. \end{aligned}$$

Fortunately, and as we have already seen, we do not have to formulate this definite integral each time, instead obtaining the answer from the ability of many calculators to output such values for normal distributions.

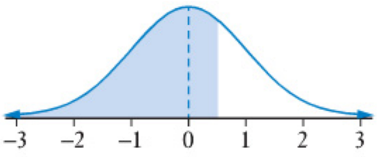


```
normCDf(40,60,10,50)
                0.6826894921
```

Exercise 4C

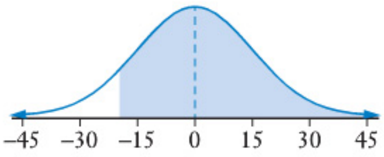
Questions **1** to **8** of this exercise refer to data sets that involve normally distributed scores, X.

1 mean = 0
standard deviation = 1



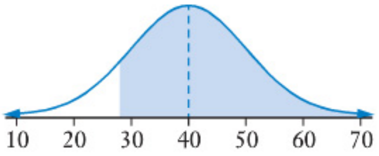
$P(X < 0.5) = k$
Find k correct to 4 decimal places.

2 mean = 0
standard deviation = 15



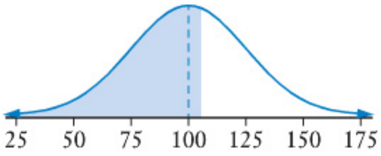
$P(X > -20) = k$
Find k correct to 4 decimal places.

3 mean = 40
standard deviation = 10



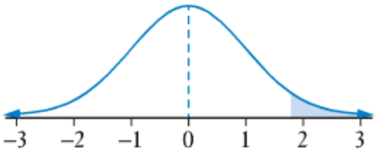
$P(X > 28) = k$
Find k rounded to 4 decimal places.

4 mean = 100
standard deviation = 25



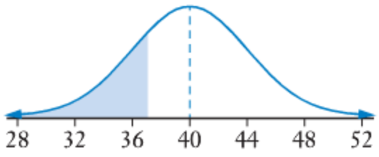
$P(X < 105) = k$
Find k rounded to 4 decimal places.

5 mean = 0
standard deviation = 1



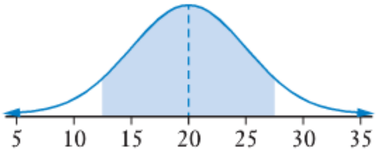
$P(X > k) = 0.0418$
Find k rounded to 2 decimal places.

6 mean = 40
standard deviation = 4



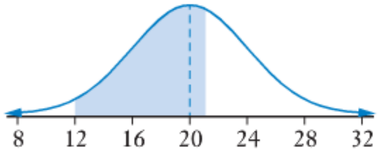
$P(X < k) = 0.2776$
Find k rounded to 2 decimal places.

7 mean = 20
standard deviation = 5



$P(20 - k < X < 20 + k) = 0.8684$
Find k rounded to 2 decimal places.

8 mean = 20
standard deviation = 4



$P(12 < X < k) = 0.6$
Find k rounded to 2 decimal places.

9 The random variable, X, is normally distributed with a mean of 12 and a standard deviation of 2, i.e. $X \sim N(12, 2^2)$. Determine $P(X \geq 13.5)$.

10 The random variable, X, is normally distributed with a mean of 240 and a variance of 400, i.e. $X \sim N(240, 20^2)$. Determine $P(218 < X < 255)$.

11 $X \sim N(62, 64)$, i.e. X, is normally distributed with a mean of 62 and a standard deviation of 8. Given that $P(X > k) = 0.8238$ determine k.

12 If $X \sim N(0, 1)$ determine

a the 0.72 quantile,

b the 0.26 quantile,

c the 89th percentile,

d the 23rd percentile.

13 If $X \sim N(20, 3^2)$ determine

a the 0.44 quantile,

b the 0.74 quantile,

c the 33rd percentile,

d the 85th percentile.

14 Using the 68%, 95%, 99.7% rule, and *not* the statistical capability of your calculator, determine the following probabilities.

a $P(-1 < X < 1), X \sim N(0, 1^2)$.

b $P(-2 < X < 2), X \sim N(0, 1^2)$.

c $P(-3 < X < 3), X \sim N(0, 1^2)$.

d $P(8 < X < 32), X \sim N(20, 6^2)$.

e $P(4 < X < 16), X \sim N(10, 2^2)$.

f $P(0 < X < 1), X \sim N(0, 1^2)$.

g $P(X < 1), X \sim N(0, 1^2)$.

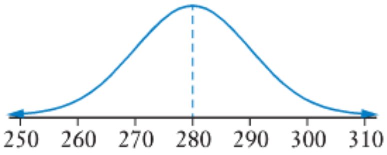
h $P(X > 1), X \sim N(0, 1^2)$.

i $P(X < 5), X \sim N(0, 5^2)$.

j $P(X > 70), X \sim N(60, 10^2)$.

15 Let us suppose that the duration of pregnancy, for a naturally delivered human baby, is a normally distributed variable with a mean of 280 days and a standard deviation of 10 days.

Using the 68%, 95%, 99.7% rule, and *not* the statistical capability of your calculator, determine estimates for the following.



a The percentage of human pregnancies, for naturally delivered babies, that are between 250 days and 310 days.

b The percentage of human pregnancies, for naturally delivered babies, that exceed 290 days.

c The percentage of human pregnancies, for naturally delivered babies, that are between 260 days and 270 days.

ISBN 9780170395205

16 A machine produces components whose weights are normally distributed with a mean of 500 g and standard deviation of 5 g.

a According to the 68%, 95%, 99.7% rule, what percentage of the components will have a weight of less than 495 g?

b According to the 68%, 95%, 99.7% rule, what percentage of the components will have a weight of less than 490 g?

17 A box of breakfast cereal has ‘contains 300 grams of breakfast cereal’ printed on it. Suppose that in fact the weight of breakfast cereal contained in these boxes is normally distributed with a mean of 310 grams and a standard deviation of 4 grams. Determine the probability that a randomly chosen box of this cereal contains

a more than 312 grams of breakfast cereal,

b less than 300 grams of breakfast cereal.

18 The lengths of adult male lizards of a particular species are thought to be normally distributed with a mean of 17.5 cm and a standard deviation of 2.5 cm.

Determine the probability that a randomly chosen adult male lizard of this species will have a length

a less than 17.5 cm,

b between 15 cm and 17.5 cm.

19 The scaled scores in a national mathematics test are normally distributed with a mean of 69 and a standard deviation of 12.

What is the probability that a randomly selected candidate who sat this test has a scaled score of

a more than 75?

b between 66 and 75?

c less than 45?

20 The heights of fully grown plants of a certain species are normally distributed with a mean of 30 cm and a standard deviation of 4 cm. If 100 fully grown plants of this species are randomly selected, approximately how many would you expect to be:

a taller than 35 cm?

b shorter than 25 cm?

c between 25 cm and 30 cm in height?

iStock.com/motorolka

21 Let us suppose that 44 mg is 110% of the recommended daily intake of a particular vitamin and that a 110 mL container of fruit juice contains approximately 44 mg of this vitamin. If in fact the weight of the vitamin in the 110 mL containers of the fruit juice is normally distributed with mean 44 mg and standard deviation 2.5 mg, determine the probability that a randomly chosen 110 mL container of this fruit juice contains less than the recommended daily intake of the vitamin.

22 Five thousand, five hundred and forty-two students sat a particular leaving exam that was marked out of 120. The scores obtained could be well modelled by a normal distribution with a mean of 62 and a standard deviation of 12.5.

a Distinction certificates were awarded to students who gained a mark of 80 or more. How many students gained distinction certificates?

b A mark of less than 40 was regarded as a fail. How many of the students failed?

c What were the minimum and maximum scores of the middle 40% of students on this test?

23 Let us suppose that the heights of the adults of a particular country are normally distributed with a mean of 1.75 m and a standard deviation of 10 cm. A car manufacturer wishes to design a new car with the space allowed for the driver, and the 'travel' on the drivers seat, suitable for every adult in the population except the tallest 5% of the adult population and the shortest 5% of the adult population. What is the height of the shortest driver and the tallest driver that the manufacturer is attempting to allow for? (Answer to nearest half-centimetre.)

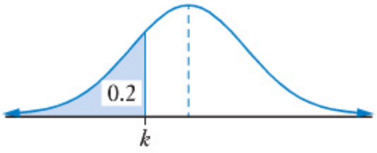
24 The marks achieved in a particular exam are normally distributed with a mean of 64 and a standard deviation of 12.

Grades are to be awarded as follows:

Top	12%	of candidates:	Grade A
Next	25%	of candidates:	Grade B
Next	40%	of candidates:	Grade C
Next	15%	of candidates:	Grade D
Remainder		of candidates:	Grade F

Determine the marks that form the A/B, B/C, C/D, and D/F grade boundaries, giving your answers correct to the nearest whole number.

25 The continuous random variable, X, is normally distributed with $\mathrm{P}(X < k) = 0.2$.



a How many standard deviations from the mean is k?

b If $k = 40$, and the standard deviation of X is 5, find the mean of the distribution, giving your answer correct to one decimal place.

26 Let us suppose that the time, in minutes, from Monica leaving home until she arrives at work is a normally distributed random variable with a mean of 45 and a standard deviation of 5. Monica's arrival at work is classified as late if it occurs after 8:30 a.m.

a One day Monica leaves home at 7:40 a.m. What is the probability of her arriving late?

b For a period of time Monica leaves home at the same time each day. During this period she finds that she arrives late approximately 8% of the time. What time is she leaving home (to the nearest minute)?

c What is the latest time (involving whole minutes) that Monica should leave home each day if she wishes to cut her late arrivals to less than 1%?

27 The annual rainfall in an area in the south west of Western Australia is normally distributed with a mean of 1200 mm and a standard deviation of 200 mm.

According to this model, and assuming the situation does not change, in every one hundred years how many years would you expect the annual rainfall to be

a less than 800 mm?

b more than 1500 mm?

c between 800 mm and 1500 mm?

d Given that a year has an annual rainfall of more than 1300 mm, what is the probability that the rainfall for the year is less than 1500 mm?

28 The weight of each apple harvested from a particular orchard determines where the apple will be sent:

If	weight of apple $\geq$ 250 g	send to premium outlet
	150g < weight of apple < 250 g	send to normal market
	weight of apple $\leq$ 150 g	send for juicing.

The weights of the apples are normally distributed with mean 180 g and standard deviation 40 g.

a In a random sample of 1000 apples how many would you expect to go to the premium outlet?

b Given that an apple does not go to the premium outlet, what is the probability that it is sent for juicing?

29 Bags of tomatoes are marked as
Contains approximately 1 kg tomatoes.

An analysis indicated that the weights of the tomatoes in the bags were approximately normally distributed with a mean of 1018 grams and standard deviation 10 grams.

Based on this normal distribution, what is the probability that a randomly chosen bag will contain tomatoes weighing

a less than 1 kg?

b at least 25 grams over 1 kg?

c If ten bags were randomly chosen, what is the probability that at least one would weigh less than 1 kg? (Give this answer correct to two decimal places.)

Shutterstock.com/iLight photo

30 Let us suppose that a particular IQ test gives results that are normally distributed with a mean score of 100 and a standard deviation of 15.

a If a randomly chosen person has a score on this test that is greater than 125, what is the probability their score is greater than 135?

b If five people were randomly selected, what is the probability that at least three would have a score greater than 120?

ISBN 9780170395205

31 Part of a breakfast cereal packing process involves a machine sending x grams of the cereal into each packet. The value of x is programmed into the machine and the packets are then filled with the weight of cereal in each being normally distributed with a mean of x grams and a standard deviation of 1.8 grams.

iStock.com/artproem

a The machine is used to fill packets that will be labelled as containing 500 grams. What should be the value of x, rounded up to the next gram, if the company wants no more than 0.5% of the packets to be underweight?

b The machine is used to fill packets that will be labelled as containing 250 grams. What should be the value of x, rounded up to the next gram, if the company wants no more than 0.5% of the packets to be underweight?

32 A machine produces components whose weights are normally distributed with a mean of 500 g and standard deviation of 5 g.

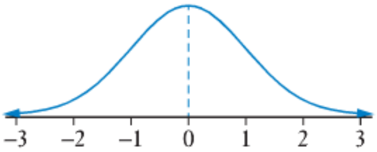
a What percentage of these components will have a weight of less than 490 g?

A new machine is being developed to produce these components more uniformly. The intention is for this new machine to produce components whose weights are normally distributed with a mean of 500 g and just 1.5% having a weight of less than 490 g.

b What does the standard deviation of the weights of components made by this new machine need to be?

Using the normal distribution to model data

The normal distribution is an extremely useful distribution and is used to model many naturally occurring random variables, for example, the blood pressures of the adult male population or adult female population of a country.



However, given a set of collected data, how would we know if it was appropriate to model the distribution of the data as a normal distribution?

One way to check the appropriateness would be to view the histogram of the distribution to see if it has the characteristic bell shaped curve.

We can also check to see if the data set gives proportions of data values lying within particular ranges similar to the proportions we would expect from a normally distributed random variable.

For example, suppose we have a distribution of scores with mean 17.2 and standard deviation 2.1. If this distribution were normally distributed we would expect approximately 68% of the scores to lie within one standard deviation of the mean.

$$\text{i.e. for } X \sim \mathrm{N}(17.2, 2.1^2),\ \mathrm{P}(15.1 < X < 19.3) \approx 0.68$$

If, for our data set, the proportion was not close to 68% we would question the wisdom of modelling the distribution as a normal distribution.

ISBN 9780170395205

Can we use the normal distribution to model discrete data?

Whilst the normal distribution is for continuous data it can be used to model discrete data if we make a 'correction, or adjustment, for continuity'.

To determine $\mathrm{P}(8 \le X \le 10)$ for a discrete distribution of integers X we would determine

$$\mathrm{P}(7.5 < Y < 10.5)$$

where Y is a suitably chosen continuous variable.

Similarly	$\mathrm{P}(X < 50)$	would become	$\mathrm{P}(Y < 49.5)$	on a continuous model,
	$\mathrm{P}(X \le 50)$	would become	$\mathrm{P}(Y < 50.5)$	on a continuous model,
and	$\mathrm{P}(8 < X < 10)$	would become	$\mathrm{P}(8.5 < Y < 9.5)$	on a continuous model.

Limitations of probability models for predicting real behaviour

If we model a particular probabilistic situation as normal, binomial, uniform, etc. we must remember that the model is still only a model. The real situation may not fit the model exactly.

Our data may not be truly representative of the real situation.

The model may fit the situation generally but be 'a bad fit' in specific circumstances.

Influences and events that we may assume to be random may not be totally random.

Etc.

Thus if we record data and then model the distribution of this data as being of a particular type and with particular characteristics, any predictions we make based on the model need to be viewed with some caution. For example, suppose we were to collect data from adult female Australians and suppose this data suggested it appropriate to model the blood pressures of adult Australian females as being normally distributed, with a particular mean and standard deviation This may not be an appropriate model to use if we were applying it to a group of super fit adult females or to a group of elderly adult females or to adult females from other countries or if our data was based on a small sample, or a biased sample or … .

Exercise 4D

Explain why each of the situations in numbers **1** to **5** would cause you to question the wisdom of choosing the normal distribution as a model for the data.

1 A data set has 80% of its data points within one standard deviation of the mean.

2 A data set involved 360 measurements of continuous data with mean 4.37 and standard deviation 2.52. Fifty-two of the 360 measurements were greater than 8.

3 A data set involved 826 measurements of a continuous variable with mean 8.9 and standard deviation 3.1.

Two hundred and fifty-seven of the measurements were within one standard deviation of the mean.

4 A data set involved 409 measurements of a continuous variable.

One hundred and five of the measurements were between the mean and one standard deviation below the mean.

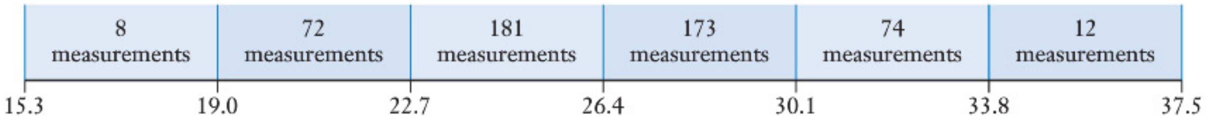
One hundred and eighty of the measurements were between the mean and one standard deviation above the mean.

5 A data set involved 180 measurements with mean 105.2 and standard deviation 31.4.

Ten of the measurements were at least two standard deviations above the mean.

None of the measurements were more than two standard deviations below the mean.

6 A scientist analyses the 520 measurements made of a particular continuous variable, X. The scientist finds that the measurements have a mean of 26.4, a standard deviation of 3.7 and can be grouped as follows:



The scientist decides to use $X \sim \mathrm{N}(26.4, 3.7^2)$ to predict probabilities if the number of measurements in the range

mean → mean + 1 st. devn.	mean – 1 st. devn. → mean
mean → mean + 2 st. devns	mean – 2 st. devns → mean
mean → mean + 3 st. devns	mean – 3 st. devns → mean

are all within 3% of the numbers predicted by $X \sim \mathrm{N}(26.4, 3.7^2)$ (This is not any standard test, just one she decides to apply.)

a Will she use $X \sim \mathrm{N}(26.4, 3.7^2)$?

b Does her 3% rule ensure that, if it is met, any predicted probabilities will be sufficiently reliable?

7 A set of data has the following summary statistics:

mean = 44.7 median = 44.9 standard deviation = 26.4

lower quartile = 19.3 upper quartile = 69.2

State, with reasoning, whether these figures suggest the data is normally distributed or not.

Shutterstock.com/Delpixel

8 For large values of n the binomial probability distribution $X \sim \text{Bin}(n, p)$ can be well modelled by the normal distribution $Y \sim \text{N}(np, np(1-p))$.

I.e., for large n, binomial probabilities involving n trials, with the probability of success at each trial being p, can be well modelled by a normal distribution with mean np and standard deviation $\sqrt{np(1-p)}$, with appropriate adjustments for continuity being made. (See page 96 for an explanation of adjustment for continuity.)

a Use you calculator to determine $P(X \leq 65)$ for $X \sim \text{Bin}(100, 0.6)$
and $P(Y \leq 65.5)$ for $Y \sim \text{N}(60, 24)$
giving each answer correct to four decimal places.

b Use your calculator to determine $P(X = 100)$ for $X \sim \text{Bin}(200, 0.5)$
and $P(99.5 \leq Y \leq 100.5)$ for the appropriate normal distribution model, giving each answer correct to four decimal places.

c Use your calculator to determine $P(20 < X \leq 25)$ for $X \sim \text{Bin}(50, 0.6)$
and the equivalent probability on the appropriate normal distribution model, giving each answer correct to four decimal places.

Miscellaneous exercise four

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

Differentiate each of numbers **1** to **6** with respect to x. For some it may be advisable to use the laws of logarithms before differentiating.

1 $y = \ln(10x)$

2 $y = 10\ln x$

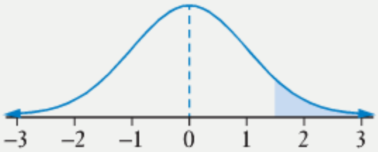
3 $y = \dfrac{x}{\ln x}$

4 $y = \log_e\left[(x^2+1)^3\right]$

5 $y = \ln\left[\dfrac{(x-1)^3}{x+1}\right]$

6 $y = \log_5 x$

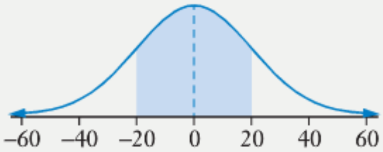
7 Normal distribution,
Mean = 0,
Standard deviation = 1.



$P(X > 1.5) = k$

Find k correct to 4 decimal places.

8 Normal distribution,
Mean = 0,
Standard deviation = 20.

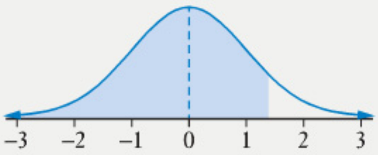


$P(-20 < X < 20) = k$

Find k correct to 4 decimal places.

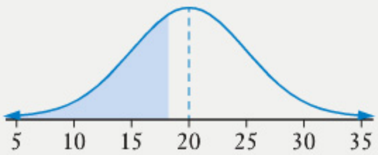
ISBN 9780170395205

9 Normal distribution,
Mean = 0,
Standard deviation = 1.



$P(X < k) = 0.9066$
Find k rounded to 2 decimal places.

10 Normal distribution,
Mean = 20,
Standard deviation = 5.



$P(X < k) = 0.3632$
Find k rounded to 2 decimal places.

Find the exact gradient of each of the following at the given point on the curve.

11 $y = 3\ln x$ at $(e, 3)$.

12 $y = x\ln x$ at (e, e).

13 The random variable, X, is normally distributed with a mean of 50 and a standard deviation of 3, i.e. $X \sim N(50, 3^2)$. Determine $P(X \geq 58)$.

14 If $X \sim N(0, 1)$ determine, to three decimal places:

a the 0.42 quantile,

b the 0.13 quantile,

c the 63rd percentile,

d the 41st percentile.

15 Let us suppose that the daily rainfall in a region is normally distributed with a mean of 11.2 mm and a standard deviation of 3.1 mm.

In a year of 365 days, how many days would you expect this region to have a rainfall that is

a less than 6 mm?

b more than 10 mm?

c between 10 mm and 15 mm?

Shutterstock.com/MK photograp55

16 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} ax^2 + k & \text{for} \quad 0 \leq x \leq 2 \\ 0 & \quad\quad\ \text{elsewhere.} \end{cases}$$

If $P(X \leq 1) = 0.2$, determine a and k.

Hence find $E(X)$, the expected value of X, and $\mathrm{Var}(X)$, the variance of X.

17 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} k\sin x & \text{for} \quad 0 \le x \le \pi \\ 0 & \quad\quad\quad \text{elsewhere.} \end{cases}$$

Determine **a** the value of k, **b** $P\left(\frac{\pi}{4} \le X \le \frac{3\pi}{4}\right)$.

18 A particular industrial process involves the extraction of a valuable metal from rock deposits containing the metal. The rock is mined and then processed in an extraction unit in five-tonne 'batches', with each batch containing approximately 100 kg of the metal. The cost of mining and then extracting x kg of the metal from each five tonne batch is \$$C$ where

$$C \approx 25\,000 - 20\,000\log_e\left(1 - \frac{x}{100}\right),\ x < 100.$$

The company carrying out this mining and extraction process has a contract with a buyer who will buy each kilogram of the extracted metal for \$1000.

a Write down an expression for $P(x)$, the profit function.

b (For this part, first determine your answers using calculus and then view the graph of $P(x)$ on a calculator and determine the answers from the graph.)
How many kilograms should be extracted from each five-tonne batch for maximum profit, and what would this maximum profit be?

19 The continuous random variable X has the cumulative distribution function

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ \dfrac{x^2 + 3x - 4}{36} & \text{for} \quad 1 \le x \le 5 \\ 1 & \text{for} \quad x > 5. \end{cases}$$

Determine: **a** $P(X \le 2)$ **b** $P(X \ge 2)$

c $P(3 \le X \le 5)$ **d** $P(X > 7)$

20 Let us suppose that the number of kilometres a new tyre of a particular brand lasts before it needs to be replaced is normally distributed with a mean of 60 000 km and a standard deviation of 8000 km.

a Determine the probability that a randomly selected new tyre of this brand lasts for less than 45 000 km before it needs to be replaced.

b Determine the probability that when four new tyres of this brand are randomly selected at least one will need to be replaced before it has travelled 45 000 km.

iStock.com/algre

5.

Random sampling

- Sampling
- How big should our sample be?
- How should we select our sample?
- Random sampling
- Generating random numbers
- Stratified sampling
- Other forms of sampling
- Capture-recapture
- Simulations
- Random number generating from other distributions
- More simulations
- Variability of random samples
- Miscellaneous exercise five

Situation One

Taking a sample

Doctors may at times need an analysis of our blood. Fortunately they can do this by taking a sample, not the whole lot!

Whilst blood samples are commonly taken from the veins around the elbow this is not always the case.

Sometimes they are taken from the wrist,
or perhaps by a pin prick to the thumb or finger,
or maybe a heel prick.

Research each of these and write a few sentences about why each location is chosen and what in particular the blood sample may be used for.

Blood samples are not the only samples members of the medical profession may request of us. Research and write a few sentences about each of the following:

Amniocentesis.
Cerebral spinal fluid sample.
Sweat testing.
Mid-stream urine sample.

Shutterstock.com/Jmiks

Situation Two

The following situation involves a technique sometimes referred to as

Capture – Recapture

This technique has been used for many years to estimate the populations of species of animals.

To estimate the population of a certain species of frog in a particular area a scientist caught and 'tagged' 40 of the frogs and then released them back into the area. Some days later a random catch of 50 frogs of this species from the area found 4 that were tagged from the earlier capture.

Estimate the number of frogs of this species in the area.

imagefolk/Zoonar/K Bain

Sampling

The situations on the previous page involved **samples** being taken from a larger **population**.

By *population* we mean the whole amount under consideration – e.g. **all** of your blood, or **all** of the frogs in a particular area.

By *sample* we mean the subgroup of the population that we will use to make inferences about the whole population.

If we take some numerical measure for the sample, perhaps red blood cell count or average length of a frog, we could then use this **sample statistic** to estimate the equivalent measure for the population. Numerical characteristics about an entire population, for example, the average age of all Australians, the mean length of all of the frogs in a region, etc, are called **population parameters**. If the sample statistic is going to give a good indication of the equivalent population parameter it is important that the sample is typical of the population. Two aspects in particular that we need to consider are:

- How big should our sample be?
  The larger our sample the more confident we can be that any information collected about the sample will be indicative of the same information about the population.
- How should we select our sample?
  It is important that the sample is a fair reflection of the makeup of the population and as free from unwanted bias as possible.

How big should our sample be?

Suppose a farmer has 1000 sheep and wants to select a random sample for testing to monitor the likely wool and meat quality of his stock.

How many sheep should he include in his sample?

How many he should choose depends on how confident he wants to feel that the results from his sample fairly reflect the characteristics of the population of 1000. A small sample of just 5 or 6 animals for example is unlikely to give him this confidence.

Getty Images/EyeEm/Maria Greenwood

One quite common 'rule of thumb' is that, when reasonable-sized populations are involved, always choose at least 30. Less than 30 will tend to leave considerable doubt as to whether data obtained from our samples will fairly reflect the whole population.

Should he choose more than 30?

Certainly the more he has in the random sample the more confident he can be that the characteristics of the sample reflect those of the population as a whole. However he must balance this desire to be confident that the results reflect the population with other aspects, two of which are mentioned on the next page.

Note: A 'rule of thumb' is a general rule often based more on experience than precise calculation.

ISBN 9780170395205

- Just how confident does he need to be that the results from his sample will reflect the population as a whole?
  How crucial is it that any data from the sample is a good reflection of the population? Is it just to have a rough idea of the standard of the wool and meat of his sheep or is it perhaps to check for something more serious? If he was worried that some of his sheep were carrying a particularly harmful and transmittable disease he may only gain peace of mind about his animals if he has them all tested – i.e. not a sample at all.
- How much will it cost him to have each animal in the sample tested?
  If the test is expensive he may be more inclined to select a small sample and accept the uncertainty about how typical the sample is. (Though there will come a point where the sample involves so few sheep that the high level of uncertainty will mean that there was little point having any tested at all.)

How should we select our sample?

There are various methods for selecting a sample so as to reduce the likelihood of unwanted bias. Consideration of a number of these methods follows.

Random sampling

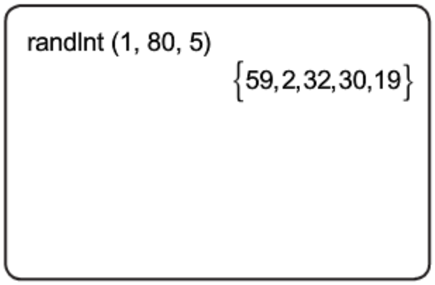
The important aspect of *random* sampling is that each member of the population has an equal chance of being chosen.

Thus to obtain a random sample from a population we could:

- Assign each member of the population a number.
  If people are involved this could simply be a list of names with each name given a number. If an area is involved, e.g. when investigating an area of grass for weed content, divide the area into equal size squares and number each square.
- Select numbers from the desired range of numbers using a random process, e.g. randomly selecting numbered balls from a bag, using random numbers from a random number generator (see below).

Generating random numbers

Some calculators can randomly generate numbers in a particular range. The display on the right for example shows 5 integers from 1 to 80 generated using the random number facility on a calculator.

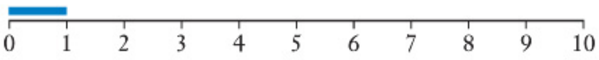


```
randInt (1, 80, 5)
                    {59,2,32,30,19}
```

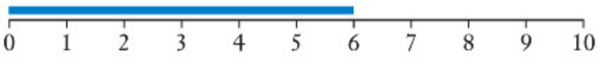
Other calculators may generate random numbers in a particular preset range, e.g. between 0 and 1. However these too can be instructed to output integer values in a desired range, as explained on the next page.

To change from a random number with an output in the range 0 to 1 to integers from the set {1, 2, 3, 4, 5, 6}:

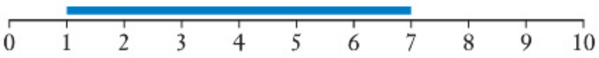
Usual output is between 0 and 1:



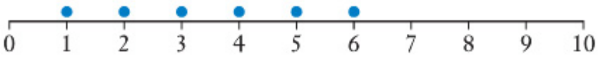
Multiply by 6 to obtain numbers between 0 and 6:

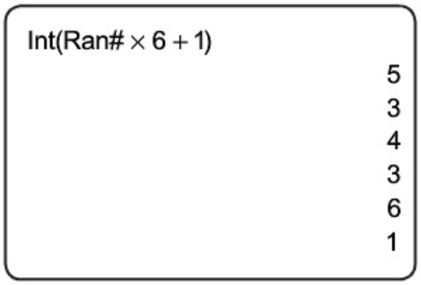


Add 1 to obtain numbers between 1 and 7:



Displaying only the integer part of such numbers will give the integers 1, 2, 3, 4, 5 or 6. (See the display below.)

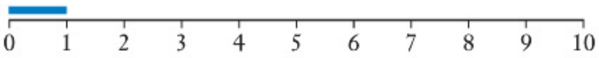




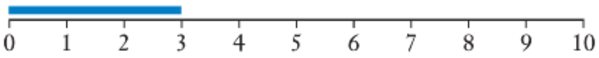
```
Int(Ran# × 6 + 1)
                        5
                        3
                        4
                        3
                        6
                        1
```

To change from a random number with an output in the range 0 to 1 to integers from the set {7, 8, 9}:

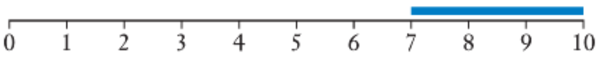
Usual output is between 0 and 1:



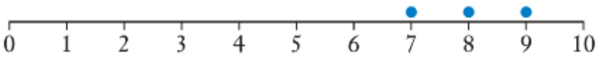
Multiply by 3 to obtain numbers between 0 and 3:

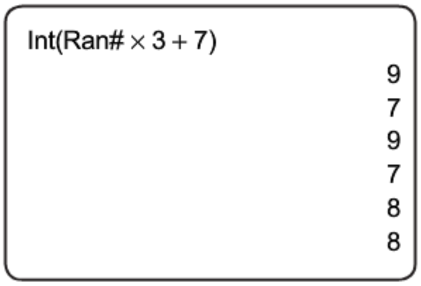


Add 7 to obtain numbers between 7 and 10:



Displaying only the integer part of such numbers will give the integers 7, 8 or 9. (See the display below.)





```
Int(Ran# × 3 + 7)
                        9
                        7
                        9
                        7
                        8
                        8
```

Use your calculator to randomly generate 5 different integers from 1 to 80.
Will you obtain the same five numbers as those shown in the display on the previous page?

Stratified sampling

In stratified sampling the population is divided up into layers, or strata, and then samples are randomly selected from each strata. For example, suppose we require a random group of 60 students from a school containing year 7 students to year 12 students. The strata could be the year levels and we then randomly select ten students from each of the 6 years.

This stratified sampling is often selected proportionally to make it more representative of the population. For example, suppose the school just mentioned had 1221 students distributed as follows:

221 in year 7, 240 in year 8, 248 in year 9,
285 in year 10, 124 in year 11, 103 in year 12.

If we want a proportional stratified sample of 60 students we choose as follows:

$$\frac{221}{1221}\times 60 \approx 10.9 \qquad \frac{240}{1221}\times 60 \approx 11.8 \qquad \frac{248}{1221}\times 60 \approx 12.2$$

11 year 7 students. 12 year 8 students. 12 year 9 students.

$$\frac{285}{1221}\times 60 \approx 14.0 \qquad \frac{124}{1221}\times 60 \approx 6.1 \qquad \frac{103}{1221}\times 60 \approx 5.06$$

14 year 10 students. 6 year 11 students. 5 year 12 students.

Other forms of sampling

Another sampling technique involves first listing the population in some order. If the sample size requires, say a 1 in 25 selection, a number between 1 and 25 is randomly selected and then every 25th person is selected after that. Thus if person number 17 is selected first then numbers 42, 67, 92, 117, …, are selected as well. This can be referred to as **systematic sampling** or **array sampling**. This method is unsuitable in cases where there is some periodic feature in the population list. For example if we were sampling components made by a machine and the work load on the machine caused it to make every 20th item faulty, (i.e., the 20th, the 40th the 60th, etc.) the above systematic selection, i.e. 17, 42, 67, … would not feature any of these defective components.

In some sampling, perhaps in an attempt to achieve stratified sampling overall, individual interviewers are given a particular number of people they must interview of various types. For example, they may be required to interview 20 people of which 5 are men aged in their twenties, 8 are married females aged over 50 and 7 are males aged over 60. This is called **quota sampling**.

Convenience sampling, as the name suggests, is when the sample is chosen because it is convenient. For example, if we wanted to collect data about primary school students the local primary school would be a convenient school to choose. Convenience sampling is sometimes used as a preliminary investigation of the situation. It is likely to be an inexpensive option compared to some others but can give some early direction to later, more involved and detailed data collection using more sophisticated sampling methods.

Television 'phone-in' surveys rely on people volunteering their opinion by phoning in. We would not expect the sample to be particularly random and the balance of the opinions expressed may be far from representative of those held by the population as a whole. This is an example of **volunteer sampling**. Those taking part select themselves to be part of the sample. This method is also called **self-selection sampling**.

Capture-recapture

Situation Two at the start of this chapter involved *capture-recapture*, a method that uses **sampling** of a population to estimate the size of the population. In both the initial capture, and again in the recapture, a sample of the whole population is taken.

In the given situation 40 frogs from an area were caught, tagged and then released back into the area. Upon their release the proportion of tagged frogs in the area is

$$\frac{40}{\text{Total number of frogs in the area}}.$$

The recapture process caught 50 frogs, of which 4 were found to be ones tagged in the earlier capture. This suggests that the proportion of tagged frogs in the area is

$$\frac{4}{50}.$$

If this second sample reasonably reflects the proportion in the whole population then

$$\frac{4}{50} \approx \frac{40}{\text{Total number of frogs in the area}}.$$

Solving this equation allows the population of frogs in the area to estimated.

Can we use capture-recapture for counting humans?

Your initial reaction to the above question might be

'Of course not! We cannot capture, tag and then release humans!'

Well in fact the capture-recapture process is used to count human populations but we do not really capture and tag people in the same way as we might do with frogs, birds or fish. With humans the first 'capture and tag' involves seeing how many of the group under consideration appear on one list and the 'recapture' is to see how many of those on this first list appear on a second list. For example:

Suppose scientists in a country where only incomplete medical records were kept of the population, wanted to estimate how many people in the capital city had suffered the amputation of a limb. Suppose that a list compiled from a number of the larger hospitals gave the identities of 150 individuals who had experienced such surgery. (These 150 amputees form the initial 'captured and tagged' group.) Hence if in the entire city there were n amputees the proportion of 'tagged' ones (i.e. the proportion appearing on the hospital listing) would be: $\frac{150}{n}$.

Now suppose we check the list of amputees attending a centre that specialises in the fitting of artificial limbs. Let us suppose that this list involved 78 people of whom 23 were also on out hospital listing (i.e. 23 of the 78 were 'tagged' from the first capture). This would indicate that the proportion of 'tagged' individuals in the entire amputee population is: $\frac{23}{78}$.

Thus if the 2nd sample is representative of the amputee population $\frac{23}{78} \approx \frac{150}{n}$,

giving an estimate of the total number in the capital city who have experienced limb amputation as approximately 510.

Note: A survey of this type was carried out for Rio de Janeiro in Brazil by Spichler at al and was published in the Pan American Journal of Public Health, 2001.

Exercise 5A

1 For each of the following state whether the sample is 'likely to introduce bias' or 'not likely to introduce bias'.

a People parking their cars at a car park are asked 'Do you think bus travel is good value?'

b People eating at *Speedy Annes* restaurant are asked 'How many times per week do you eat at a restaurant?'

c Every fourth person on a school's alphabetical roll of students is asked 'How many times do you use the school canteen in a week?'

d The colours of 2000 cars on a freeway are noted in an attempt to determine the most common colour of car.

e The heights of all the year eights in an Australian school of 1800 pupils were noted to give an indication of the average heights of Australian year eights.

2 The 497 members of a sports club comprise 100 in the under 20 age range, 154 in their twenties, 175 in their thirties and 68 aged 40 or over.

If the committee is to consist of 10 members and the age balance on the committee is to reflect the age balance of the membership, how many of each of the four age ranges should the committee consist of?

3 **a** Use the random number generator on a calculator or computer to simulate 10 rolls of a normal six-sided die. Tabulate your results and also determine the mean score from the 10 rolls and compare your results to those of others in your class.

b Repeat the above for 20 rolls, 30 rolls, 50 rolls and 100 rolls, each time comparing your results to those of others in your class.

4 A stratified sample of 80 students is to be selected from the year 8 to 12 student population of a school. If the school student population for these years consists of 420 in year eight, 407 in year nine, 389 in year ten, 270 in year eleven and 258 in year twelve how many of each year should be in the sample for it to reflect the proportion in each year group.

5 To estimate the population of a certain species of frog in a particular area a scientist caught and 'tagged' 34 of the frogs and then released them back into the area. Some days later a random catch of 28 frogs of this species from the area found 5 that were tagged from the earlier capture.

Estimate the number of frogs of this species in the area.

6 In an attempt to estimate how many fish were in a particular lake, 64 fish from the lake were netted, tagged and released back into the lake to mix with the rest of the population.

A 'recapture' carried out one week later netted 83 fish and, of these, 7 were found to carry tags indicating they were in the first netting.

According to these figures, estimate the number of fish in the lake at this time.

Auscape/Reg Morrison

7 To estimate the numbers of a particular species of bird visiting a favoured breeding ground at breeding time, scientists catch and tag 123 of the birds and then release then back into the population in the breeding ground.

A second capture catches 154 of the birds of which 6 showed the tag that indicated they were in the first catch too.

Use these figures to estimate the number of these birds in the area.

Shutterstock.com/Gerald A. DeBoer

8 Shane used the random number generator on his calculator to simulate the rolling of a normal, fair, six-sided die 12 times.

Similarly, Christine simulated the rolling of a normal, fair, six-sided die 150 times.

The results obtained by one of these students had a mean of 4.08 (maybe rounded).

The results of the other student had a mean of 3.42 (maybe rounded).

Which mean value is likely to belong to which student? Explain your reasoning.

9 Portia used the random number generator on her calculator to simulate the rolling of two normal, fair, six-sided die 12 times, each time noting the sum of the two numbers obtained.

Similarly, Horace simulated the rolling of two normal, fair, six-sided die 150 times, each time noting the sum of the two numbers obtained.

The results obtained by one of these students had a mean of 7.17 (maybe rounded).

The results of the other student had a mean of 6.42 (maybe rounded).

Which mean value is likely to belong to which student? Explain your reasoning.

10 To estimate the number of people missed from the population census carried out in a region, statisticians targeted a particular subsection of the region that they believed to be typical of the region as a whole, interviewed everyone in the subsection, and then checked how many of this targeted group featured on the census.

They found that: 1 235 067 people completed a census form.

1345 people were in the subsection and of these 1338 people had completed a census form.

Based on these figures suggest an approximation for the number in the whole region who did not complete a census form.

11 As part of an attempt to estimate the population of long-necked turtles living in a particular lakeland area, scientists intend to catch and tag a number of the turtles from the area and then release them back into the area. Explain how this can be used to estimate the population, including in your explanation what the process involves, why it works and what possible sources of error might need to be considered and, if possible, avoided?

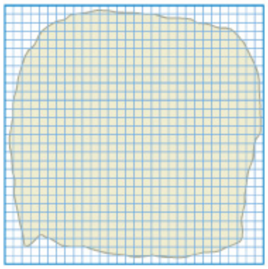
Getty Images/Ted Mead

What is a **census**?

The random numbers generated by a calculator or computer spreadsheet are sometimes referred to as **Pseudo-random numbers**. Why is this? Do some research to find out why.

Counting seals

In order to estimate the number of seals on a particular island favoured by seals, a number of aerial photographs are to be taken of the island.



A '30 × 30' grid of 900 squares is placed over a map of the entire island with each square covering 10 000 m^2 in real life (i.e. 100 m × 100 m).

220 of these squares were at the edge of the island and showed some sea or all sea, all the others were entirely covering land.

The plan was to randomly select some of the squares showing only land, to photograph these areas and to use the photographs to count the number of seals present in each of the selected squares.

Knowing that the island had a total area of 7.34 km^2 an estimate of the population of seals on the island could be made.

- Explain how we could 'randomly select' the squares to be photographed and suggest how many squares should be selected.
- Will your random selection guarantee that the sample is an accurate representation of the population of seals on the island at the time the photographs were taken? Explain.
- How could the 'seal counts' from the selected photographs be used to estimate the seal population on the island at the time the photographs were taken?
- Suggest any improvements that could be made to the plan.

Dreamstime/Johncarnemolla

ISBN 9780170395205

Simulations

Some questions in the previous exercise mentioned the idea of using a random number generator to **simulate** the rolling of a die.

Just as sampling allows us to 'get a feel for' the characteristics of a population without actually surveying every member of the population, then so running a simulation allows us to 'get a feel for' how a situation might evolve without actually running the real situation – just as a flight simulator attempts to recreate for a trainee pilot the conditions and experiences they will encounter when flying a real aircraft, without actually flying a real plane. Computer games try to make the player feel the excitement of playing a game of soccer, golf, cricket, etc without playing the real game.

A simulation of something attempts to resemble or mimic the real thing without actually being the real thing.

In mathematics we may run a simulation to collect data about some event without actually carrying out the real event. If our simulation is a good imitation of the real thing then the data collected from the simulation may help us predict what might happen in the real thing. If we need our simulation to involve some randomly occurring event we can simulate the outcome with the toss of a coin, the roll of a die or the ability of some calculators and computer spreadsheets to generate random numbers.

Simulation I: Overbooking

An airline company finds that, due to late cancellations, flights often take off with empty seats.

iStock.com/bkindler

The company considers deliberately overbooking on flights so that the cancellations will bring the flight back to capacity and reduce the number of empty seats. Should insufficient cancellations occur on any overbooked flight any passengers who have to miss the flight will be offered an alternative flight plus a refund. One particular route has a plane with a capacity of just 25 passengers. Past records indicate that on average, on this route, one in every 15 customers cancels late. The company wants to investigate the likely consequences of booking 26 passengers for the flight.

We can simulate the situation if we use a calculator or spreadsheet to randomly generate 26 integers taken from the integers 1 to 15, and take the number 15 to indicate a passenger who cancels. Run such a simulation at least ten times and see how many will result in an overbooking problem.

Of course it might be the case that on some flights there would be less than 25 booking anyway. Suppose instead we expect bookings to be in the range 15 to 27 (i.e. up to 2 overbooked) and the cancellation rate is anything from 0% to 10%.

Consider how you might run a simulation to investigate this situation.

ISBN 9780170395205

Simulation II: Spread of illness

If one student in a class of twenty-five students has an infectious illness and comes to school how many others in the class are likely to catch it?

Let us suppose that the class sits in the five-by-five layout shown below:

Dreamstime/Johncarnemolla

Let us further suppose that if one student has the illness the probability of one of his or her immediate neighbours catching the illness is one-sixth for each neighbour.

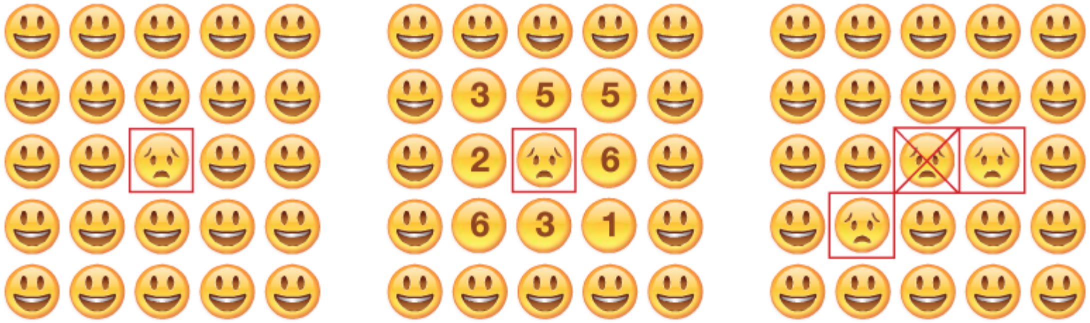
We can simulate the spread of disease in the class using a normal die.

Suppose the student in the middle is the one initially having the illness. This student is the 'active' spreader of the illness to immediate neighbours. (See diagram below left.)

This student has 8 immediate neighbours.

Starting with the neighbour 'below' the initial carrier we roll a normal die and, if a six results, we assume the student catches the illness. We then work clockwise around the 8 neighbours.

The eight rolls: 3, 6, 2, 3, 5, 5, 6, 1 shown in the middle situation leads to the situation below right – three people infected and two 'active' spreaders of the illness whose immediate neighbours now need to be considered.



If this process continues will the entire class become infected? If not how many will? (Assume an infected person does not become active a second time.)

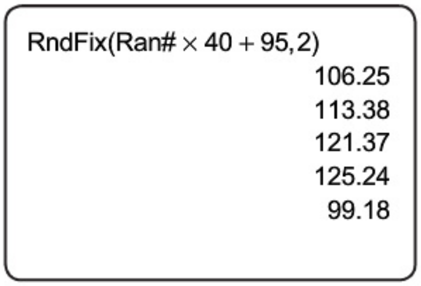
Carry out the simulation a number of times and report on your findings.

You might also like to consider the following:

- Suppose the student initially having the illness is not the middle student?
- Suppose the infection rate is something other that one-sixth? Perhaps one-tenth or one-half or …?

Random number generating from other distributions

Suppose we wanted to randomly select five fit adult males aged between 20 and 40, and note their systolic blood pressure. If we were to know that this blood pressure reading for almost all fit adult males aged between 20 and 40 lies in the range 95 mm Hg to 135 mm Hg, we could ask our calculator or computer spreadsheet to generate five random numbers in the range 95 to 135:



```
RndFix(Ran# × 40 + 95,2)
                        106.25
                        113.38
                        121.37
                        125.24
                         99.18
```

However, such a list would give numbers taken from a uniform distribution. The histogram on the next page shows how the the 150 random numbers listed below tend to demonstrate this uniformity.

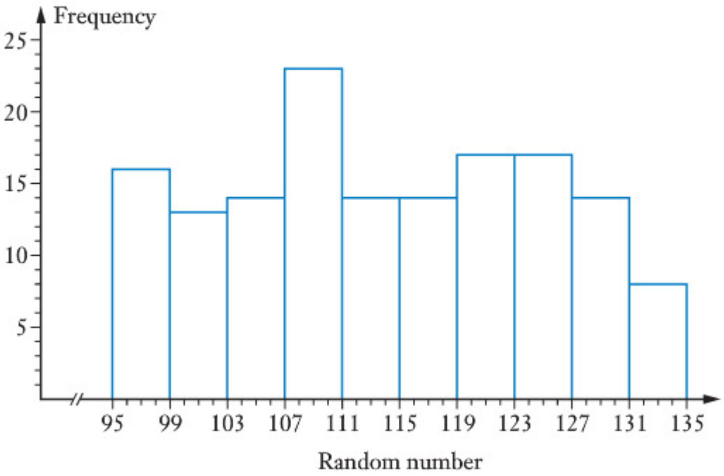
104.51	114.53	109.87	134.79	118.51	132.97	96.35	112.51	108.65	95.95
119.07	113.01	119.7	98.18	122.45	117.73	99.46	132.85	124.45	132.3
102.3	110.43	96.17	132.97	121.19	104.96	98.25	123.73	123.59	101.28
121.27	100.52	124.9	125.47	107.13	97.63	113.06	105.94	130.74	124.39
117.61	122.33	106.05	97.36	115.34	106.3	110.06	122.14	115.86	112.51

97.21	98.65	116.61	109.98	131.75	120.99	128.93	104.34	126.62	104.07
114.07	110.91	129.82	110.57	128.44	110.7	100.98	111.84	110.06	120.2
128.16	114.84	119.55	125.87	95.83	118.18	120.95	117.23	107.93	108.68
130.23	108.66	95.48	108.5	107.91	116.17	105.76	130.74	118.66	105
107.82	95.7	108.34	100.54	133.05	103.72	117.71	118.39	129.04	130.18

120.1	127.89	107.45	128.73	115.64	111.52	111.04	119.93	108.02	128.29
134.13	124.05	98.29	112.2	108.06	113.63	98.21	102.22	115.88	120.42
125.8	110.25	122.02	104.43	110.4	125.95	102.04	103.74	126.44	101.66
112.05	123.47	100.81	125.16	114.17	120.74	100.54	102.31	123.15	95.69
129.28	101.69	103.37	126.59	110.62	106.6	125.1	96.05	127.94	122.17

Number (x)	$95 \le x < 99$	$99 \le x < 103$	$103 \le x < 107$	$107 \le x < 111$	$111 \le x < 115$
Frequency	16	13	14	23	14

Number (x)	$115 \le x < 119$	$119 \le x < 123$	$123 \le x < 127$	$127 \le x < 131$	$131 \le x < 135$
Frequency	14	17	17	14	8

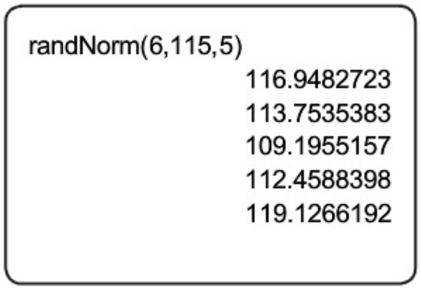


However, with blood pressure tending to be normally distributed it would be better if our sample still involved each member of the appropriate population having an equal chance of being selected but, with more people having a blood pressure close to the mean, we need our selection to be more likely to give a random number close to the mean. We can achieve this if our random selection is made from a normal distribution.

Let us suppose that for the population under consideration, systolic blood pressure is normally distributed with mean 115 mm Hg and standard deviation 6 mm Hg.

Some calculators, computer spreadsheet programs and interactive websites can generate random numbers from a normal distribution.

Explore the capability of your calculator or computer spreadsheet in this regard.



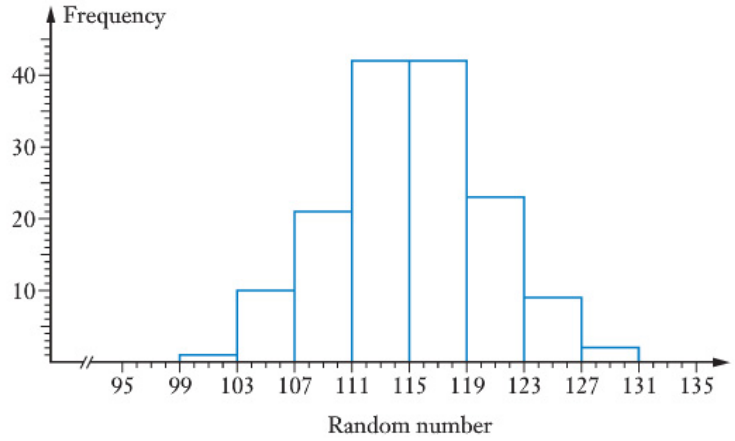
```
randNorm(6,115,5)
                        116.9482723
                        113.7535383
                        109.1955157
                        112.4588398
                        119.1266192
```

The 150 numbers below have been randomly generated from a normal distribution with a mean of 115 and a standard deviation 6. Notice on the next page how the characteristic bell shaped curve is evident on the histogram.

111.9	115.25	104.81	118.62	120.84	126.37	118.56	111.76	123.68	113.57
109.26	119.48	114.52	118.12	112.29	111.5	112.06	113.73	117.54	115.84
122.34	115.45	118.22	107.86	121.95	110.74	128.94	123.25	115.22	122.9
124.96	120.05	111.7	110.67	116.02	116.02	118.31	120.49	111.63	111.98
106.76	115.09	118.94	108.17	112.49	112.02	113.72	120.02	114.7	117.67
119.92	113.83	122.02	117.85	110.13	115.94	115.64	111.35	106.46	121.06
115.93	118.26	109.34	117.61	106.37	118.61	126.94	107.73	118.42	120.01
106.48	112.76	117.17	107.6	108.96	114.01	119.38	114.76	109.11	118.36
110.85	108.49	119.11	113.69	112.55	120.18	109.01	104.66	111.04	119.8
112.59	117.84	117.26	112.44	111.3	106.78	113.24	120.05	111.23	119.5
116.98	123.68	118.97	116.14	110.89	106.31	113.37	110.4	114.03	112.33
111.87	102.84	124.09	121.17	113.7	115.92	111.83	114.86	112.38	116.82
106.49	122.2	116.04	105.36	110.77	117.33	123.75	113.73	114.37	112.96
110.34	117.25	115.14	111.71	117.57	113.67	119.45	108.63	109.53	118.28
123.21	122.06	115.08	117.1	117.4	120.54	130	113.73	108.78	118.39

Number (x)	$95 \le x < 99$	$99 \le x < 103$	$103 \le x < 107$	$107 \le x < 111$	$111 \le x < 115$
Frequency	0	1	10	21	42

Number (x)	$115 \le x <119$	$119 \le x < 123$	$123 \le x < 127$	$127 \le x < 131$	$131 \le x < 135$
Frequency	42	23	9	2	0



More simulations

Earlier we considered two simulations:

Simulation I: Overbooking.
Simulation II: Spread of illness.

We will now consider two more but now our generation of random numbers can come from non-uniform distributions.

Simulation III: Investing funds

A person wishes to invest \$30 000 into share funds, property trusts and cash.

This person believes that in any one year:

the share funds are likely to do anything from losing about 12% to gaining 18%,
i.e. the multiplication factor will be somewhere between 0.88 and 1.18.

the property trusts are likely to do anything from losing 8% to gaining 16%,
i.e. the multiplication factor will be somewhere between 0.92 and 1.16.

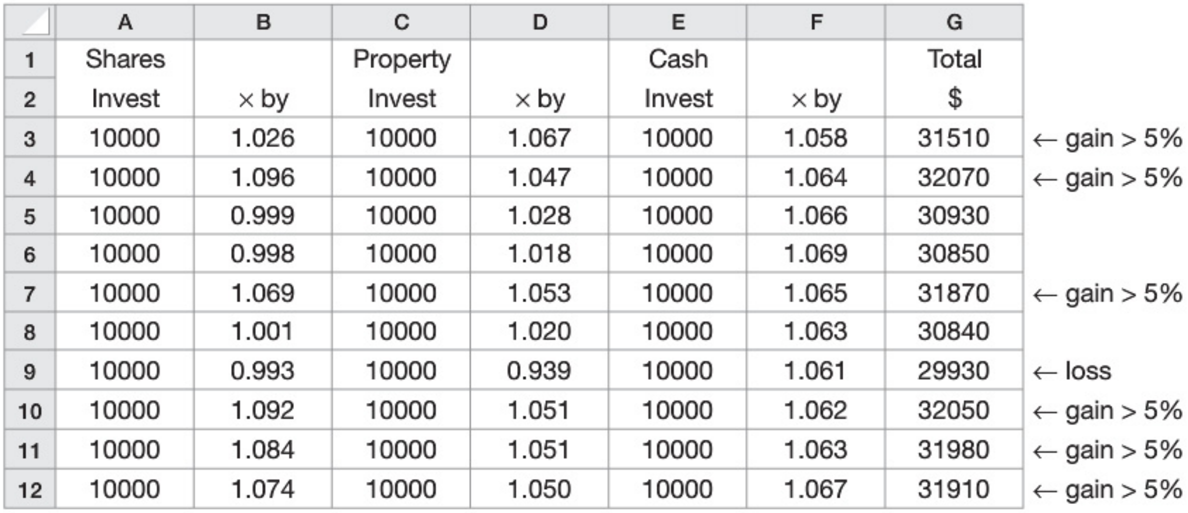
and the cash is likely to do anything from gaining 5% to gaining 8% in a bank account.
i.e. the multiplication factor will be somewhere between 1.05 and 1.08.

Let us suppose that the multiplication factors are each normally distributed as follows:

Share funds multiplication factor ~ $N(1.03, 0.05^2)$.
Property trusts multiplication factor ~ $N(1.04, 0.04^2)$.
Cash multiplication factor ~ $N(1.065, 0.005^2)$.

The person decides to simulate a number of years with $10 000 invested in each of the three types of investment, and the multiplication factors randomly generated from appropriate normal distributions.

Ten such simulations are shown below.



	A	B	C	D	E	F	G	
1	Shares		Property		Cash		Total	
2	Invest	× by	Invest	× by	Invest	× by	$	
3	10000	1.026	10000	1.067	10000	1.058	31510	← gain > 5%
4	10000	1.096	10000	1.047	10000	1.064	32070	← gain > 5%
5	10000	0.999	10000	1.028	10000	1.066	30930	
6	10000	0.998	10000	1.018	10000	1.069	30850	
7	10000	1.069	10000	1.053	10000	1.065	31870	← gain > 5%
8	10000	1.001	10000	1.020	10000	1.063	30840	
9	10000	0.993	10000	0.939	10000	1.061	29930	← loss
10	10000	1.092	10000	1.051	10000	1.062	32050	← gain > 5%
11	10000	1.084	10000	1.051	10000	1.063	31980	← gain > 5%
12	10000	1.074	10000	1.050	10000	1.067	31910	← gain > 5%

Based on these ten runs we might suggest that

- the probability that by the end of a year the total value of the investment will have reduced is 0.1.
- the probability that by the end of a year the total value of the investment will have increased by at least 5% is 0.6.

However such statements would not necessarily be regarded as particularly reliable when based on just ten runs of the simulation.

- Perform this simulation yourself for at least 50 runs and use your results to put forward some statements like to the two given earlier. (If you feel that the range of percentage losses or gains used above are not applicable for the financial climate at the time of reading then research and adjust accordingly.)
- Carry out the simulation a number of times but divide the $30 000 between the three types of investment differently to the even split shown above. Consider for example $20 000 to shares, $10 000 to the property trusts and $0 to cash.

Using random numbers to simulate an event involving a number of variables, in the above case the various multiplication factors, and running the simulated event many times, usually with the aid of a computer, is called a **Monte Carlo simulation**.

RESEARCH

Do some research to find out who named such a process a Monte Carlo simulation and why.

Simulation IV: Will the mineral extraction be profitable?

Let us suppose that an Australian company involved with the extraction and processing of minerals wants to investigate the viability of extracting a particular mineral, which we shall call X, from a newly discovered deposit located overseas. The country involved has granted a mining licence to the Australian company.

An ore containing X will have to be mined and then processed to yield the X it contains.

The company wishes to analyse the likely profitability of such a venture.

Whether this new venture will be profitable or not will depend on many variables. For example:

- **The cost of mining and processing each tonne of the ore.**
  This will vary according to the difficulty of extraction and processing, local costs for labour, time lost due to mechanical failure, availability of sufficient skilled labour, etc. If paid in local currency the exchange rate introduces another variable but for simplicity we will assume payments are in American dollars (US$).
- **The amount of X extracted from each tonne of ore.**
  This will vary according to the concentration of X in the ore which could vary across the expanse of the deposit.
- **The amount the company will receive for each tonne of X.**
  The company will sell the X in the 'market place' where the price is quoted in US dollars (US$). This price will vary according to the pressures of supply and demand.
- **The US dollar to Australian dollar exchange rate.**
  The company needs to be able to justify any decision to go ahead with the venture by predicting likely profits in Australian dollars (A$) and the exchange rate will vary with time.

There may well be all sorts of other variables too but for now let us restrict our attention to the ones just listed.

Suppose the company analyses these four aspects using two different models, one which assumes the variables are uniformly distributed across particular ranges and the other that assumes the variables are normally distributed.

Item	Uniform model	Normal model
Mining and processing 1 tonne of the ore (US$)	From 120 to 180	$\sim N(150, 10^2)$
Number of kg of X extracted per tonne of ore	From 135 to 225	$\sim N(180, 15^2)$
US$ received for each tonne of X	From 1200 to 2400	$\sim N(1800, 200^2)$
Exchange rate, i.e. what 1 US$ buys in A$	From 0.96 to 1.44	$\sim N(1.2, 0.08^2)$

Thus, for the uniform model, the worst and best case scenarios, for 1 tonne of ore, would be as follows:

Worst case scenario	**Best case scenario**
Costs US$180 to obtain	Costs US$120 to obtain
This produces 135 kg of X	This produces 225 kg of X
Sale of this raises US$162	Sale of this raises US$540
Hence LOSS in US$ is $18	Hence PROFIT in US$ is $420
Loss in A$ is $25.92	Profit in A$ is $604.80

Five simulations of each model are shown below with the profit (or loss) involved from each tonne calculated, in Australian dollars.

	Cost (in US$) of mining and processing 1 tonne of ore	**Amount (kg) of *X* produced per tonne of ore**	**US$ received for each tonne of *X***	**A$ bought by 1 US$**	**Profit (in A$) for each tonne of ore**
Uniformly distributed model	150.93	218.93	1435.63	1.1036	180.30
	121.69	166.49	1308.53	1.0801	103.87
	140.76	143.84	2271.59	1.2089	224.84
	124.26	204.89	2199.99	1.1528	376.38
	177.80	182.53	1549.11	1.3960	146.52
Normally distributed model	156.28	141.60	1781.58	1.1496	110.35
	151.34	198.30	1758.71	1.1958	236.07
	149.62	157.23	1803.02	1.2400	166.00
	139.12	194.80	1422.09	1.1841	163.29
	165.80	194.81	1997.46	1.2687	283.33

Using a computer to simulate thousands of runs the likelihood of the profit falling into a particular range could be investigated for each model.

Use a spreadsheet to run the above simulations and write a report of your findings.

Shutterstock.com/Andriy Solovyov

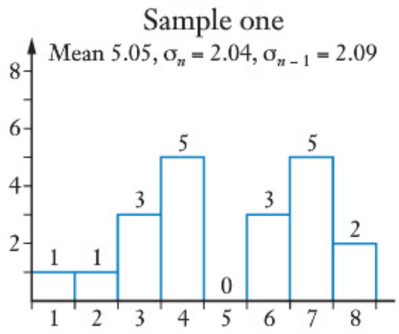
Variability of random samples

Whether we are considering a random sample of numbers taken from a uniform distribution or from a normal distribution, or indeed from any other distribution, we should not expect every such sample of numbers taken from that distribution to have the same characteristics. We would expect a certain amount of variation between samples even when they are taken from the same distributions.

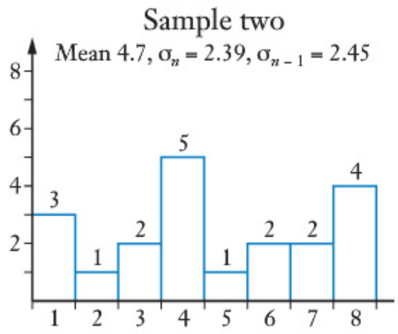
The following graphs each show the distribution of a sample of 20 numbers generated from a random variable, X, with X uniformly distributed across the integers

$$1, 2, 3, 4, 5, 6, 7, 8.$$

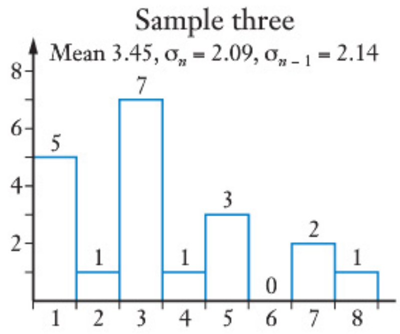
Note: $E(X) = 4.5$, $SD(X) = 2.29$ (2 decimal places).


Sample one

Mean 5.05, $\sigma_n = 2.04$, $\sigma_{n-1} = 2.09$

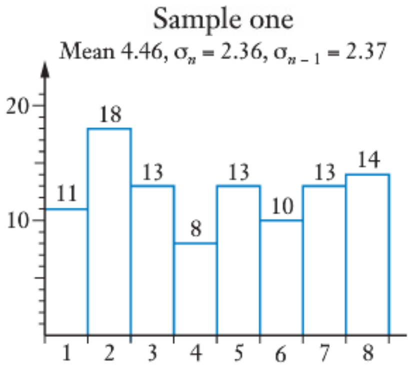

Sample two

Mean 4.7, $\sigma_n = 2.39$, $\sigma_{n-1} = 2.45$

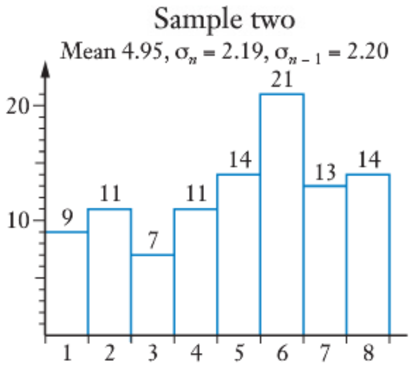

Sample three

Mean 3.45, $\sigma_n = 2.09$, $\sigma_{n-1} = 2.14$

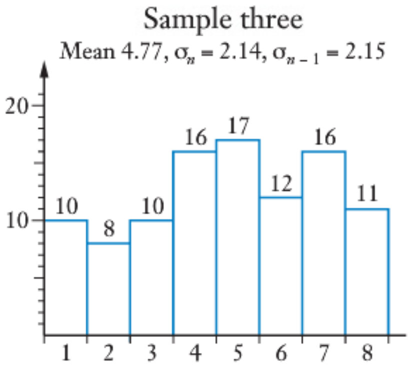
The following graphs each show the distribution of a sample of 100 numbers generated from a random variable, X, with X uniformly distributed across the integers 1 to 8.


Sample one

Mean 4.46, $\sigma_n = 2.36$, $\sigma_{n-1} = 2.37$

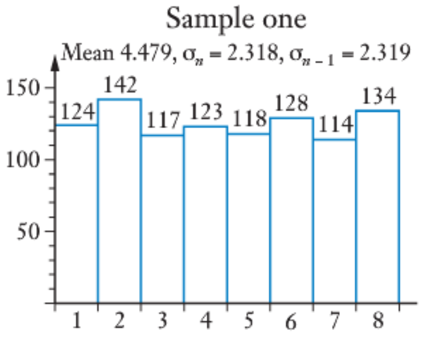

Sample two

Mean 4.95, $\sigma_n = 2.19$, $\sigma_{n-1} = 2.20$

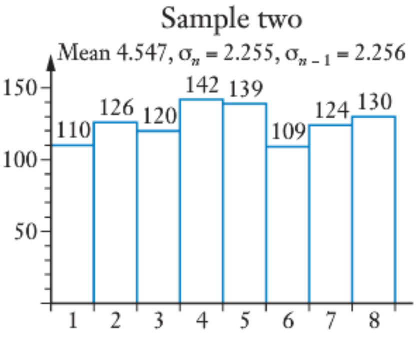

Sample three

Mean 4.77, $\sigma_n = 2.14$, $\sigma_{n-1} = 2.15$

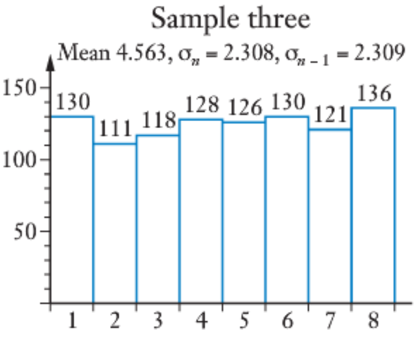
The following graphs each show the distribution of a sample of 1000 numbers generated from a random variable, X, with X uniformly distributed across the integers 1 to 8.


Sample one

Mean 4.479, $\sigma_n = 2.318$, $\sigma_{n-1} = 2.319$


Sample two

Mean 4.547, $\sigma_n = 2.255$, $\sigma_{n-1} = 2.256$

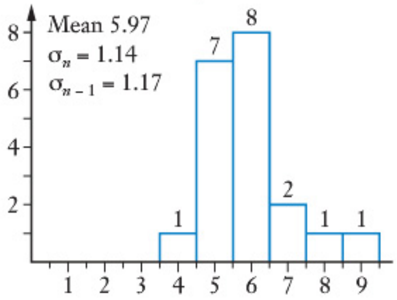

Sample three

Mean 4.563, $\sigma_n = 2.308$, $\sigma_{n-1} = 2.309$

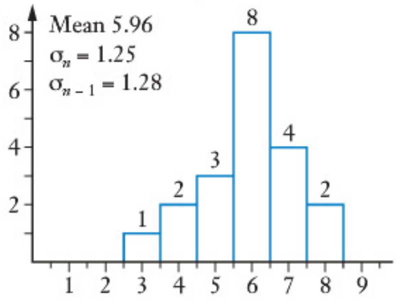
The following graphs each show the distribution of a sample of 20 numbers generated from a random variable, X, with $X \sim N(6, 1.2^2)$

Note: On this page, whilst each graph shows the numbers grouped into columns, each mean and standard deviation have been calculated from the original numbers.

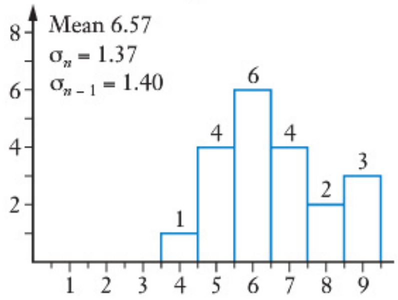
Sample one



Sample two

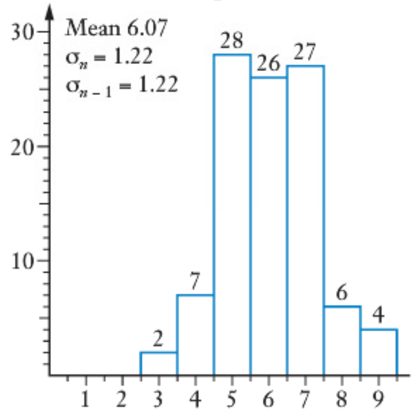


Sample three

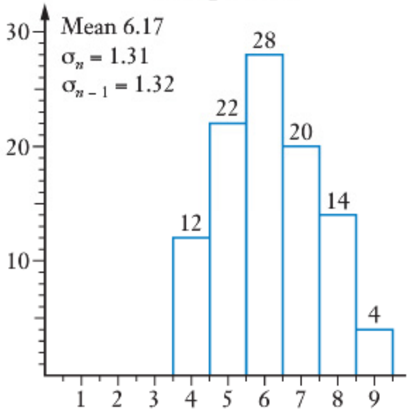


The following graphs each show the distribution of a sample of 100 numbers generated from a random variable, X, with $X \sim N(6, 1.2^2)$.

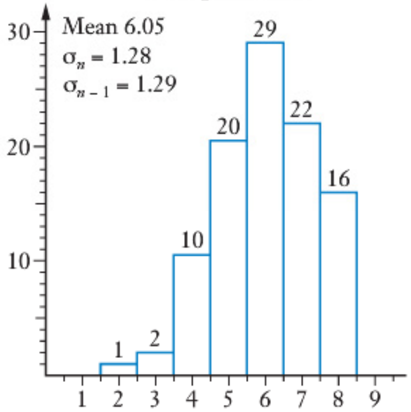
Sample one



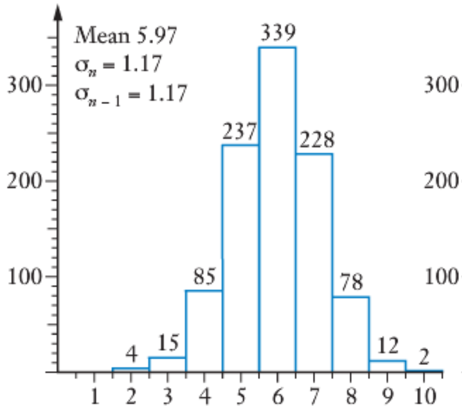
Sample two

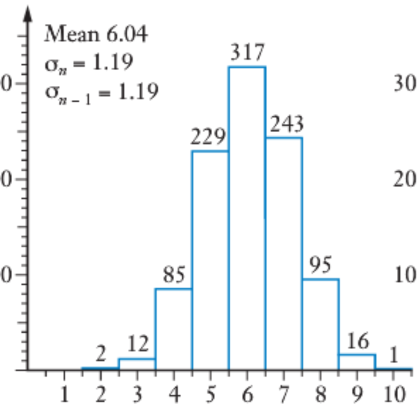


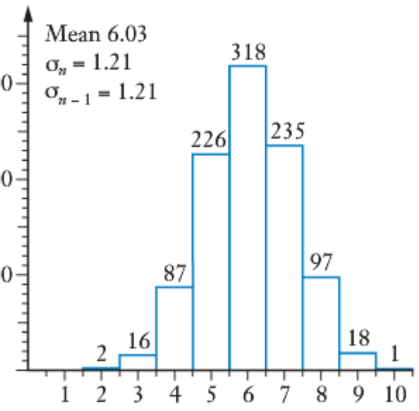
Sample three



The following graphs each show the distribution of a sample of 1000 numbers generated from a random variable, X, with $X \sim N(6, 1.2^2)$.

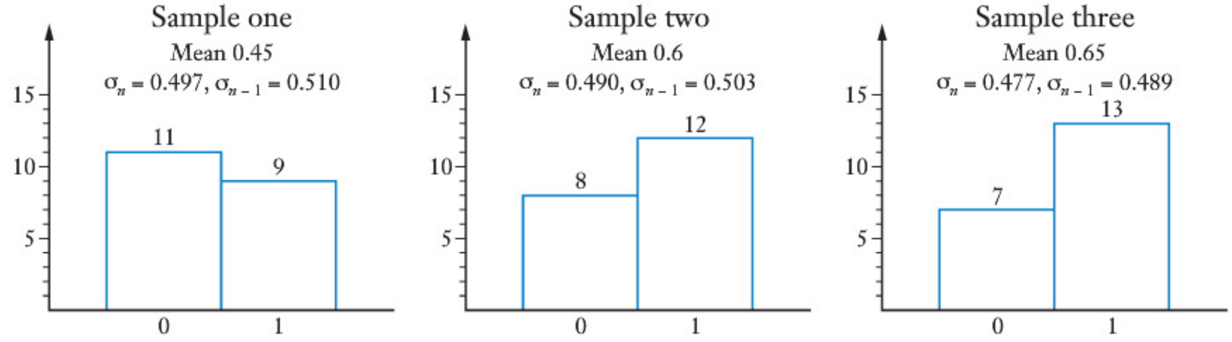






The following graphs each show the distribution of a sample of 20 numbers generated from a Bernoulli distribution, X, with $P(0) = 0.4$ and $P(1) = 0.6$.

$$\text{Note: } E(X) = 0.6,\ SD(X) = \sqrt{0.6(1-0.6)} \approx 0.49.$$



Sample one

Mean 0.45

$\sigma_n = 0.497, \sigma_{n-1} = 0.510$

Sample two

Mean 0.6

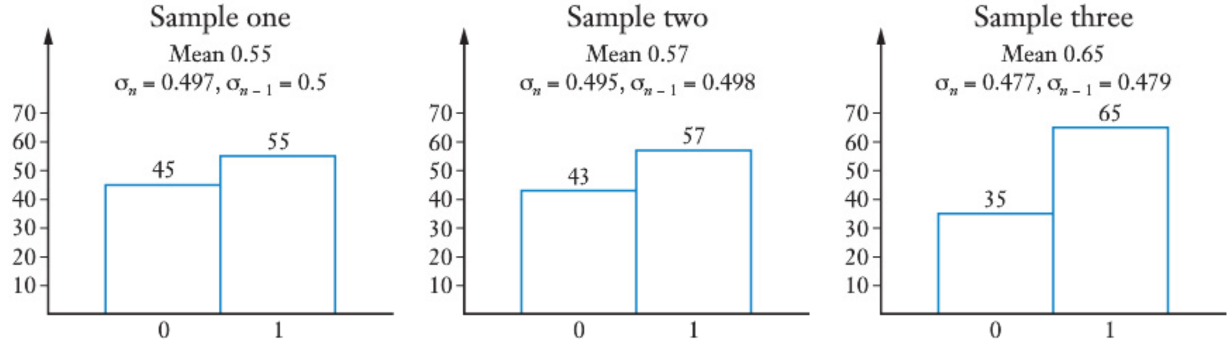
$\sigma_n = 0.490, \sigma_{n-1} = 0.503$

Sample three

Mean 0.65

$\sigma_n = 0.477, \sigma_{n-1} = 0.489$

The following graphs each show the distribution of a sample of 100 numbers generated from a Bernoulli distribution with $P(0) = 0.4$ and $P(1) = 0.6$.



Sample one

Mean 0.55

$\sigma_n = 0.497, \sigma_{n-1} = 0.5$

Sample two

Mean 0.57

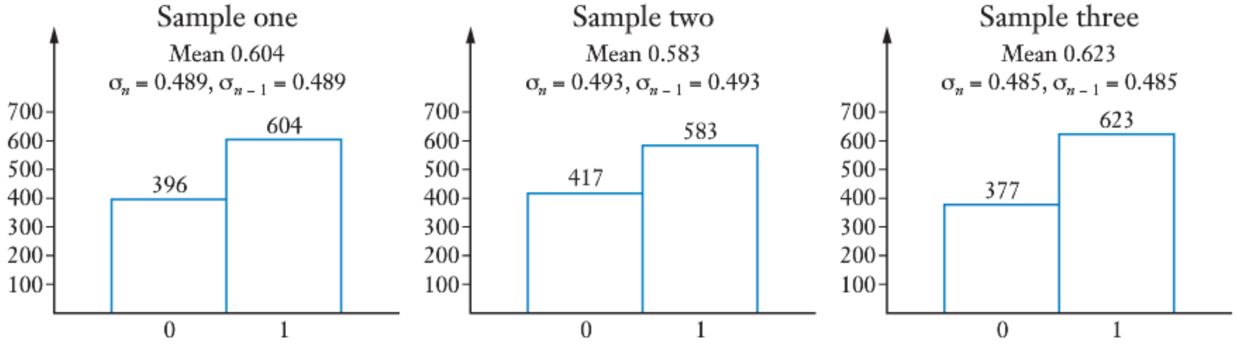
$\sigma_n = 0.495, \sigma_{n-1} = 0.498$

Sample three

Mean 0.65

$\sigma_n = 0.477, \sigma_{n-1} = 0.479$

The following graphs each show the distribution of a sample of 1000 numbers generated from a Bernoulli distribution with $P(0) = 0.4$ and $P(1) = 0.6$.



Sample one

Mean 0.604

$\sigma_n = 0.489, \sigma_{n-1} = 0.489$

Sample two

Mean 0.583

$\sigma_n = 0.493, \sigma_{n-1} = 0.493$

Sample three

Mean 0.623

$\sigma_n = 0.485, \sigma_{n-1} = 0.485$

Use the graphs of this page and the previous two pages to answer the following:

- Are all samples of the same size, and from the same distribution, identical?
- Are the means of samples from the same distribution the same as each other? If not the same are they 'close'? Are the sample means close to the population mean? How close?
- Are the standard deviations of samples from the same distribution the same as each other? If not the same are they 'close'? Are the sample standard deviations close to the population standard deviation?
- As sample size increases does the 'shape' of the graph get closer to the 'shape' of the population distribution?

Miscellaneous exercise five

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

1 Evaluate $\log(100) - \ln(e^{-3})$ without the use of a calculator.

2 If $P = 9e^{(t+1)}$ find an exact expression for t in terms of P and evaluate it, correct to three decimal places if rounding is necessary, for

a $P = 180$,

b $P = 3600$,

c $P = 9e^3$.

3 If $\log a = p$ and $\log b = q$, express each of the following in terms of p or q or both p and q.

a $\log(ab)$ **b** $\log\left(\frac{a}{b}\right)$ **c** $\log(a^2b^3)$

d $\log\sqrt{a}$ **e** $\ln a$ **f** $\log_5(b^2)$

4 The discrete random variable X is binomially distributed with $X \sim \text{Bin}(n, p)$.
If $E(X) = 60$ and $SD(X) = 6$ find n, p and $P(X \le 50)$ giving the last of these correct to three decimal places.

Differentiate each of the following with respect to x.

5 $y = x\ln(5x)$ **6** $y = (\log_e x)^2$ **7** $y = x^2 \ln x$

8 $y = (3 + \ln x)^2$ **9** $y = \dfrac{2}{x} + 2\ln x$ **10** $y = \dfrac{1}{\ln x}$

11 If $f(x) = x^3 \ln x$ determine $f''(x)$, the second derivative of the function with respect to x.

12 A particle moves in a straight line such that its displacement from a fixed point O, at time t seconds ($t \ge 0$), is x metres where $x = 9\ln(1+t) - 4t$.

Find t when

a the velocity is zero,

b the velocity (in m/s) is numerically equal to the acceleration (in m/s^2).

13 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} k(4-x) & \text{for } 1 \le x \le 3 \\ 0 & \text{elsewhere.} \end{cases}$$

Determine

a the value of k,

b $E(X)$, the expected value, or long-term mean, of X,

c $\text{Var}(X)$, the variance of X,

d $SD(X)$, the standard deviation of X.

e Define $P(X \le x)$, the cumulative distribution function for X, for $-\infty < x < \infty$.

14 As part of a population estimation exercise, 127 birds of a particular species visiting a swamp region favoured by migrating birds of this species are caught, tagged and released back into the region.

A few days later a second capture of a sample of this species of bird from the same swamp region caught 89 birds and it was found that amongst these 89 birds there were 3 carrying tags from the first group of 127 birds.

Releasing these 89 birds back into the region and then, a few days later, carrying out a third capture, saw 99 being caught of which 4 carried tags from the first group of 127.

Use these figures to estimate the number of birds of this species living in the area.

What might be a problem with using capture-recapture techniques on migratory birds visiting a particular region?

15 Which of the following statements are true for all $p > 0$ and $q > 0$?

a $\log_p q = \log_q p$

b $\log(p + q) = \log p + \log q$

c $\log(p - q) = \log p - \log q$

d $\log(pq) = \log p \times \log q$

e $(\log p)^q = q \log p$

f $\dfrac{\log p}{\log q} = \log p - \log q$

g $\log\left(\dfrac{p}{q}\right) = \dfrac{\log p}{\log q}$

h $\dfrac{1}{\log p} = \log\left(\dfrac{1}{p}\right)$

i $\log(pq) = \log p + \log q$

j $\log\left(\dfrac{p}{q}\right) = \log p - \log q$

k $\log(p^q) = q \log p$

l $\log_p q \times \log_q p = 1$

16 The random variable, X, is normally distributed with a mean of 1240 and a variance of 56^2, i.e. $X \sim \mathrm{N}(1240, 56^2)$. Determine $\mathrm{P}(1200 < X < 1300)$.

17 If $X \sim \mathrm{N}(50, 10^2)$ determine, to three significant figures:

a the 0.34 quantile,

b the 0.82 quantile,

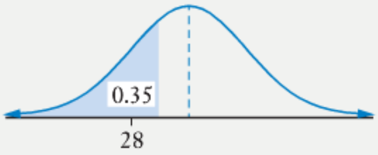
c the 43rd percentile,

d the lower quartile.

18 The continuous random variable, X, is normally distributed with $\mathrm{P}(X < 28) = 0.35$.

a How many standard deviations from the mean is a score of 28?

b If the standard deviation of X is 5.74, find the mean of the distribution, giving your answer correct to two decimal places.



6.

Sample proportions

- Variation between samples
- Sample proportion distribution
- How would we determine the distribution of $\hat{p}$ if we don't know p?
- Why is it useful to know how the sample proportions are distributed?
- Confidence intervals
- Margin of error
- Sample size
- Increasing the level of confidence increases the margin of error
- Let's check
- Miscellaneous exercise six

Did you know that (at the time of writing):

Approximately 78% of the people living in Western Australia live in Perth.

Approximately 64% of the people who live in New South Wales live in Sydney.

Approximately 38% of Labor party MPs (federal) are women.

Approximately 10% of people in the world are left handed.

In Australia approximately 32% of human births are by caesarean section.

China won almost 13% of the gold medals awarded at the London Olympics.

Approximately 17% of the world's population live in India.

Approximately 71% of the Earth's surface is covered by water.

More than 80% of Australians live within 100 kilometres of the sea.

In discussion with others suggest proportions that you think appropriate for each of the following:

What proportion of the world's population are Chinese?

What proportion of Australian adults have never held a driver's licence?

What proportion of homes in Australia are double storey?

What proportion of Australians live in Victoria?

What proportion of Australians are female?

What proportion of Australian families own a dog?

What proportion of Australian married couples never have children?

What proportion of Australian families own more than one car?

What proportion of the world's adults are vegetarian?

What proportion of the world's population celebrate Christmas?

iStock.com/izarek

ISBN 9780170395205

As you may have realised from the items on the previous page, and indeed the title of the chapter, we are now considering proportions of a population.

Suppose we wanted to know what proportion of the Australian population engaged in a particular activity, for example following a vegetarian diet or being a regular attender at a gym, or what proportion possessed a particular characteristic, for example being aged over 60 or being left handed etc. We could include a suitable question in the census so that all of the population is checked for this activity or characteristic, or we could consider a sample that is representative of the whole population and use the proportion from the sample to suggest the proportion for the population. For example, if we found that in a sample of 200 Australians, 19 were left handed, i.e. 9.5% of our sample were left handed, we could suggest that 9.5% of Australians were left handed, or, as a proportion of the whole, 0.095. However, if we were to choose a different sample we could well obtain a different proportion. How different?

How different might the proportions be from one sample to another?

How might the proportions from a number of such samples be distributed?

How different might the proportion in a sample be from the proportion that exists in the population?

If we had the proportions from a number of samples we could combine the results to suggest the proportion of Australians who were left handed. (If the other samples also each involved 200 Australians this would simply be the mean of the proportions.)

The proportion of left handers in samples we choose is a random variable, capable of taking any value from 0 to 1, but probably close to the proportion that exists in the whole population, especially if the sample is large. The proportion in the population is the **population proportion**, sometimes simply referred to as p. The proportion from our random sample varies according to the sample we choose and is called the **sample proportion**, sometimes written as $\hat{p}$ (pronounced 'p hat').

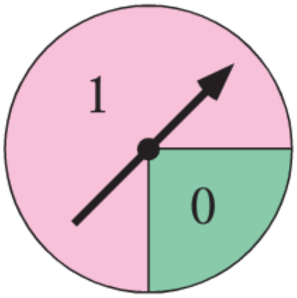
How close would our sample proportion, $\hat{p}$, be to the population proportion, p?

Is the size of our sample of any significance?

We have posed a lot of questions about the proportion of left handed people in our sample. Let us now consider this idea of 'sample proportions' further, and let us start by considering something for which we know the value of p, the population proportion, or long term proportion.

Variation between samples

Consider the spinner on the right.

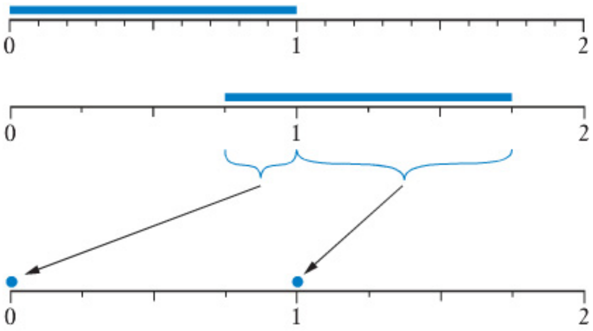


From the geometry of the spinner we know that in the long term, the proportions of spins that will result in a 1 will be 0.75.

Now suppose we simulate 20 spins of such a spinner, a number of times, and see how the proportion of 1s in our samples of 20 spins vary.

Question: How could we simulate the spinning of such a spinner?
Answer: We could use random numbers, as shown on the next page.

To change from a random number with an output in the range zero to one, to an output of either zero or one, with P(1) = 0.75 we can proceed as follows:



Usual random number output is between 0 and 1:

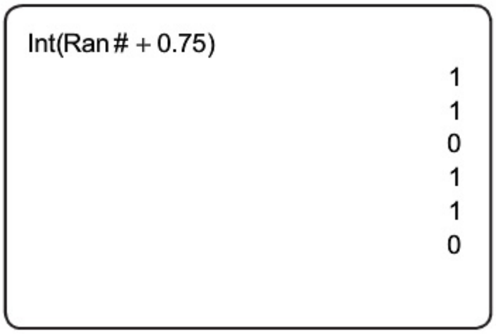
Add 0.75 to obtain numbers between 0.75 and 1.75.

Display only the integer part of such numbers.

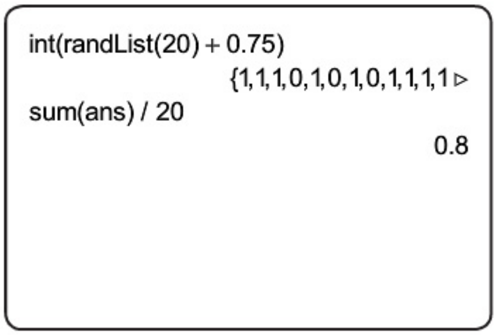
This would make 1 three times as likely as 0.

The display below left shows six ones or zeros generated in this way.

The display below right shows the generation of 20 such numbers (not all of which are displayed), summing them (to determine the number of 1s) and expressing the number of 1s as a proportion of the 20 spins. In this case, 80% of the 20 spins were 1s. For this sample $\hat{p} = 0.8$.



```
Int(Ran# + 0.75)
                                1
                                1
                                0
                                1
                                1
                                0
```



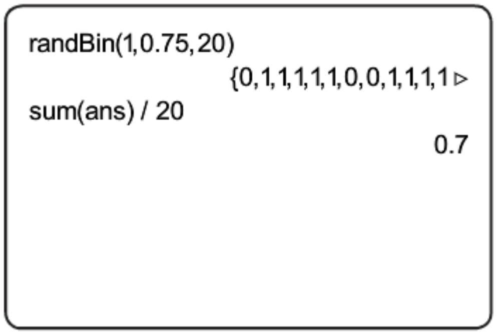
```
int(randList(20) + 0.75)
              {1,1,1,0,1,0,1,0,1,1,1,1 ▷
sum(ans) / 20
                                    0.8
```

Alternatively we could generate zeros and ones according to a Binomial distribution, Bin (1, 0.75). With one trial involved we have either 0 successes or 1 success and with P(success) = 0.75 we will have

$$P(0) = 0.25 \text{ and } P(1) = 0.75$$

as required.

For the sample on the right $\hat{p} = 0.7$.



```
randBin(1,0.75,20)
              {0,1,1,1,1,1,0,0,1,1,1,1 ▷
sum(ans) / 20
                                    0.7
```

Three further samples of twenty spins, together with the proportion of 1s obtained, are given below:

1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1

Proportion of 1s: 0.85, i.e., for this sample, $\hat{p} = 0.85$.

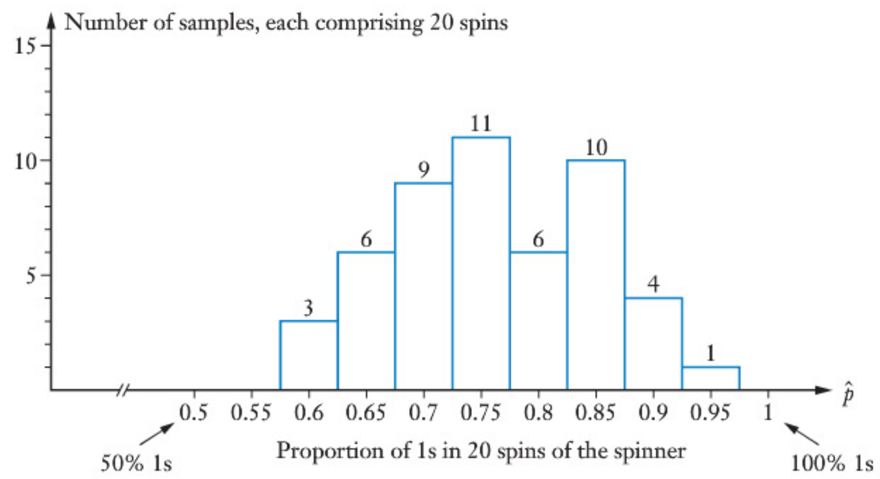
1 0 1 1 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1

Proportion of 1s: 0.8, i.e., for this sample, $\hat{p} = 0.8$.

0 0 1 0 1 0 0 1 0 1 1 1 1 0 1 0 1 1 1 1

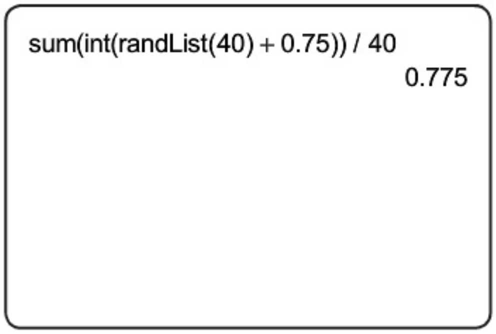
Proportion of 1s: 0.6, i.e., for this sample, $\hat{p} = 0.6$.

The graph below shows the results of 50 such samples, each involving 20 simulated spins of the spinner, with the distribution of the proportion of 1s as shown.



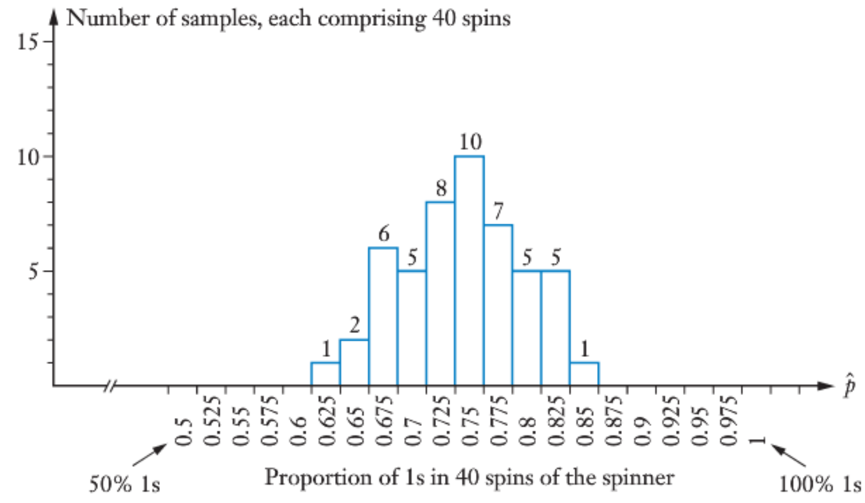
Suppose instead we again collected 50 samples but we now increase the number of spins in each sample to, say 40, rather than 20.

Check that you understand the display on the right, which indicates that the proportion of 1s in the single simulated sample of 40 spins of our spinner was 0.775.



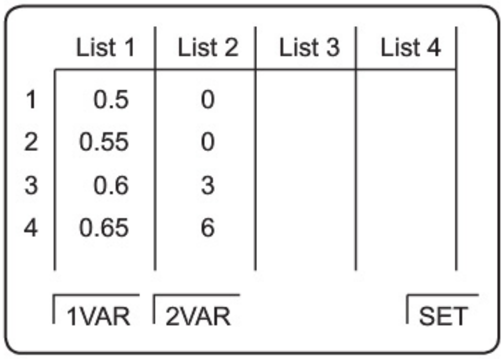
```
sum(int(randList(40) + 0.75)) / 40
                              0.775
```

The graph below shows the distribution of the proportion of 1s for a simulation involving 50 samples, each sample comprising 40 spins. (The display above, with its sample proportion of 0.775, is just one of these 50 samples, there being 7 with sample proportion of 0.775.)

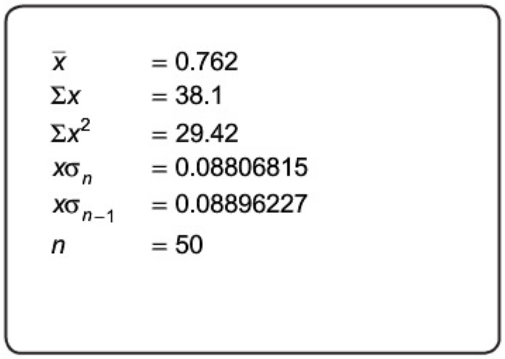


Notice that both graphs of $\hat{p}$ for multiple samples appear somewhat 'bell shaped' and both peak around the theoretical long term proportion of 0.75.

Using the information given on each graph, and with the assistance of a calculator we can determine that the first graph has a mean of 0.762 and a standard deviation of 0.09,

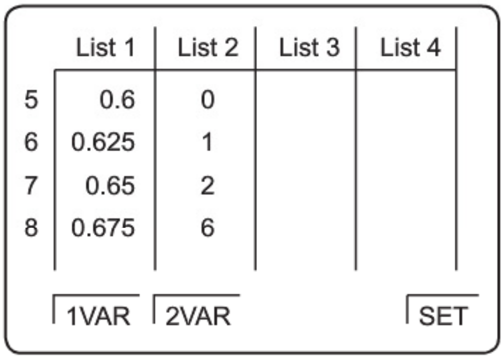


	List 1	List 2	List 3	List 4
1	0.5	0		
2	0.55	0		
3	0.6	3		
4	0.65	6		

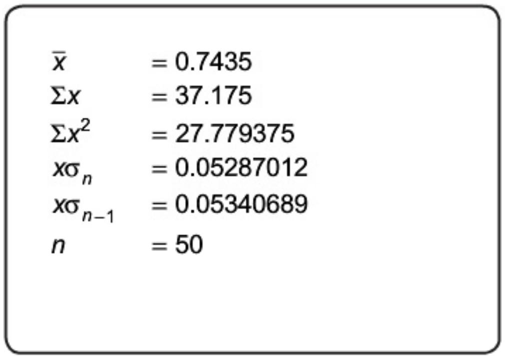


$\bar{x}$ = 0.762
Σx = 38.1
Σx^2 = 29.42
$x\sigma_n$ = 0.08806815
$x\sigma_{n-1}$ = 0.08896227
n = 50

whilst the second graph has a mean of 0.7435 and a standard deviation of 0.053.



	List 1	List 2	List 3	List 4
5	0.6	0		
6	0.625	1		
7	0.65	2		
8	0.675	6		



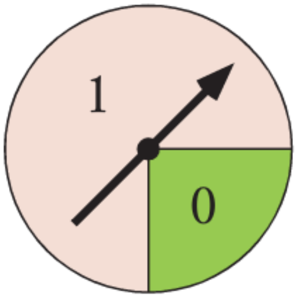
$\bar{x}$ = 0.7435
Σx = 37.175
Σx^2 = 27.779375
$x\sigma_n$ = 0.05287012
$x\sigma_{n-1}$ = 0.05340689
n = 50

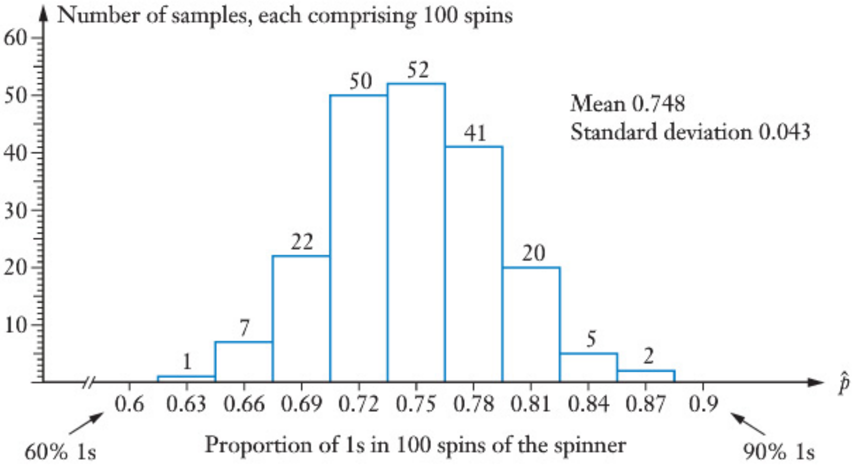
Thus it would appear that as we increase the sample size, the mean of the sample proportions gets closer to the theoretical mean and the sample proportions are less spread out.

If we increase both the number of spins in each sample, i.e. we increase the sample size, and we increase the number of samples, the shape of the histogram becomes more like the characteristic bell shape of the normal distribution.

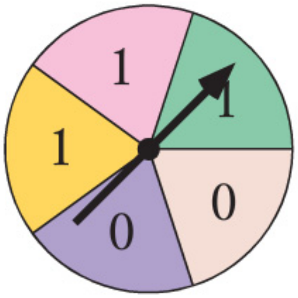
The graph on the next page shows the distribution of the proportion of 1s in two hundred samples with each sample involving one hundred spins of the spinner.

(The graph involves grouped data on the horizontal axis with the mean and the standard deviation calculated from the grouped data.)





SIMULATION – YOUR TURN



If we made repeated spins of the spinner shown on the right we would expect, in the long run, that the proportion of spins giving a 1 would be 0.6, i.e. we would expect that approximately 60% of the spins would result in a 1.

Step 1: Using the random number generating facility of some calculators or computer spreadsheets, simulate twenty spins of this spinner and note the proportion of 1s.

Step 2: Repeat step 1 nine more times, to give ten samples in all, with each sample involving twenty spins of this spinner, and for each sample note the proportion of 1s.

Step 3: Find the mean and standard deviation of the ten proportion values you have from carrying out the previous steps.

Is your mean close to the expected long term value of 0.6?

Step 4: Simulate fifty spins of this spinner and note the proportion of 1s.

Is your proportion close to the expected long term value of 0.6?

Step 5: Repeat step 4 twenty-nine more times, to give thirty samples in all, with each sample involving fifty spins of this spinner, and for each sample note the proportion of 1s.

Step 6: Find the mean and standard deviation of the thirty proportion values you have from carrying out steps 4 and 5.

Is your mean close to the expected long term value of 0.6?

Is the standard deviation for the proportions of 1s with a sample size of 50 less than the standard deviation of the proportions with a sample size of 20?

Compare your response to this question with those of others in your class.

ISBN 9780170395205

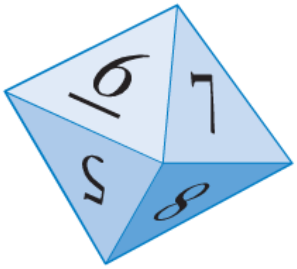
If we were to roll a normal fair octagonal die, ten times, what proportion of the ten rolls would you expect to give an odd number?

Well, with an equal number of odd {1, 3, 5, 7} and even {2, 4, 6, 8} numbers we would expect approximately half of the ten rolls to give an odd number. But what variation should we expect in this proportion of odd numbers? The results of twenty sample rolls of such a die, with each sample involving ten rolls are shown below, together with the proportion of odd numbers obtained in each sample.

Sample 1:	3	4	1	2	5	5	3	6	4	1	Proportion of odd numbers	0.6
Sample 2:	8	4	2	7	8	4	5	7	6	1	Proportion of odd numbers	0.4
Sample 3:	2	7	5	1	8	7	8	3	8	5	Proportion of odd numbers	0.6
Sample 4:	4	2	7	6	1	1	6	3	1	6	Proportion of odd numbers	0.5
Sample 5:	4	8	1	1	2	2	6	4	5	4	Proportion of odd numbers	0.3
Sample 6:	2	1	8	1	7	4	6	4	5	2	Proportion of odd numbers	0.4
Sample 7:	2	2	7	8	6	8	3	4	8	4	Proportion of odd numbers	0.2
Sample 8:	8	5	5	6	8	1	5	1	1	3	Proportion of odd numbers	0.7
Sample 9:	2	1	1	3	2	6	3	5	8	1	Proportion of odd numbers	0.6
Sample 10:	4	2	7	1	5	1	3	4	8	5	Proportion of odd numbers	0.6
Sample 11:	7	1	2	1	4	7	3	7	1	1	Proportion of odd numbers	0.8
Sample 12	3	3	7	8	2	2	7	7	8	5	Proportion of odd numbers	0.6
Sample 13:	5	3	1	8	6	7	7	1	6	1	Proportion of odd numbers	0.7
Sample 14:	7	2	4	3	6	2	5	1	8	3	Proportion of odd numbers	0.5
Sample 15:	5	7	7	2	2	3	1	8	8	1	Proportion of odd numbers	0.6
Sample 16:	1	3	2	1	1	8	1	6	7	6	Proportion of odd numbers	0.6
Sample 17:	7	7	4	4	7	8	5	4	5	1	Proportion of odd numbers	0.6
Sample 18:	7	2	4	2	2	5	2	6	6	7	Proportion of odd numbers	0.3
Sample 19:	8	7	8	7	5	2	2	6	2	3	Proportion of odd numbers	0.4
Sample 20:	6	5	7	1	3	1	4	3	1	2	Proportion of odd numbers	0.7

Our 20 sample proportions have a mean of 0.535
and a standard deviation, σ_n, of 0.153.

SIMULATION – YOUR TURN



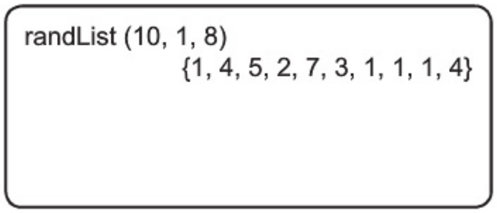
What might the mean and standard deviation of the sample proportions be if we still have 20 samples but now the sample size is 20, or 50, or …?

Investigate.

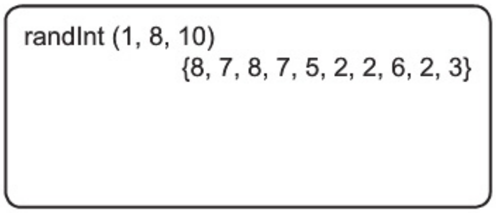
However, before you do, read the next page for some suggestions that may help you create the simulated die rolling.

Simulation method 1

Some calculators can produce a list of random integers between two values, for example 1 to 8 inclusive. Hence, to simulate the rolling of the octagonal die ten times we could instruct such a calculator to list ten integers in this way. We would then note what proportion of the numbers are odd numbers.



```
randList (10, 1, 8)
                {1, 4, 5, 2, 7, 3, 1, 1, 1, 4}
```

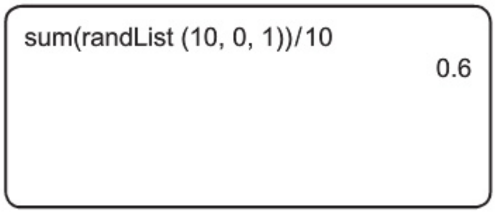


```
randInt (1, 8, 10)
                {8, 7, 8, 7, 5, 2, 2, 6, 2, 3}
```

(Alternatively we could use the ability of computer spreadsheets and calculators to generate random numbers between 0 and 1 but alter the output appropriately, as explained earlier in this book.)

Simulation method 2

With each roll of the die equally likely to produce an odd as it is an even, we could simply output 10 numbers that are either 0 (for even) or 1 (for odd), sum the list (to give the total number of odd numbers) and then divide by 10 to give the proportion of odd numbers.



```
sum(randList (10, 0, 1))/10
                                    0.6
```

Simulation method 3

Suppose you wish to keep the authentic feel to the simulation by actually simulating the ten rolls and then checking for each number being even or odd, which method 2 lacks. In that case first note that if we were to divide a randomly generated integer by 2 the remainder is 0 when the integer is even, and 1 when the integer is odd. Some calculators and computer spreadsheets have the facility to return a zero if a number is even and a 1 if the number is odd. Applying this facility to a list of numbers, and then summing the 0s and 1s, gives a count of the odd numbers.

Such a function is sometimes called *remain* (as in *remain*der) *mod* (as in *mod*ulo arithmetic) or perhaps *iMod*.

So if we

- generate an appropriate list of randomly generated integers,
- apply the remain, mod, or iMod function, to return a 1 for an odd number and a 0 for an even number,
- sum the 0s and 1s produced,
- divide by the sample size,

we can quickly generate appropriate proportions.

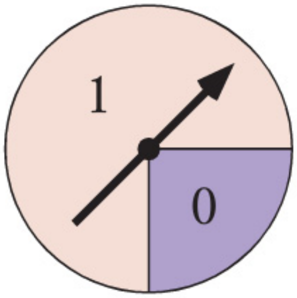
See the displays at the top of the next page for the sort of calculator language needed for this task. The display top left shows the step by step process for one sample of ten numbers and top right shows two 'all at once' formulae each generating the proportion of odd numbers in a sample of ten random integers between 1 and 8 inclusive.

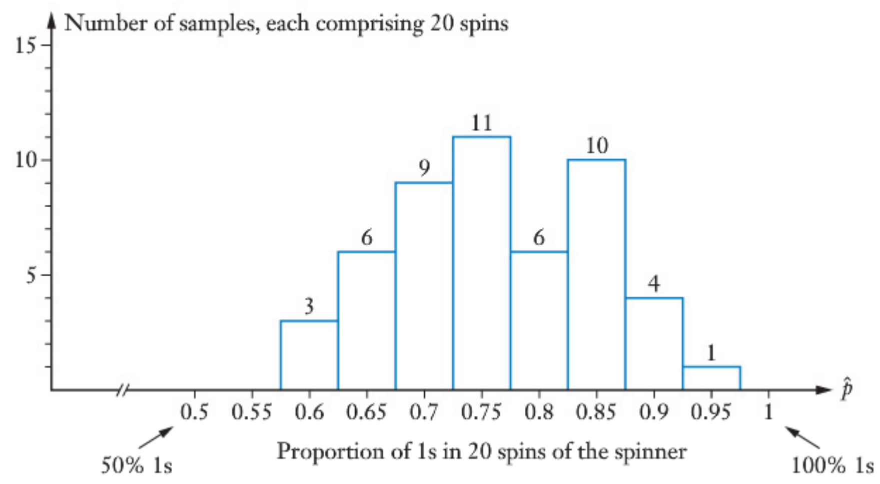
```
randList (10, 1, 8)
                          {6, 4, 4, 2, 3, 5, 3, 7, 4, 1}
iMod (ans, 2)
                          {0, 0, 0, 0, 1, 1, 1, 1, 0, 1}
sum (ans)/10
                                                     0.5
```

```
sum(mod(randInt (1, 8, 10), 2))
              10
                                       0.6
sum(remain(randInt (1, 8, 10), 2))
              10
                                       0.5
```

Sample proportion distribution

Earlier in this chapter we considered 50 samples each involving 20 spins of the spinner shown below. For each sample we determined the proportion of the 20 spins resulting in a 1.





Now we know from the geometry of the spinner that in the long term the proportion of 1s will be 0.75. Our sample values estimated this long term proportion and gave us a distribution with a mean of 0.762 and a standard deviation of 0.09.

Counting the number of 'successes' in n trials involves a binomial distribution Bin(n, p), where p is the probability of success. This has mean np, standard deviation $\sqrt{np(1-p)}$.

The *proportion* of successes simply divides the number of successes by n. Hence the distribution of sample proportions will have mean p, standard deviation $\sqrt{p(1-p)/n}$.

In fact, according to a statistical rule called the Central Limit Theorem:

As the sample size, n, increases, the distribution of the sample proportions will approach that of a normal distribution, mean p, standard deviation $\sqrt{p(1-p)/n}$.

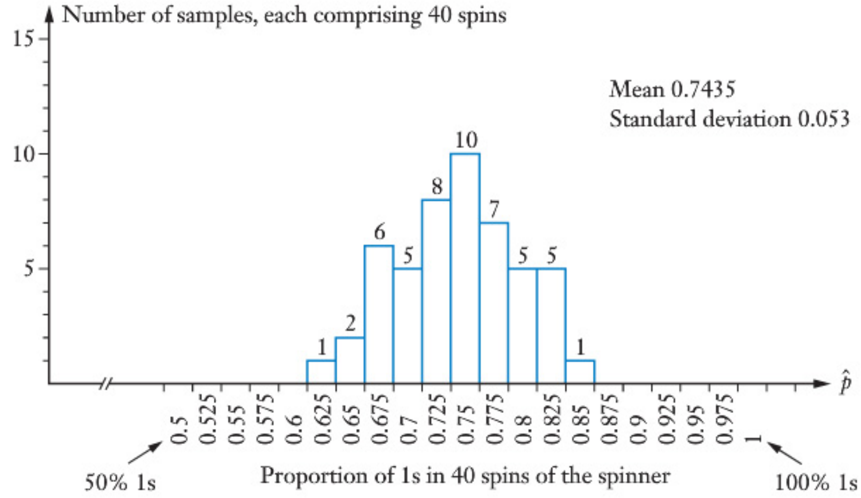
Thus in our situation with $p = 0.75$, and $n = 20$, the theory suggests that the proportion of 1s would approach a normal distribution with mean 0.75 and standard deviation 0.097. Our values of 0.762 and 0.09 are reasonably close to the theoretical model. However, the graph does not look especially normal. This is because the approximation to the normal distribution improves as the sample size increases. To be confident that the sample distributions will approximate to the stated normal distribution we need both $np \geq 10$ and $n(1-p) \geq 10$. (Indeed some statistical textbooks even suggest ≥ 15.)

For this same spinner we also considered sample sizes of 40 and 100:

With $p = 0.75$ and $n = 40$ the rule suggests that our distribution of sample proportions should be approximately normal with mean 0.75 and standard deviation

$$\sqrt{\frac{0.75(1-0.75)}{40}} = 0.068$$

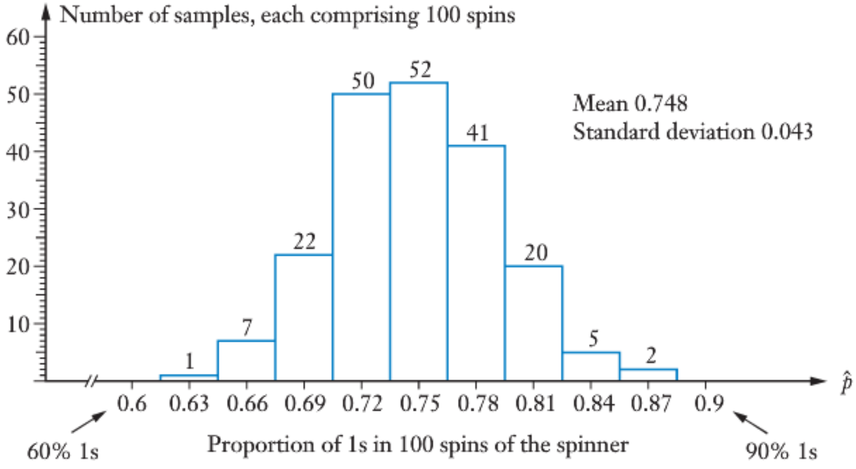
We obtained:



With $p = 0.75$ and $n = 100$ the rule suggests that our distribution of sample proportions should be approximately normal with mean 0.75 and standard deviation

$$\sqrt{\frac{0.75(1-0.75)}{100}} = 0.0433.$$

We obtained:

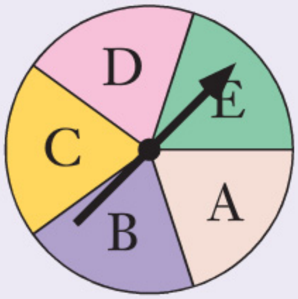


For increasing sample size, and with both np and $n(1 - p) \geq 10$, the graphs above do appear to be getting more normally distributed and the means and standard deviations are roughly as the rule would have us expect.

ISBN 9780170395205

How do your results for the simulation of the proportion of odd numbers obtained when rolling an octagonal die compare with the normal distribution that we would expect?
Does the distribution of sample proportions approximate towards the expected normal distribution as the sample size gets larger, with np and $n(1-p) \geq 10$?

EXAMPLE 1



When the spinner on the right was spun 200 times an A occurred on 43 occasions.

a What is the value of p, the population proportion of As?

b What is the value of $\hat{p}$, the sample proportion of As?

c Calculate the mean and standard deviation of $\hat{p}$ for such samples of 200 spins.

Solution

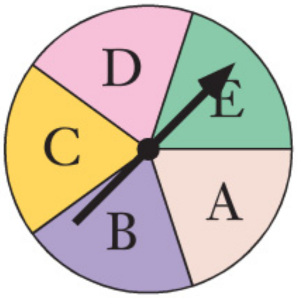
a In the long term we would expect an A to occur 20% of the time.
Thus $p = 0.2$

b In our sample of 200 spins an A occurred on 43 occasions.
Thus $\hat{p} = 0.215$.

c $\hat{p}$ will have a mean equal to p and a standard deviation of $\sqrt{\frac{p(1-p)}{n}}$, with $p = 0.2$, and $n = 200$.

Thus $\hat{p}$ has a mean of 0.2 and a standard deviation of 0.0283, correct to 4 dp.

How would we determine the distribution of $\hat{p}$ if we don't know p?



It is all very well considering the proportion of 1s that might occur when a normal die is rolled, or the proportion of As that will occur when the spinner on the right is spun a number of times, because for these sort of random events the long term proportion, or population proportion, p, is known. However it will often be the case that the population proportion, p, will not be known and we will be sampling from the population in order to estimate p. If we need to give a single value estimate of the population proportion we would use the best estimate we have available, i.e our sample proportion. (Though we will see later in this chapter that we can instead give an *interval* in which the population proportion is likely to lie.) Similarly, to estimate the standard deviation of the distribution of sample proportions we can again use our best estimate for p, i.e. $\hat{p}$, in the formula for standard deviation. For large n this use of $\hat{p}$ for p, is still likely to give a good approximation for the standard deviation (and after all, if we only had one sample, and not knowing p, it would be the best we can do).

EXAMPLE 2

A survey was carried out to investigate the number of fifty to sixty year old males who had suffered back problems. The survey found that in a sample of 221 fifty to sixty year old males, 124 had suffered back problems.

a Calculate the sample proportion, $\hat{p}$, of those surveyed who had suffered back problems.

b Estimate the standard deviation of the random variable $\hat{p}$ for such samples of size 221.

Solution

a In the sample of 221 fifty to sixty year old males, 124 had suffered back problems.

Thus $\hat{p} = \dfrac{124}{221} \approx 0.561.$

b The random variable $\hat{p}$ will have standard deviation $\approx \sqrt{\dfrac{\frac{124}{221}\left(1-\frac{124}{221}\right)}{221}} \approx 0.0334.$

Why is it useful to know how the sample proportions are distributed?

Knowing that sample proportions are normally distributed with mean p, the population proportion, and standard deviation $\sqrt{p(1-p)/n}$, where n is the sample size, allows us to apply our understanding of the normal distribution to sample proportions, as the following examples show.

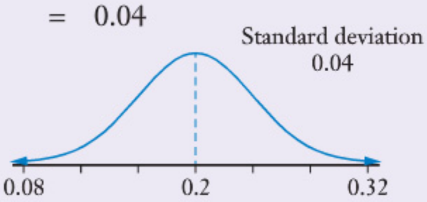
EXAMPLE 3

If we classify as 'extremely unlikely' the likelihood that something that is normally distributed is more than three standard deviations from the mean explain why it is extremely unlikely that, when sampling 100 light bulbs from a batch that is thought to have 20% of the bulbs defective, we would find more than 35 of our sample defective.

Solution

The sample proportions would be normally distributed with mean $= 0.2$

$$\text{and standard deviation} = \sqrt{\frac{0.2(1-0.2)}{100}} = 0.04$$



For 35 defective globes in 100 we have a sample proportion of 0.35.

Now $\dfrac{0.35-0.2}{0.04} = 3.75.$

Thus a sample proportion of 0.35 is 3.75 standard deviations above the mean. To find more than 35 defective bulbs in a sample of 100 is therefore extremely unlikely.

If we did get more than 35 defective light bulbs in our sample of 100 we would question the statement that 20% were defective, or we would question whether our sample was truly representative of the batch.

ISBN 9780170395205

Shutterstock.com/Ambient Idea

Remember:

- The actual proportion (p) in a population, for example the proportion of people who are left handed, is fixed for a given population.
- The proportion who are left handed in a sample, $\hat{p}$, will vary according to the sample. Hence $\hat{p}$ forms a random variable.
- These sample proportions are normally distributed with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$ where n is the sample size.

EXAMPLE 4

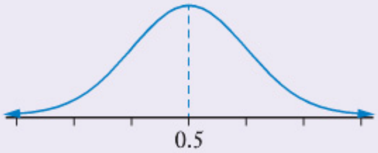
For this question use the rule that for a normal distribution we would expect approximately 95% of the 'scores' to be within 2 standard deviations of the mean.

If we were to roll a normal six sided die 80 times we would expect to get an even number approximately 50% of the time. Between what two values, situated symmetrically either side of the 50% long term average, would we expect the proportion of even numbers to lie for approximately 95% of the time?

Solution

The sample proportions would be normally distributed with mean $= 0.5$

$$\text{and standard deviation} = \sqrt{\frac{0.5(1-0.5)}{80}} \approx 0.0559$$



Standard deviation 0.0559

If $\hat{p}$ is within two standard deviations of the mean then

$$0.5 - 2 \times 0.0559 < \hat{p} < 0.5 + 2 \times 0.0559$$

i.e. $$0.3882 < \hat{p} < 0.6118$$

On approximately 95% of the occasions that we roll a normal six sided die 80 times we would expect the proportion of even numbers to be between 39% and 61%.

The answer for the previous example stated that:

> *On approximately 95% of the occasions that we roll a normal six sided die*
> *80 times we would expect the proportion of even numbers to be between 39% and 61%.*

The 95% assumes that rolling the die 80 times is carried out numerous times. However even if we were to carry out the 80 rolls just 100 times we would still expect to find that the number of times the proportion of even numbers fell between 0.39 and 0.61 was pretty close to 95.

Simulating 80 rolls of a normal die, noting the proportion of times an even number was achieved and then repeating this simulation a further 99 times gave the following 100 proportions.

0.5125	0.5	0.6125	0.55	0.425	0.4125	0.5625	0.4	0.45	0.5125
0.45	0.55	0.5625	0.4625	0.5	0.5875	0.5375	0.525	0.525	0.575
0.4875	0.4	0.625	0.5375	0.55	0.4375	0.5375	0.5625	0.4625	0.4625
0.375	0.5	0.5	0.5375	0.55	0.525	0.5125	0.475	0.525	0.4
0.4375	0.5	0.475	0.4375	0.4	0.55	0.5	0.5375	0.5875	0.6
0.45	0.5	0.6	0.4875	0.5125	0.4625	0.5625	0.45	0.525	0.45
0.5375	0.4	0.4375	0.6625	0.5875	0.5125	0.525	0.4375	0.5625	0.4875
0.4875	0.425	0.5375	0.525	0.475	0.425	0.5375	0.5875	0.525	0.525
0.5375	0.4375	0.5375	0.475	0.525	0.5375	0.5125	0.4875	0.5	0.4625
0.5125	0.525	0.5125	0.5375	0.4875	0.5	0.525	0.475	0.5	0.475

Repeating this process again gave the following 100 proportions.

0.525	0.6	0.4125	0.5375	0.55	0.475	0.5875	0.575	0.575	0.625
0.5	0.475	0.525	0.5625	0.4875	0.5125	0.525	0.475	0.5625	0.5625
0.575	0.4875	0.5125	0.475	0.5375	0.6125	0.55	0.525	0.5	0.45
0.45	0.35	0.625	0.4375	0.475	0.6125	0.3375	0.525	0.55	0.45
0.4875	0.4875	0.4375	0.575	0.475	0.5375	0.525	0.5375	0.475	0.4625
0.5	0.5	0.4625	0.5375	0.4875	0.5875	0.475	0.45	0.5	0.4375
0.4625	0.45	0.525	0.45	0.5	0.475	0.4625	0.45	0.5375	0.6
0.5125	0.5	0.4625	0.5	0.5625	0.55	0.5125	0.5875	0.525	0.55
0.575	0.5625	0.4875	0.5	0.45	0.55	0.5875	0.4625	0.525	0.525
0.4375	0.4625	0.5125	0.475	0.5125	0.3625	0.6	0.5125	0.525	0.45

Are the above results consistent with the statement given earlier? i.e.:

> *On approximately 95% of the occasions that we roll a normal six sided die*
> *80 times we would expect the proportion of even numbers to be between 39% and 61%.*

Shutterstock.com/de-n-is

ISBN 9780170395205

EXAMPLE 5

Let us suppose that 54% of a community are in favour of a particular development occurring.

A sample of 500 people from the community are to be surveyed to see if they are in favour of the development occurring.

With A and B symmetrically placed either side of the 54% population proportion copy and complete the following statement:

> *There is an 80% chance that in a sample of 500 people from this community, the sample proportion in favour of the development occurring will be between _A%_ and _B%_.*

Solution

The sample proportions would be normally distributed with mean $= 0.54$

and standard deviation $= \sqrt{\dfrac{0.54(1-0.54)}{500}}$

≈ 0.0223

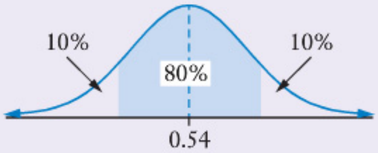
For $X \sim \mathrm{N}(0.54, 0.0223^2)$

If $\mathrm{P}(X < k) = 0.1$

$k = 0.5114$

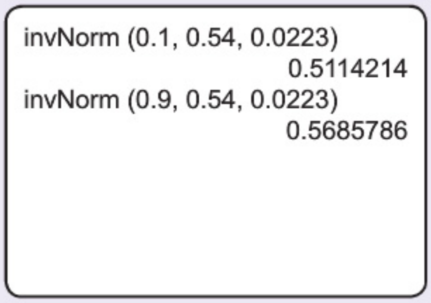
If $\mathrm{P}(X < k) = 0.9$

$k = 0.5686$



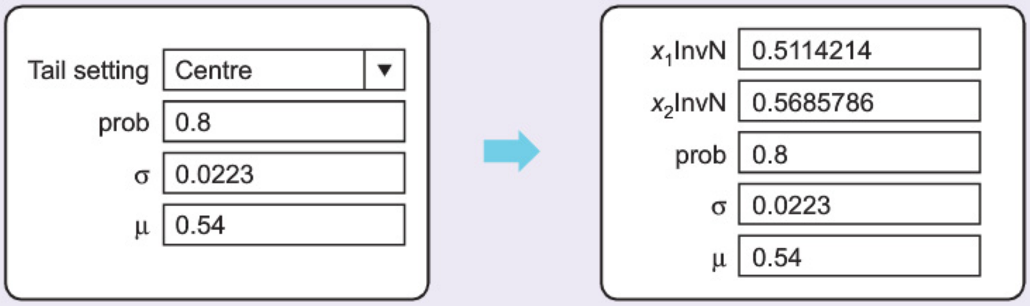
Standard deviation 0.0223

There is an 80% chance that in a sample of 500 people from this community, the sample proportion in favour of the development occurring will be between 51.1% and 56.9%.



```
invNorm (0.1, 0.54, 0.0223)
                        0.5114214
invNorm (0.9, 0.54, 0.0223)
                        0.5685786
```

Some calculators give both values 'in one go' as suggested below (and as you may have discovered when working through chapter four):



Tail setting: Centre ▼
prob: 0.8
σ: 0.0223
μ: 0.54

x_1InvN: 0.5114214
x_2InvN: 0.5685786
prob: 0.8
σ: 0.0223
μ: 0.54

Exercise 6A

1 A survey of 123 randomly selected Australians found that 49 of them were aged over 50.

A second survey involving 2348 randomly selected Australians found that 761 of them were aged over 50.

Determine the sample proportion of people aged over 50 for each survey.

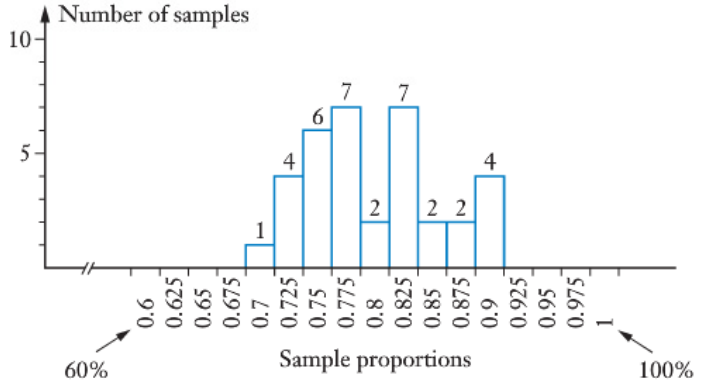
Which sample proportion is likely to be the better estimate of the proportion of Australians aged over 50?

2 During the course of one day ten samples of 'shoppers' were surveyed. Each sample involved 25 shoppers who had purchased at least one item from a particular supermarket, and for each sample the proportion of the 25 shoppers purchasing milk from the supermarket was noted. The proportions for the ten samples were as follows:

Sample	1	2	3	4	5	6	7	8	9	10
Proportion purchasing milk	0.72	0.68	0.88	0.56	0.60	0.76	0.64	0.84	0.72	0.80

For shoppers buying at least one item from this supermarket, estimate the population proportion of the shoppers buying milk from this supermarket and explain your answer.

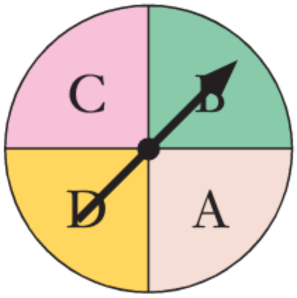
3 A number of sample proportions, for samples of the same size, gave rise to the graph shown below.



a How many samples were involved?

b Estimate p, the population proportion.

4 When the spinner on the right was spun 320 times, an A occurred on 72 occasions.



a What is the value of $\hat{p}$, the sample proportion of As for the 320 spin sample mentioned?

b What is the value of p, the population proportion of As?

c Calculate the mean and standard deviation of the random variable, $\hat{p}$, for samples involving 320 spins.

5 Two normal fair six sided dice are rolled and the two numbers this process gives are added together.

This is repeated a further 99 times and the 100 trials produce 'a total less than 9' on 67 of the 100 occasions.

a What is the value of p, the population proportion of obtaining a total less than 9 when two normal fair dice are rolled?

b What is the value of $\hat{p}$, the sample proportion of a total less than 9 for the 100 trials mentioned?

c Calculate the mean and standard deviation of $\hat{p}$, the sample proportions for samples of 100 rolls of the two dice.

d How many standard deviations from p is our value for $\hat{p}$?

6 The census for a particular country showed that 84% of the households in that country had internet access.

At about the same time as the census data was collected, a sample of 240 households in a particular region of the country showed that 147 of the households had internet access.

a Determine p, the population proportion of households with internet access.

b Determine $\hat{p}$, the sample proportion of households with internet access.

c Comment on your results from **a** and **b**.

7 The abstract for a paper by Curtis Hardyck and Lewis Petrinovich, on left handedness and published by the American Psychological Association in 1977 states that *left-handedness, ranging from moderate through strongly left-handed, is found in approximately 10% of the population*.

Comment on each of the following:

a A survey of 1000 school students classified 112 of the students as being left handed.

b According to an article by L. J. Harris in the Psychology Press publication 'Laterality':

In a survey of participants in the 1981 World Fencing Championship, 35% of the athletes in the foil competition were left-handed.

8 A survey was carried out to investigate the gender of teachers in Australian schools. The survey found that in a sample of 1247 teachers, 461 were male.

a Calculate the sample proportion, $\hat{p}$, of those teachers surveyed who were male.

b Estimate the standard deviation of the random variable $\hat{p}$, for such samples of size 1247.

iStock.com/monkeybusinessimages

9 A survey involving 248 midwives found that 143 of them were aged between 46 and 60.

a Calculate the sample proportion, $\hat{p}$, of those midwives surveyed who were aged between 46 and 60.

b Estimate the standard deviation of the random variable $\hat{p}$ for such samples of size 248.

ISBN 9780170395205

10 Twenty samples, not all involving the same number of people, involved males of 18 years and over and considered the proportion of males in the sample who had high blood pressure, according to a particular classification of high blood pressure. The proportion having high blood pressure in each sample, and the number of people in the sample are given below.

Sample	Number in sample	Sample proportion having high blood pressure
1	8	0.5
2	10	0.1
3	50	0.28
4	25	0.24
5	10	0.2
6	80	0.2375
7	56	0.286
8	10	0.1
9	50	0.2
10	180	0.261
11	20	0.35
12	10	0.1
13	8	0.375
14	25	0.2
15	8	0.5
16	150	0.2
17	20	0.3
18	25	0.32
19	10	0.8
20	1	1

a Why would it be unwise to estimate the population proportion of males aged 18 years and over who have high blood pressure, according to the classification, simply by finding the mean of the above sample proportions?

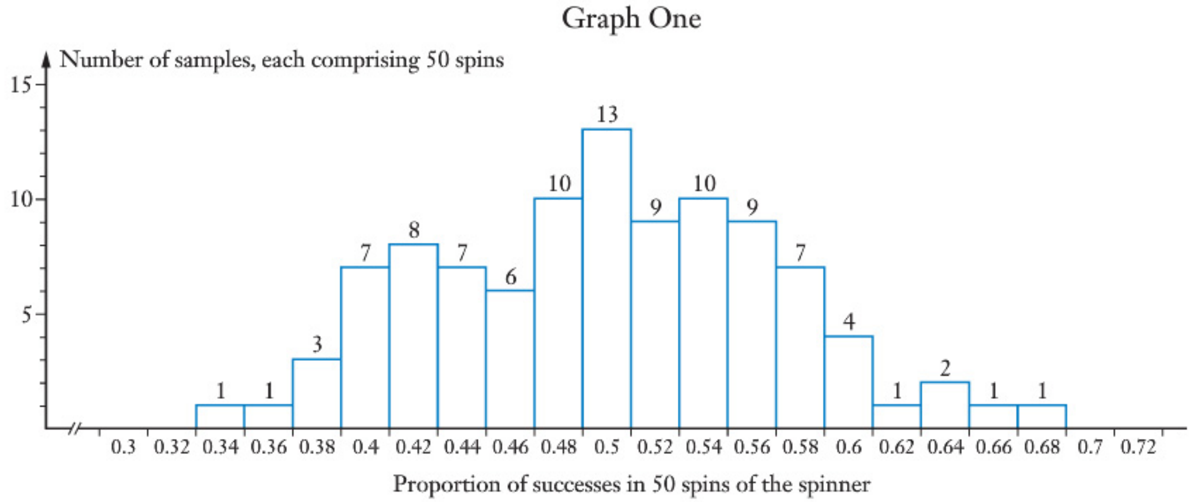
b Explain a better way to use the above figures to determine an estimate for the population proportion of males aged 18 years and over who have high blood pressure, according to the classification, and determine that estimate.

11 If we classify as 'extremely unlikely' the likelihood that something that is normally distributed is more than three standard deviations from the mean, explain why it is extremely unlikely that, when sampling 200 seeds from a batch that is thought to have 10% of the seeds unable to germinate, we would find 35 or more of our sample unable to germinate.

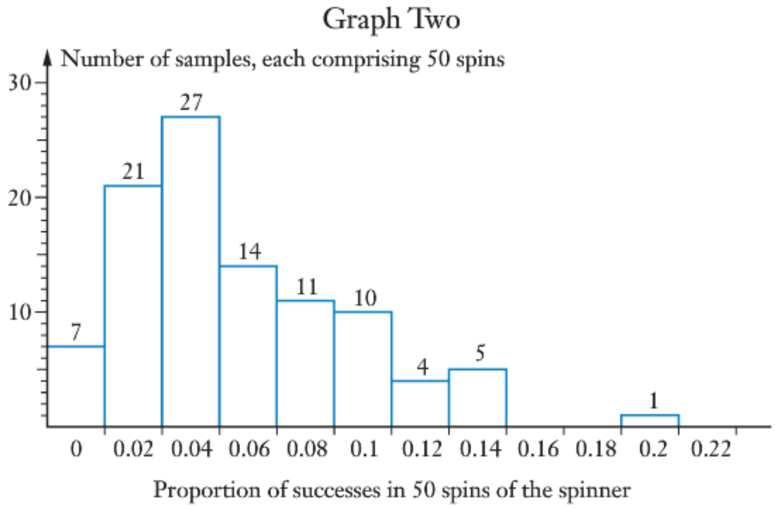
12 A random process for which P(success) = 0.5 was carried out 50 times and the proportion of successes recorded.

This same sampling process was carried out a further 99 times to give 100 sample proportions, each with sample size of 50.

The 100 sample proportions are shown as Graph One below.



Graph Two below also shows 100 sample proportions of successes for samples each of size 50 but this time for a random process for which P(success) = 0.05.



Graph One seems to have much more of the characteristic Normal Distribution ‘bell shape’ about it than Graph Two.

Why should this not really be a surprise?

13 For this question use the rule that for a normal distribution we would expect approximately 68% of the 'scores' to be within 1 standard deviation of the mean.

If we were to roll a normal six sided die 100 times we would expect to get an even number approximately 50% of the time. Between what two values, situated symmetrically either side of the 50% long term average, would we expect the proportion of even numbers to lie for approximately 68% of samples of size 100?

14 A politician claimed that he expected to win 52% of the votes in the forthcoming election for the seat of Dasha.

Wishing to check this, a newspaper carried out a survey and found that out of 200 people who intended to vote in the election for the seat of Dasha, 81 said they would be voting for the politician.

Comment.

15 Let us suppose that 24% of cars produced by the XYZ car manufacturing company are blue.

A random sample of 800 cars produced by this company are surveyed and the proportion of blue cars in the sample noted.

With A and B symmetrically placed either side of the 24% population proportion copy and complete the following statement:

There is a 90% chance that in a sample of 800 cars produced by this company the sample proportion that are blue will be between _*A%*_ *and* _*B%*_.

iStock.com/rawpixel

Confidence intervals

In statistics, samples are taken in an attempt to estimate information regarding the population of which the sample forms a part. Thus if we took a sample from a population and found that 24% of our sample possessed some characteristic, be it left handedness, being over a particular age, supporting a particular political party, or whatever, we could estimate the proportion of the population possessing this characteristic as being 24%. This is a **point estimate** of the population proportion because it gives a 'one value estimate', 0.24. However, we know that the sample proportions over many samples are normally distributed with mean p and standard deviation $\sqrt{p(1-p)/n}$, where p is the population proportion and n is the sample size.

This allows us, with a particular level of confidence, to give a range of values, or *interval*, that we can expect the population proportion to lie within. This is called an **interval estimate**.

First let us revisit the **standard normal distribution**, i.e. $Z \sim N(0, 1^2)$, with its associated 'z scores' and establish some important numbers (sometimes called **critical scores**) relating to what we will call the

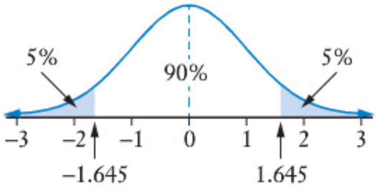
90%, 95% and 99% **confidence intervals**.

ISBN 9780170395205

For a *90% confidence interval*

Solving	$P(Z < k)$	$=$	0.95
gives	k	$=$	1.645

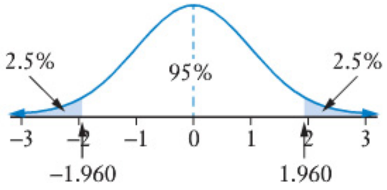
90% of the scores from a normal distribution lie within 1.645 standard deviations of the mean.



For a *95% confidence interval*

Solving	$P(Z < k)$	$=$	0.975
gives	k	$=$	1.960

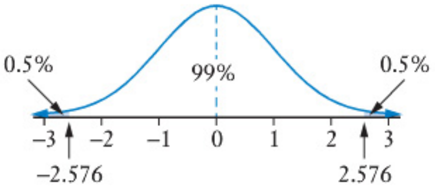
95% of the scores from a normal distribution lie within 1.960 standard deviations of the mean.



For a *99% confidence interval*

Solving	$P(Z < k)$	$=$	0.995
gives	k	$=$	2.576

99% of the scores from a normal distribution lie within 2.576 standard deviations of the mean.



When we take one sample of a sufficiently large size, we know that the proportion of our single sample comes from a distribution of sample proportions that approximate to a normal distribution with a mean equal to the population proportion. Hence we can be 90% confident that our sample proportion is within 1.645 standard deviations of the population proportion, (1.645 being the critical score for the 90% confidence interval). Now if A is within 1.645 units of some fixed value B then B is within 1.645 units of A.

Therefore:

We can be 90% confident that the population proportion is within 1.645 standard deviations of the sample proportion.

Similarly:

We can be 95% confident that the population proportion is within 1.960 standard deviations of the sample proportion.

And:

We can be 99% confident that the population proportion is within 2.576 standard deviations of the sample proportion.

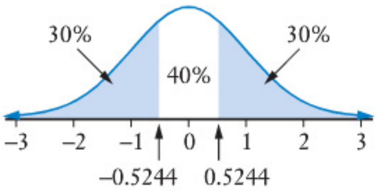
Thus, to infer **an interval estimate for the population proportion**, p, from a sample proportion, $\hat{p}$:

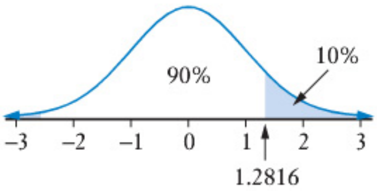
- Assume that the sample proportions are normally distributed with mean p and standard deviation $\sqrt{\hat{p}(1-\hat{p})/n}$. (This standard deviation is the *best we can do*).
- The interval estimate for p is then $\hat{p} - k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

  where k is the appropriate number of standard deviations for the required confidence interval.

i.e., to 3 dp,	k	$=$	1.645 for a 90% confidence interval
	k	$=$	1.960 for a 95% confidence interval
and	k	$=$	2.576 for a 99% confidence interval.

- **If we were to construct many 95% (or 90% or 99%) confidence intervals we would expect approximately 95% (or 90% or 99%) of them to contain the population proportion.**
- Whilst we have concentrated here on the 90%, 95% and 99% confidence intervals we could have intervals involving other percentages, say 40% as shown below left, or intervals not centred on the mean value, as shown below right. However we will tend to concentrate on the 90%, 95% and 99% intervals centred on the mean.





EXAMPLE 6

A survey of 1000 people found that 143 said they were satisfied with the way the government was running the country. Within what range of values (centred on the sample proportion) can we be 95% confident that the population proportion lies?

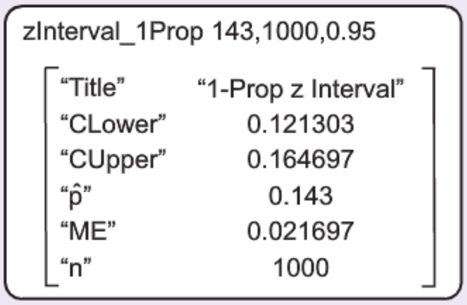
Solution

Using a standard deviation of 0.011, ($\sqrt{0.143(1-0.143)/1000} \approx 0.011$), the sample proportion of 0.143, and the critical value for the 95% confidence interval of 1.960 we have:

$$\begin{aligned} 0.143 - 1.960 \times 0.011 &= 0.121 \\ 0.143 + 1.960 \times 0.011 &= 0.165 \end{aligned}$$

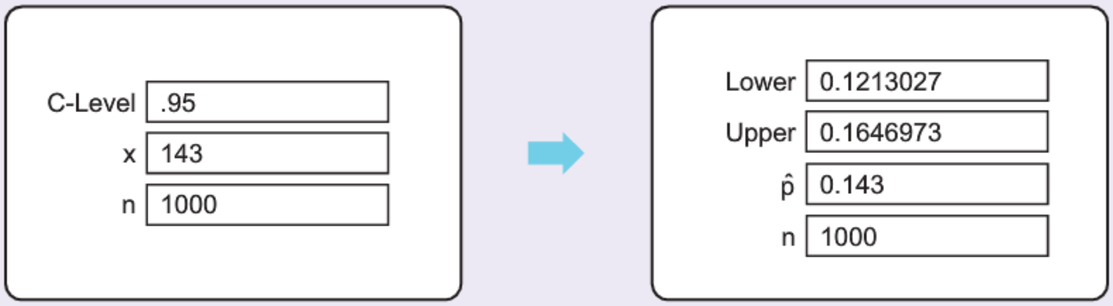
We can be 95% confident that the population proportion lies between 0.121 and 0.165.

Some calculators, given the sample proportion and sample size, can give confidence intervals for population proportions, as the displays shown below suggest.



```
zInterval_1Prop 143,1000,0.95

 “Title”      “1-Prop z Interval”
 “CLower”        0.121303
 “CUpper”        0.164697
 “p̂”               0.143
 “ME”            0.021697
 “n”               1000
```



```
C-Level  .95                Lower  0.1213027
      x  143                Upper  0.1646973
      n  1000                   p̂  0.143
                                n  1000
```

Note: The 95% confidence interval means that for such samples, 95% of such confidence intervals would contain the population proportion. We tend to express this by saying that we are 95% confident that our interval contains the population proportion, as in the previous example. However, if asked to interpret or explain the confidence interval it is perhaps best to include the fuller interpretation, as in the next example.

EXAMPLE 7

A survey of 500 drivers asks each person if they think the current penalties for using a mobile phone when driving are too harsh.

184 of the 500 say they do think the penalties are too harsh.

Find the 90% confidence interval for the population proportion and interpret your answer.

Solution

Using a standard deviation of 0.0216, ($\sqrt{0.368(1-0.368)/500}) \approx 0.0216$), the sample proportion of 0.368, and the critical value for the 90% confidence interval of 1.645:

$$\begin{aligned} 0.368 - 1.645 \times 0.0216 &= 0.3325 \\ 0.368 + 1.645 \times 0.0216 &= 0.4035 \end{aligned}$$

The 90% confidence interval for the population proportion is 0.3325 to 0.4035.

Interpretation: We could expect 90% of the 90% confidence intervals constructed in this way to contain the population proportion. Thus, with 90% confidence we estimate that between 33.25% and 40.35% of all drivers think that the current penalties for using a mobile phone when driving are too harsh.

The reader should confirm that these same values can be obtained from a calculator capable of giving confidence intervals.

Margin of error

In the previous example we could have expressed the 90% confidence interval as:

$$0.368 \pm 0.0355.$$

Using the minus gives us the lower bound: $0.368 - 0.0355 = 0.3325$

Using the plus gives us the upper bound: $0.368 + 0.0355 = 0.4035$

The value 0.0355 is the **margin of error**. It is the maximum amount that p can differ from $\hat{p}$ whilst still being in the confidence interval.

With the confidence interval given by:

$$\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \le \quad p \quad \le \quad \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

with a suitably chosen value of k, it follows that the margin of error is given by

$$k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

i.e.: The appropriate critical value from the standard normal distribution multiplied by the standard deviation of the sample proportions.

Sample size

Suppose that, in the previous example, instead of having the 90% confidence interval as

$$0.368 \quad \pm \quad 0.0355$$

we wanted it to be $\quad 0.368 \quad \pm \quad 0.025.$

I.e., we wanted the margin of error to be 0.025.

Then $\quad k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad = \quad 0.025$

For the 90% confidence interval we have $k = 1.645$, and the example had $\hat{p} = 0.368$.

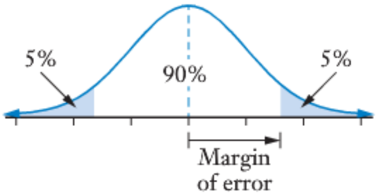
Thus $\quad 1.645 \times \sqrt{\frac{0.368(1-0.368)}{n}} \quad = \quad 0.025$

Solving gives $\quad n \quad = \quad 1006.97$

Remembering that n must be an integer, $\quad n \quad = \quad 1007$

The sample size should be 1007.

Check:



$$\text{Margin of error} \quad = \quad 1.645 \times \sqrt{\frac{0.368(1-0.368)}{1007}}$$

$$\approx \quad 0.025 \text{ as required.}$$

In this way, if we are planning a survey and already have some idea of what level of confidence we want, and to within what margin of error, we can choose the sample size appropriately. This would also require us to have some idea of the likely sample proportion but, to get started we could choose

$$\hat{p} \quad = \quad 0.5$$

as this will give the maximum value for $\hat{p}(1-\hat{p})$. This will make sure that whatever the value of $\hat{p}$, n will be such that the margin of error is at or within the desired level.

[Can you prove that $x(1-x)$ is maximised for $x = 0.5$?]

ISBN 9780170395205

EXAMPLE 8

A survey is to be carried out with the aim of having the 95% confidence interval on the population proportion with margin of error just 0.022. Taking the sample proportion as 0.5, find the sample size.

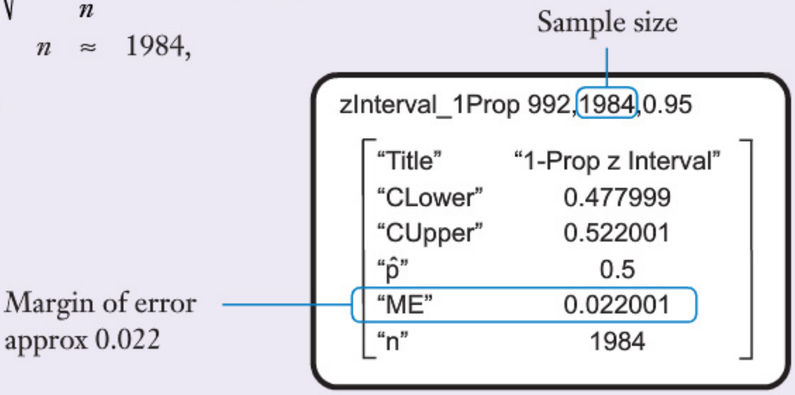
Solution

For the 95% confidence interval we have $k = 1.960$, and we are told that $\hat{p} = 0.5$.

Thus $$1.960\sqrt{\frac{0.5(1-0.5)}{n}} = 0.022.$$

Solving gives $$n \approx 1984,$$

The sample size needs to be 1984.



Sample size

```
zInterval_1Prop 992,1984,0.95

 "Title"      "1-Prop z Interval"
 "CLower"         0.477999
 "CUpper"         0.522001
 "p̂"              0.5
 "ME"             0.022001
 "n"              1984
```

Margin of error approx 0.022

- If it is important that our margin of error does not exceed 0.022 we would round n to the next integer **up**, i.e. 1985.

Increasing the level of confidence increases the margin of error

Suppose a survey of 500 people in Australia finds that 122 reply that they have lost confidence with the democratic system.

Using the 90% confidence interval we could say that:

We are 90% confident that the proportion of people in the whole Australian population who would say that they have lost confidence with the democratic system would lie in the interval 0.2124 to 0.2756. (Check that you can obtain these figures.)

I.e., with 90% confidence our estimate for the population proportion is 0.244 ± 0.0316.

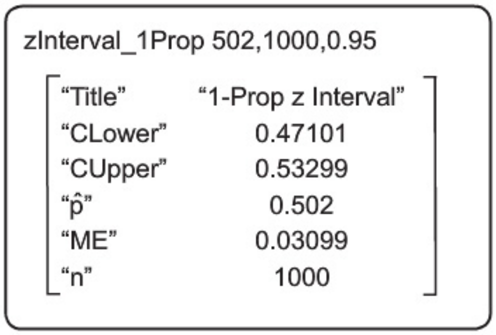
If we want to increase our level of confidence, to say 95% or even 99%, still using the 122 out of 500 from our sample, we have to accept the increase in the margin of error. In other words, by increasing the size of our interval we can be more confident that it includes the population proportion.

Check that you agree with the following statements for this situation and notice how the margin of error increases as the level of confidence increases (for fixed n and $\hat{p}$).

The 95% confidence interval is 0.2064 to 0.2816.
I.e., with 95% confidence our estimate for the population proportion is 0.244 ± 0.0376.

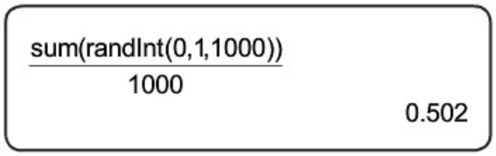
The 99% confidence interval is 0.1945 to 0.2935.
I.e., with 99% confidence our estimate for the population proportion is 0.244 ± 0.0495.

Let's check



```
zInterval_1Prop 502,1000,0.95

 "Title"     "1-Prop z Interval"
 "CLower"        0.47101
 "CUpper"        0.53299
 "p̂"             0.502
 "ME"            0.03099
 "n"             1000
```



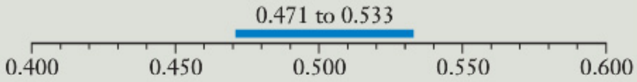
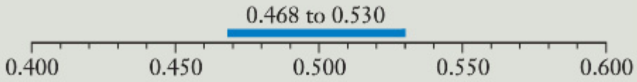
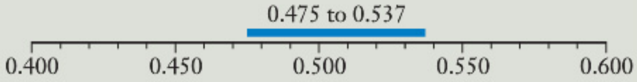
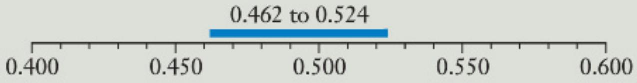
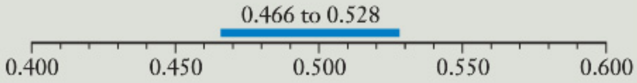
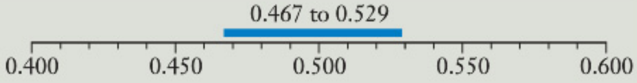
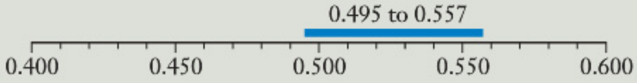
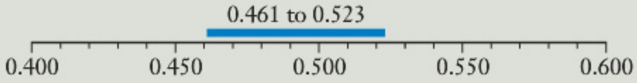
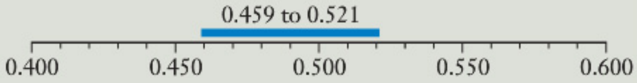
```
sum(randInt(0,1,1000))
        1000
                                0.502
```

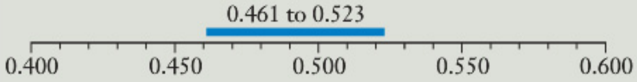
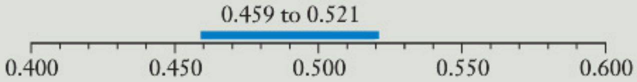
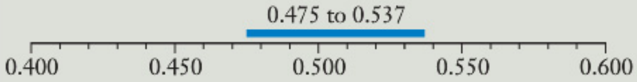
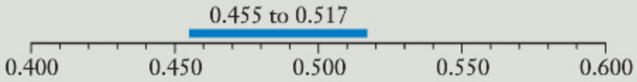
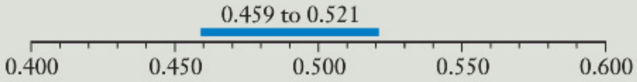
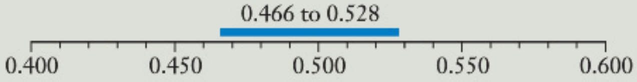
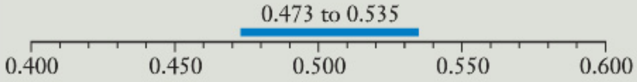
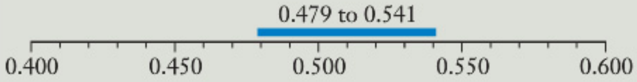
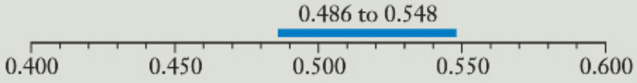
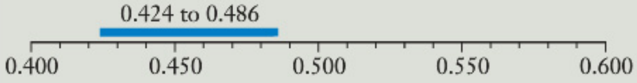
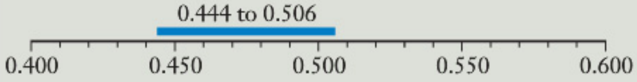
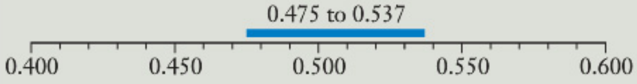
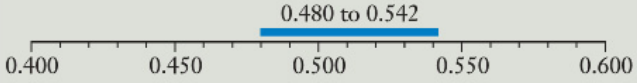
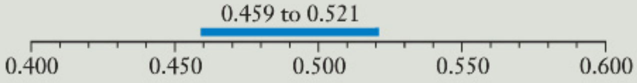
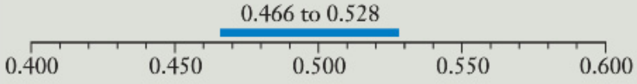
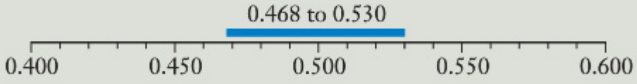
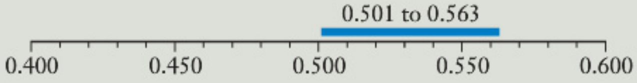
Suppose we toss a fair coin 1000 times and obtain a head on 502 occasions, i.e. the proportion of heads is 0.502, or 50.2%. Using this proportion, the 95% confidence interval would be 0.471 to 0.533, i.e. 0.502 ± 0.031.

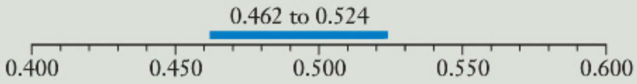
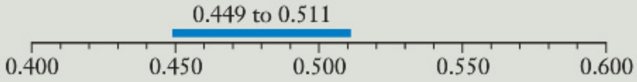
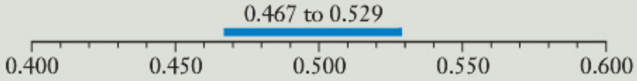
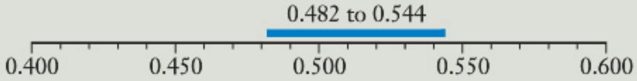
If we took many such samples of 1000 tosses of the coin we would expect approximately 95% of the *95% confidence intervals* so produced to contain the population proportion.

Of course, in most situations, we would probably be carrying out the sampling to estimate the population proportion. However, in this fair coin situation we have the luxury of knowing the long term proportion so let us simulate the situation a number of times and see if we do indeed find that most, but not all of the 95% confidence intervals so produced do indeed contain 0.5, the population proportion.

Thirty such samples each of size 1000 are shown below, and on the next two pages. As can be seen, at least for these 30 samples, it is indeed the case that most, but not all of the 95% confidence intervals do contain 0.5, the population proportion, *p*.

Sample	Sample proportion	95% confidence interval	Interval contains 0.5
1	0.502		✓
2	0.499		✓
3	0.506		✓
4	0.493		✓
5	0.497		✓
6	0.498		✓
7	0.526		✓
8	0.492		✓
9	0.490		✓

10	0.492	 0.461 to 0.523	✓
11	0.490	 0.459 to 0.521	✓
12	0.506	 0.475 to 0.537	✓
13	0.486	 0.455 to 0.517	✓
14	0.490	 0.459 to 0.521	✓
15	0.497	 0.466 to 0.528	✓
16	0.504	 0.473 to 0.535	✓
17	0.510	 0.479 to 0.541	✓
18	0.517	 0.486 to 0.548	✓
19	0.455	 0.424 to 0.486	✗
20	0.475	 0.444 to 0.506	✓
21	0.506	 0.475 to 0.537	✓
22	0.511	 0.480 to 0.542	✓
23	0.490	 0.459 to 0.521	✓
24	0.497	 0.466 to 0.528	✓
25	0.499	 0.468 to 0.530	✓
26	0.532	 0.501 to 0.563	✗

27	0.493		✓
28	0.480		✓
29	0.498		✓
30	0.513		✓

Carry out some similar simulations yourself, perhaps for coin tossing or die rolling and for 95%, 90% or even 80% confidence intervals, and comment on your findings.

Exercise 6B

(Note: Survey results referred to in this exercise are for the purpose of this exercise only and do not necessarily reflect any real situation.)

1 A survey of 1200 people living in Australia found that 450 were in favour of the idea of introducing compulsory national service. Find the 95% confidence interval for the population proportion and interpret your answer.

2 A survey involving 800 people living in Australia, and who had recently contacted their bank for online help, found that 680 of the 800 were either satisfied or very satisfied with the service they received. Find the 90% confidence interval for the population proportion and interpret your answer.

3 A survey of 250 people who all regularly played a particular sport found that 190 of the 250 agreed that the recent rule changes were a good idea. Find the 99% confidence interval for the population proportion and interpret your answer.

4 With a sample size of 880 and a sample proportion of 70% state the 95% confidence interval for the population proportion in the form $a\% \pm b\%$, with b given correct to the nearest integer.

5 A national opinion poll surveyed 2000 people and found that 45% wanted to see changes to the current daylight saving rules.

Find the 90% confidence interval for the population proportion in the form $a\%$ to $b\%$, with a and b each given correct one decimal place. Interpret your answer.

When all 200 people in a particular community were surveyed it was found that 140 of the 200 wanted to see changes to the daylight saving rules. Comment.

6 The testing of 1000 seeds from a particular batch of millions of such seeds found that 28% failed to germinate. Use this information to write a statement about the batch involving the idea of being 95% confident.

7 For a sample size of 2000 and sample proportion 0.45, find the margin of error at the 95% confidence level.

8 For a sample size of n and sample proportion b, which out of the 90% confidence interval and the 99% confidence interval has the larger margin of error?

9 A sample of 200 machine components made by a particular machine finds that 36 of the components are deemed to be of an unacceptable standard. (A component can be judged to be either acceptable or unacceptable.)

- **a** What is the proportion of acceptable components in this sample?
- **b** Write a statement about the percentage of acceptable components in the population of components made by this machine including in your statement the idea of being 90% confident.
- **c** Write a statement about the percentage of acceptable components in the population of components made by this machine including in your statement the idea of being 99% confident.

10 A survey is to be carried out with the aim of having the 95% confidence interval on the population proportion with margin of error just 0.065. Taking the sample proportion as 0.5 find the sample size.

11 A survey is to be carried out with the aim of having the 90% confidence interval on the population proportion with margin of error just 0.03. Taking the sample proportion as 0.5 find the sample size.

12 A survey is to be carried out with the aim of having the 95% confidence interval for the population proportion equal to, or within, 0.60 to 0.70. Taking the sample proportion as 0.65 find the sample size.

13 Let us suppose that in a survey of 500 Australian males aged between 20 and 30 it was found that 76% of the males were taller than their father.

According to the results of this survey copy and complete the following statement (give percentages to the nearest whole percent):

We can be 95% confident that of all Australian males between the ages of 20 and 30,
between __% and __% are taller than their father.

If we wanted to be 99% confident would our interval be larger or smaller than the 95% interval? Explain why.

14 A government inquiry wanted to estimate the proportion of Australians who possessed a particular attribute. The results from a random sample of Australians led to the calculation of the 90% confidence interval for the population proportion who have the particular attribute as 0.241 ± 0.060.

Copy and complete the following statement, giving percentages to the nearest whole percent.

We can be 90% confident that the proportion of Australians having
the particular attribute lies between __% and __%.

Determine the sample size and the number in the sample possessing the attribute.

15 Let us suppose that a random sample of 480 year twelve Australian school students were surveyed and 168 of the students said that they intended proceeding to University the following year.

- **a** Calculate the sample proportion of these year 12s intending to proceed to University the following year.
- **b** We would expect all such sample proportions for samples of this size to approximate to a normal distribution. Calculate the standard deviation of this normal distribution (rounded correct to four decimal places).
- **c** Calculate the 95% confidence interval for the population proportion and interpret your answer.
- **d** A second random sample is planned but this time the organisers would like the 95% confidence interval to involve a margin of error of, at most, 3%.
  Calculate the sample size.

Miscellaneous exercise six

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

Answer questions **1**, **2**, **3** and **4** *without* the assistance of your calculator.

1 Determine the value of x in each of the following:

a $\log_x 64 = 6$ **b** $\log_2 x = 3$

c $\log x = 2$ **d** $\ln e = x$

2 Differentiate

a $x\sqrt{x}$ **b** $4x^5 + \log_e x$

c $7\ln x$ **d** $\ln(5x^3 - 6x)$

3 Find $\dfrac{dy}{dx}$ for each of the following

a $y = \log_e(5x - 1)$ **b** $y = \ln(x^4 + 1)$ **c** $y = \ln[(x + 1)(x - 1)]$

4 Determine the equation of the tangent to $y = 3 - \ln x$ at the point $(e, 2)$.

5 Two fair coins are flipped 160 times and come down with both heads facing upwards on 46 of the 160 occasions.

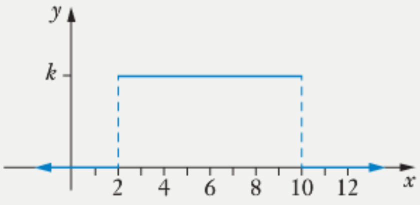
a What is the value of p, the population proportion of obtaining two heads?

b What is the value of $\hat{p}$, the sample proportion of two heads for the 160 flips?

c Calculate the mean and standard deviation of the random variable $\hat{p}$, for samples involving 160 flips of two coins.

d How many standard deviations from p is our value for $\hat{p}$?

6 The continuous random variable X has the probability density function shown on the right.



Determine **a** k

b $P(X \geq 4)$

c $P(X < 8)$

d $P(X > 3 \mid X < 7)$

7 A parent suggests that when four year olds are given a new fibre-tip pen the number of minutes until they manage to lose the cap of the pen forms a random variable with probability density function:

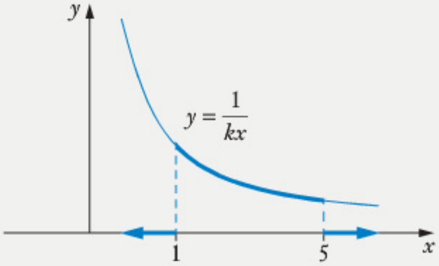
$$f(x) = \begin{cases} 0.5e^{-0.5x} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

According to this suggested rule, find the probability of a four year old losing the cap within 5 minutes of being given the pen.

8 If $f'(x) = \dfrac{x+5}{x}$ and $f(1) = 5$ determine expressions for $f''(x)$ and $f(x)$.

9 A continuous random variable, X, has the probability density function

$$f(x) = \begin{cases} \dfrac{1}{kx} & \text{for } 1 \le x \le 5 \\ 0 & \text{for all other values of } x. \end{cases}$$



Determine each of the following, giving your answers as exact values.

a The value of k,

b $P(2 \le X \le 4)$.

10 $X \sim N(24, 64)$, i.e. X, is normally distributed with a mean of 24 and a standard deviation of 8. Given that $P(X > k) = 0.2266$ determine k.

11 The response times on a psychological test were found to be approximately normally distributed with mean 2.32 seconds and standard deviation 0.48 seconds.

a Using this normal distribution as the basis for prediction, what is the probability that a randomly chosen individual would achieve a time of less that 1 second in this test?

b Given that a randomly chosen individual achieved a time that was less than the mean, what is the probability that they achieved a time that is less than 1 second.

12 A random variable X is normally distributed with mean μ and standard deviation σ. If $P(X < 25)$ is 0.1082 and $P(X > 42)$ is 0.1303 determine

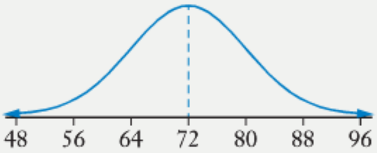
a $P(25 < X < 42)$, giving your answer correct to two decimal places,

b the values of μ and σ, giving answers correct to 1 decimal place,

c $P(X > 22)$, giving your answer correct to two decimal places.

13 A coin is tossed 1000 times and a head is obtained on 521 occasions. Does this suggest the coin is unfair? Comment.

14 In an examination for which the marks are approximately normally distributed with a mean score of 72 and a standard deviation of 8, 85% of the candidates pass.



What is the pass mark? (Round your answer to the nearest 0.5%.)

15 If we roll five normal fair six sided dice and add the numbers on the uppermost faces our answer could be from a low of 5 to a high of 30.

Use a calculator or computer to simulate approximately 100 such rolls of 5 dice and record your answers grouped as follows:

Score	≤ 9	$10 \rightarrow 13$	$14 \rightarrow 17$	$18 \rightarrow 21$	$22 \rightarrow 25$	≥ 26
Frequency						

The probability distribution of possible total scores can be modelled using a normal distribution with mean 17.5 and standard deviation 3.8.

Compare your distribution with that given theoretically by the suggested normal distribution model, i.e. $N(17.5, 3.8^2)$, with the necessary adjustment for continuity being made. (See page 96 for an explanation of adjusting for continuity.)

16 Let us suppose that a survey of mobile phone use amongst Australian adults found that the 95% confidence interval for the proportion of adult Australians who were 'mobile only', i.e. they did not have land line phone access in their place of residence but relied totally on their mobile phone, was 0.19 ± 0.02.

Explain what this 0.19 ± 0.02 confidence interval means.

In view of the above survey comment on the following:

- A survey of 500 Australians aged 65 or over found that 20 were 'mobile only'.
- A survey of 800 Australians aged in their twenties found that 336 were 'mobile only'.

17 At an election one of the minor parties polled 22% of the vote. Six months later a survey of 400 people in the electorate showed that just 20% supported the party. Could we conclude that the party has lost popularity or could the change be explained by expected variation in sample proportions? Explain.

Suppose the 20% figure had instead come from a survey involving 4000 people.

18 A company makes a batch of batteries and rejects them on the basis that tests show that about 15% of the batteries don't work.

A second company buys the batteries and sells them cheaply with the statements:

> Batteries for sale at half price:
> 15% of these batteries don't work but at half price that
> means the 85% you are getting that do work are cheap.

Joe complains to a consumer protection organisation claiming that he bought 100 of the batteries and found that 17 were faulty.

Comment on the above.

ANSWERS

Exercise 1A PAGE 8

1 $2^3 = 8$ **2** $7^2 = 49$ **3** $49^{0.5} = 7$ **4** $10^3 = 1000$ **5** $5^4 = 625$ **6** $4^{2.5} = 32$

7 $5^{-2} = 0.04$ **8** $3^{-2} = \frac{1}{9}$ **9** $a^y = x$ **10** $b^c = y$ **11** $x^p = a$ **12** $a^3 = x$

13 $3^y = 5$ **14** $2^x = 3$ **15** $x^4 = 5$ **16** $3^p = 5$ **17** $\log_2 64 = 6$ **18** $\log_3 81 = 4$

19 $\log_9 81 = 2$ **20** $\log_9 27 = \frac{3}{2}$ **21** $\log_2 0.5 = -1$ **22** $\log_2 0.25 = -2$ **23** $\log 100 = 2$ **24** $\log 0.01 = -2$

25 $\log_p r = q$ **26** $\log_r q = p$ **27** $\log_2 y = x$ **28** $\log_3 z = y$ **29** $\log_5 4 = k$ **30** $\log_7 3 = y$

31 $\log_3 7 = p$ **32** $\log_e x = y$ **33** 2 **34** 7 **35** 4 **36** 5

37 -1 **38** -4 **39** -3 **40** -3 **41** $\frac{5}{2}$ **42** -3

43 1 **44** 0 **45** 0 **46** $\frac{5}{2}$ **47** 1 **48** 3

49 0.699 **50** 1.398 **51** 0.845 **52** 1.690 **53** 1.301 **54** 1

55 1.322 **56** 1

57 **a** Yes **b** No (If we are restricting our attention to real numbers, as we are in this unit.)

Exercise 1B PAGE 11

1 $\log(xz)$ **2** $\log(x^2y)$ **3** $\log(x^2y^3)$ **4** $\log\left(\frac{x^2}{y}\right)$ **5** $\log\left(\frac{ab}{c}\right)$ **6** $\log\left(\frac{a^3b^4}{c^2}\right)$

7 $\log(c^2a)$ **8** $\log(100x)$ **9** $\log\left(\frac{1000}{xy}\right)$ **10** $\log\left(\frac{1000y}{x}\right)$ **11** 3 **12** 4

13 3 **14** 1 **15** 2 **16** -4 **17** -1 **18** 2

19 1.5 **20** 4

21 **a** $p+q$ **b** $p+2q$ **c** $2p+q$ **d** $p-q$ **e** $2q+4$ **f** $p-2q$

22 **a** $2a$ **b** $a+2b$ **c** $a-2b$ **d** $b+2$ **e** $2a+b+1$ **f** $a+2b+2$

23 $y = a^x$ **24** $y = 2x$ **25** $y = x^3$ **26** $y = x^{\frac{3}{2}}$ **27** $y = ax$ **28** $y = a^2x$

29 $y = \frac{1}{x}$ **30** $y = \frac{a^2}{x}$

31 **a** 75 **b** ~58 **c** ~51 **d** 9

32 **a** 3 **b** $10^{5.4}I_0$ **c** 10 **d** $10^{1.8} (\approx 63)$

33 **a** 4 **b** 4.5 (approximately) **c** 6.6 (approximately) **d** 7.8 (approximately)

e 7.4 (approximately), 0.000 005 6 moles/litre (approximately).

34 **a** $10^4 I_0$ **b** $10^7 I_0$ **c** 10^7

Exercise 1C PAGE 14

1 $x = \dfrac{\log 7}{\log 3}$

2 $x = \dfrac{3}{\log 7}$

3 $x = \log 27$ (i.e. $3\log 3$)

4 $x = \dfrac{\log 11}{\log 2}$

5 $x = \dfrac{\log 17}{\log 3}$

6 $x = \dfrac{\log 80}{\log 7}$ $\left(\text{i.e. } \dfrac{1+3\log 2}{\log 7}\right)$

7 $x = \dfrac{\log 21}{\log 5}$

8 $x = \log 15$

9 $x = \dfrac{\log 70}{\log 2}$ $\left(\text{i.e. } \dfrac{1+\log 7}{\log 2}\right)$

10 $x = \dfrac{\log 17}{\log 6} - 2$ $\left(\text{i.e. } \dfrac{\log\left(\frac{17}{36}\right)}{\log 6}\right)$

11 $x = \dfrac{\log 17}{\log 3}$

12 $x = \dfrac{\log 7}{\log 8} + 1$ $\left(\text{i.e. } \dfrac{\log 56}{3\log 2}\right)$

13 $x = \dfrac{\log 5}{\log\left(\frac{5}{9}\right)}$

14 $x = \dfrac{\log 2}{\log 1.5}$

15 $x = \dfrac{2\log 5}{\log\left(\frac{64}{5}\right)}$

16 $x = \dfrac{\log 6}{\log\left(\frac{8}{9}\right)}$

17 $x = -\dfrac{\log 3}{\log 2}$

18 $x = \dfrac{\log 3}{\log 5}$

19 $x = \dfrac{\log 3}{\log 2}$

20 $x = \dfrac{\log 3}{\log 2}, x = \dfrac{\log 5}{\log 2}$

21 $x = \dfrac{\log 7}{\log 2}$

22 **a** $\log_3 5 = \dfrac{\log 5}{\log 3}$ **b** $\log_2 12 = \dfrac{\log 12}{\log 2}$ **c** $\log_9 15 = \dfrac{\log 15}{\log 9}$

d $\log_9 4 = \dfrac{\log 4}{\log 9}$ $\left(\text{i.e. } \dfrac{\log 2}{\log 3}\right)$ **e** $\log_{2.5} 6.8 = \dfrac{\log 6.8}{\log 2.5}$ **f** $\log_{5.4} 9 = \dfrac{\log 9}{\log 5.4}$

23 The metal should be passed through the rollers 20 times.

24 **a** ~270 **b** ~330 **c** The population first exceeded 1000 on the 17th day ($t \approx 16.2$).

25 The risk of an accident is 51% for $a = 0.191$

26 **a** Approximately 211 000. **b** Approximately 182 000.
Sales fall to 135 000 bars per week approximately 19 weeks after the first campaign ceases.

27 **a** \$12 597.12 **b** \$17 138.24 **c** 21 years **d** **i** 17 years **ii** 14 years
e 14.9%

Exercise 1D PAGE 18

1 1

2 –1

3 3

4 0.5

5 $\dfrac{1}{3}$

6 $-\dfrac{1}{2}$

7 –3

8 $-\dfrac{1}{3}$

9 $\ln 7 - 1$

10 $\ln 50 - 3$

11 $2\ln 10 + 3$

12 $\dfrac{\ln 15 - 1}{2}$

13 $\dfrac{\log_e 600 + 1}{3}$

14 $3\ln 10 - 2$

15 $\ln 10, \ln 20$

16 $\dfrac{\ln 2}{\ln 7}$

17 $\dfrac{\ln 3 + \ln 7}{\ln 2}$

18 $\dfrac{3\ln 2 + 2\ln 5}{\ln 3}$

19 $\dfrac{\ln 2 + 2\ln 5}{\ln 5}$

20 $\dfrac{2\ln 3}{\ln 2 + \ln 3}$

21 $\dfrac{\ln 2 + \ln 3}{2\ln 3}$

22 $\dfrac{\ln 3 + 2\ln 2 + 2\ln 5}{2\ln 2}$

23 $\dfrac{\ln 11 + 2\ln 2 + \ln 5}{3\ln 2}$

24 $\ln\left(\dfrac{2000}{A}\right)$ **a** 0.288 **b** 1.386 **c** 3.689

25 **a** 2028 **b** 2045 **26** **a** 5 days **b** 9 days

Exercise 1E PAGE 20

1 **a** $(-7, 0)$ **b** $(0, 3)$

2 $(1, 0)$ **3** $(a, 1)$

4 **a** $x = 0$, i.e. the y-axis. **b** $x = 3$ **c** $x = 0$

5 Difficult to be accurate from the graph but answers for **a** to **d** should be close to the following:

a $x \approx 2.2$ **b** $x \approx 11.2$ **c** $x \approx 3.6$ **d** $x \approx 9.1$

e Rounded to 3 dp: 2.236, 11.180, 3.624, 9.103

6 $a = 2, b = 4, c = 3$

Exercise 1F PAGE 23

1 **a** 7.19 (correct to 2 dp) **b** 1.58×10^{-10} (correct to 3 sig figs)

2 **a** Approximately 1.32 octaves **b** Higher frequency $= 8f_1$

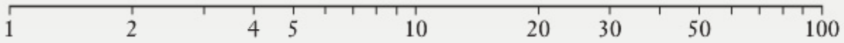
3 **a** 10^{-7} moles per litre **b** 2

4 **a** -1.39 **b** 0.98 **c** The event is more likely not to occur than to occur.

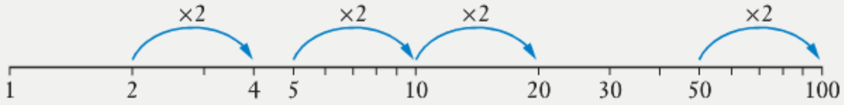
d Compare your answer to that of others.

5 Whilst it is true that the amplitude of the vibrations caused by an earthquake of magnitude 7 will be ten times that of an earthquake of magnitude 6 the cost of the damage will depend on other things too. For example, an earthquake measuring 6 on the Richter scale with its epicentre at a location of high density housing and infrastructure could cause more costly damage than an earthquake of scale 7 occurring in an uninhabited desert area. Hence the cost of damage caused by an earthquake of Richter scale 7 will not necessarily be ten times that of one with a Richter scale of 6.

6 In a logarithmic scale, if 1 to 10 is 1 unit of length then 1 to 2 will be log 2 units, 1 to 5 will be log 5 units, 1 to 20 will be log 20 units, etc. Thus on our diagram, with the distance from 1 to 10 being 5 cm the distance from 1 to 2 will be 5 cm × log 2, the distance from 1 to 5 will be 5 cm × log 5 etc.



Note also that the distance from 2 to 4 (a doubling) is the same as the distance from 5 to 10 (also a doubling) and the same as the distance from 10 to 20 and the same as the distance from 50 to 100.



This is as we would expect for a logarithmic scale because

$$\begin{aligned} \log 4 - \log 2 &= \log 10 - \log 5 \\ &= \log 20 - \log 10 \\ &= \log 100 - \log 50 \end{aligned}$$

because all are of the form

$$\begin{aligned} \log(2a) - \log a &= \log(2a \div a) \\ &= \log 2 \end{aligned}$$

Miscellaneous exercise one PAGE 25

1 $15x^2$ **2** $3x^2 + 1$ **3** $\dfrac{11}{(2x+5)^2}$ **4** $12x^2(x^3+1)^3$ **5** e^x

6 $2e^x$ **7** $10e^x$ **8** $e^x + 6x + 3x^2$ **9** $5e^{5x}$ **10** $12e^{4x}$

11 $6e^{2x}$ **12** $6e^{3x} + 6e^{2x}$ **13** $3^4 = 81$ **14** $6^3 = 216$ **15** $2^{-2} = 0.25$

16 $a^c = b$ **17** $a^b = c$ **18** $b^c = a$ **19** $c^b = a$ **20** $x^5 = 2$

21 $\log_2 8 = 3$ **22** $\log_5 25 = 2$ **23** $\log_4 0.25 = -1$ **24** $\log_2 0.125 = -3$ **25** $\log_7 y = x$

26 $\log_a p = 2$ **27** $\log_{10} z = y$ **28** $\log_e x = y$ **29** 5 **30** 3

31 1 **32** 3 **33** 6 **34** 2 **35** 3

36 2 **37** 0 **38** 1 **39** 3 **40** 0.5

41 $\ln 12 - 1$ **42** $\ln 25 - 2$ **43** $\ln 150 + 1$ **44** $\frac{\ln 34 - 1}{2}$ **45** $\ln 25 - 1$

46 $\ln 5, \ln 7$ **47** $\log(x^3y)$ **48** $\log\frac{x^2}{y^3}$ **49** $\log\frac{a^2b}{c^3}$ **50** $\log(1000x)$

51 $\ln(e^2x)$ **52** $\ln\frac{e^3y^2}{x}$ **53** 2026

54 **a** $(e-1)$ m/s, (Approximately 1.72 m/s.) **b** $10(e-2)$ m, (Approximately 7.18 m.)
c $10e^{0.1T}(e^{0.1}-1)-1$ m **d** 0.285 m **e** 1.587 m

Exercise 2A PAGE 32

1 $\frac{1}{x}$ **2** $\frac{1}{x}$ **3** $10x + \frac{1}{x}$ **4** $1 + e^x + \frac{1}{x}$ **5** $\frac{3}{3x+2}$

6 $\frac{2}{2x+3}$ **7** $\frac{2}{2x-3}$ **8** $\frac{2x}{x^2+1}$ **9** $-\tan x$ **10** $\frac{2}{x}$

11 $\frac{1}{3x}$ **12** $\frac{1}{2x}$ **13** $\frac{1}{x}$ **14** $\frac{2x+3}{x(x+3)}$ **15** $\frac{2x+1}{(x+4)(x-3)}$

16 $1 + \log_e x$ **17** $\frac{3}{x}(\log_e x)^2$ **18** $-\frac{1}{x}$ **19** $-\frac{1}{x(\log_e x)^2}$ **20** $e^x\log_e x + \frac{e^x}{x}$

21 $\frac{1-\log_e x}{x^2}$ **22** $\frac{3(1+\log_e x)^2}{x}$ **23** $\frac{1}{x} + \frac{1}{x+5} + \frac{1}{x+3}\left(= \frac{3x^2+16x+15}{x(x+5)(x+3)}\right)$

24 $\frac{1}{x+1} - \frac{1}{x+3}\left(= \frac{2}{(x+1)(x+3)}\right)$ **25** $\frac{8x}{x^2+5}$ **26** $\frac{1}{x} - \frac{2x}{x^2-1}\left(= \frac{1+x^2}{x(1-x^2)}\right)$

27 $\frac{3}{x+2} - \frac{1}{x-2}\left(= \frac{2(x-4)}{(x+2)(x-2)}\right)$ **28** 7 **29** 3 **30** 7

31 -2 **32** $(4, \log_e 4)$ **33** $(0.5, -\log_e 4)$ **34** $(5, \log_e 25)$ **35** $(3, \log_e 18)$

36 $y = x - 1$ **37** $ey = x$ **38** $\frac{1}{x\ln 4}$ **39** $\frac{1}{x\ln 6}$

40 Approximate change in y is 0.5
$50\ln 10.1 - 50\ln 10 = 0.4975$, correct to four decimal places.

41 1.5 m/s, –0.25 m/s^2

42 Minimum point at $(5, 25 - 50\ln 10)$

Exercise 2B PAGE 37

Note: At the time of writing the syllabus for this unit includes

$$\int\frac{1}{x}\,dx \text{ for } x > 0, \text{ and } \int\frac{f'(x)}{f(x)}\,dx \text{ for } f(x) > 0.$$

Thus whilst $\int\frac{1}{x}\,dx = \ln|x| + c,\ x \neq 0$, see the note on page 34, the restriction $x > 0$ (and $f(x) > 0$) makes the absolute value unnecessary.

Hence answers to this exercise are given here without the absolute value symbol.

1 $5\ln x + c$ **2** $4\ln x + c$ **3** $\frac{x^2}{2} + 2\ln x + c$ **4** $\frac{1}{2}\ln x + c$

5 $\ln(x^2+1)+c$ **6** $\dfrac{x^3}{3}+5\ln x+c$ **7** $2x^2+e^x+2\ln x+c$ **8** $2\ln(x+1)+c$

9 $4\ln(x^2-3)+c$ **10** $\ln(5x-3)+c$ **11** $5\ln(2x+1)+c$ **12** $3\ln(x^2+1)+c$

13 $\ln(x^2+x+3)+c$ **14** $3\ln(x^2+5x)+c$ **15** $10\ln(x^2+4)+c$ **16** $-\ln(\cos x)+c$

17 $\ln(\sin x)+c$ **18** $-\dfrac{1}{2}\ln(\cos 2x)+c$ **19** $-\ln(\cos x)+c$ **20** $-\dfrac{1}{5}\ln(\cos 5x)+c$

21 $-3\ln(\cos 2x)+c$ **22** $-\ln(\sin x+\cos x)+c$ **23** $\dfrac{1}{2}\ln(4x+\sin 2x)+c$ **24** $\ln(e^x+x)+c$

25 $\ln 3$ **26** $3\ln 1.5$ **27** $e^2-e+\ln 2$ **28** $x=\ln\left(\dfrac{t+2}{2}\right)$

29 $(4+\ln 3)$ units^2 **30** $(1-\ln 2)$ units^2 **31** $(2e\ln 2+1)$ units^2 **32** $\ln\left(\dfrac{2}{\sqrt{3}}\right)$ units^2

33 $a=3,\ b=5,\ 3\ln(x+4)+5\ln(x+2)+c$

34 a $e^{0.5}$ **b** $e^{0.25}$ **c** $2\ln\left(\dfrac{1+e^{0.5}}{2}\right)$

Miscellaneous exercise two PAGE 39

1 $\dfrac{dy}{dx}=2\cos 2x$ **2** $\dfrac{dy}{dx}=-3\sin 3x$ **3** $\dfrac{dy}{dx}=4e^{4x}$ **4** $\dfrac{dy}{dx}=20e^{4x}$

5 $\dfrac{dy}{dx}=\dfrac{5}{(x+1)^2}$ **6** $\dfrac{dy}{dx}=12(3x-1)^3$ **7** $\dfrac{dy}{dx}=\dfrac{2}{x}$ **8** $\dfrac{dy}{dx}=2x\log_e x+x$

9 $\dfrac{dy}{dx}=-\dfrac{1}{x^2}+6e^{2x}$ **10** $\dfrac{dy}{dx}=\dfrac{1+2x}{1+x+x^2}$ **11** $\dfrac{\log 11}{\log 2}$

12 a $2p$ **b** $3p+q$ **c** $p+2q$ **d** $p+0.5q$ **e** $p+q+3$ **f** $\dfrac{q}{p}$

13 a 4 **b** 8 **c** 2 **d** 2 **e** 8.5 **f** 34
g 0.5 **h** 8

14 a $p=xy$ **b** $p=x^y$ **c** $p=\dfrac{x^3}{y}$ **d** $p=100\sqrt{y}$

15 $e^2y=x+e^2$ **16** 0.88^tQ_0. The pump must work for approximately 23.4 minutes.

17 a $6x\ln(3x+2)+\dfrac{9x^2}{3x+2}$ **b** $1.8+6\ln 5$

18 a $A(e^{-1},0)$, $B(e,0)$ **b** $(1,-1)$ **c** Point $B(e, 0)$ is the only point of inflection.

Exercise 3A PAGE 47

1 a $\dfrac{163}{186}$ **b** $\dfrac{64}{93}$ **c** $\dfrac{2}{93}$

2 a 36% **b** 15% **c** 24% **d** 48% **e** 24% **f** 53%

3 a 20 **b i** 0.82 **ii** 0.18 **iii** $\dfrac{8}{41}$

4 a 0.65 **b** 0.65 **c** 0.825 **d** $\dfrac{5}{26}$ **e** 0.8

5 a 0.28 **b** 0.72 **c** 0.67 **d** 0.64

Exercise 3B PAGE 54

1 0.25 **2** 0.05 **3** 2 **4** 2.5 **5** 1.8 **6** 4.5

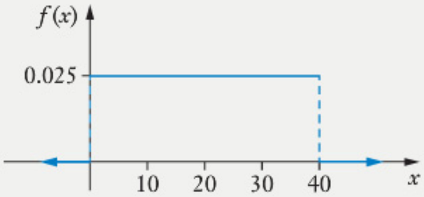
7 15 **8** 17.5 **9** $f(x)=\begin{cases} 0.5 & \text{for} \quad 1\le x\le 3 \\ 0 & \text{for all other values of } x. \end{cases}$

10 **a** 0.25 **b** 0 **c** 0.75 **d** $\frac{2}{3}$

11 **a** 1 **b** 0.3 **c** 0 **d** 1 **e** 0.625

12 **a** 25 **b** 0 **c** 0.4 **d** 0.4 **e** 0.8 **f** 1
g 0.2 **h** 0

13



a 0.5 **b** 0.625 **c** 0.2

Exercise 3C PAGE 59

1 0.5 **2** 0.25 **3** 1.25 **4** 2.5 **5** $\frac{1}{15}$ **6** $\sqrt{\frac{2}{\pi}}$

7 0.5 **8** 1.8 **9** 1.6 **10** 2.5 **11** 24 **12** $\frac{4}{15}$

13 $\frac{1}{\sqrt{e}-1}$ **14** 2.4 **15** 0.25 **16** 2

17 **a** $f(x)=\begin{cases} 0.125x & \text{for} \quad 0\le x\le 4 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$ or $f(x)=\begin{cases} 0.125x & \text{for} \quad 0< x\le 4 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$

b $f(x)=\begin{cases} 0.25 & \text{for} \quad 1\le x\le 2 \\ 0.5 & \text{for} \quad 2< x< 3 \\ 0.25 & \text{for} \quad 3\le x\le 4 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$

18 **a** 0.25 **b** 0.4375 **c** 1 **d** $\frac{5}{9}$

19 **a** 0.5 **b** 0.125 **c** 0.875 **d** $\frac{3}{7}$

20 $h(x)$ could be a probability density function for $0\le x\le 5$.

$\int_0^5 f(x)dx \ne 1$. Thus $f(x)$ cannot be a probability density function for $0\le x\le 5$.

$g(x)$ is negative for $\frac{25}{6}< x\le 5$.

Thus $g(x)$ cannot be a probability density function for the interval $0\le x\le 5$.

21 **a** 0.84 **b** 0.28 **c** $\frac{7}{16}$

22 **a** 0.08 **b** 0.64 **c** 0.48 **d** 0.75

23 **b** 0.25 **c** 0.15 **d** 0.6

24 **a** $-\frac{8}{3}$ **b** No. For some values in $1\le x\le 4$, $f(x)$ is –ve. Thus $f(x)$ cannot be a pdf for $1\le x\le 4$.

ISBN 9780170395205

25 **a** 0.75 **b** 0.5 **c** 0.84 **d** 1.87

26 $a = -5, b = 6$

27 **a** 0.4 **b** $\frac{2}{3}$

28 **a** $\frac{2}{9}$ **b** $\frac{5}{9}$ **c** $\frac{8}{9}$

29 **a** 0.8 **b** 0.8 **c** 0.95

30 **a** 0.2019 **b** Approximately 0.25 (= ${}^6C_2$ $(0.2019)^2$ $(0.7981)^4$)

Exercise 3D PAGE 70

1 Mean 4.5, variance $\frac{25}{12}$. **2** Mean 0.75, variance $\frac{3}{80}$. **3** Mean 0.25, variance $\frac{3}{80}$. **4** Mean 2.4, variance $\frac{192}{175}$.

5 **a** $\frac{16}{3}$ **b** $\frac{5\sqrt{2}}{3}$

6 100 metres. **7** Mean 3, variance 0.8, standard deviation $\sqrt{0.8}$.

8 Mean 6.5 (i.e. 4.5 + 2), variance $\frac{25}{12}$ (i.e. no change).

9 Mean 9 (i.e. 4.5 × 2), variance $\frac{25}{3}$ (i.e. $\frac{25}{12} \times 2^2$).

10 **a** $\mathrm{E}(Y) = 36, \mathrm{SD}(Y) = 9$ **b** $\mathrm{E}(Y) = 15, \mathrm{SD}(Y) = 3$ **c** $\mathrm{E}(Y) = 29, \mathrm{SD}(Y) = 6$

11 **a** $\mathrm{E}(Z) = 102, \mathrm{SD}(Z) = 20$ **b** $\mathrm{E}(Z) = 45, \mathrm{SD}(Z) = 8$ **c** $\mathrm{E}(Z) = 64, \mathrm{SD}(Z) = 12$

12 Mean 118.4, variance 51.84, standard deviation 7.2

13 Mean ÷ 100, standard deviation ÷ 100.

In questions **14** to **19** the choice of ≤ or <, and ≥ or >, could differ from that shown here.

14 $$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x \le 0 \\ 0.25x & \text{for } 0 < x \le 4 \\ 1 & \text{for } x > 4 \end{cases}$$

15 $$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x \le 2 \\ 0.25(x-2) & \text{for } 2 < x \le 6 \\ 1 & \text{for } x > 6 \end{cases}$$

16 $$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x \le 0 \\ x^3 & \text{for } 0 < x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$$

17 $$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x \le 1 \\ \ln x & \text{for } 1 < x \le e \\ 1 & \text{for } x > e \end{cases}$$

18 $$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 - e^{-x} & \text{for } x > 0 \end{cases}$$

19 $$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x < 5 \\ 0.5x - 0.02x^2 - 2 & \text{for } 5 < x \le 10 \\ 1 & \text{for } x \ge 10 \end{cases}$$

20 **a** 0.7 **b** 0.3 **c** 0.4 **d** 0.7

21 **a** 0.7364 **b** 0.2636 **c** 0.2835 **d** 0.7165 **e** 0.1184

Miscellaneous exercise three PAGE 73

1 $\frac{\log 6}{\log 3}$

2 **a** 0.6 **b** 0.2 **c** 0.5

3 **a** $q - p$ **b** $2p + q$ **c** $p + 2q$ **d** $3p + 1$ **e** $\frac{q}{p}$ **f** $\frac{p}{q}$

4 **a** $\ln\left(\frac{17}{e+1}\right)$ **b** $\frac{7\ln 50 + 1}{\ln 50 - 2}$

5 **a** ${}^5C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx 0.032\,15$ **b** $\left(\frac{1}{6}\right)^3 \approx 0.004\,63$ **c** 0.003 34 (5 dp) **d** 0.180 87 (5 dp)

6 $\frac{1}{x}$ **7** $3+\frac{1}{x}$ **8** $\frac{2}{x}$ **9** $\frac{6}{x}$ **10** $\frac{1}{2x}$ **11** $-\frac{1}{x}$

12 Approximately 31 years.

13 **a** $(2, 2+\ln 4)$ **b** $(3, \ln 18)$

14 **a** $\frac{200}{1+x}$ dollars per unit **b** \$15.36

15 $y=\sqrt{3}x-\frac{\sqrt{3}\pi}{6}$

16 **a** 0.6321 **b** 0.9502 **c** 0.0498

17 $(0.5, 0.5+\log_e 2)$, minimum. **18** Mean 4, variance 16

Exercise 4A PAGE 78

1 **a** 1 **b** 1.7 **c** –2 **d** 0.5 **e** –0.75

2 Test A: 2.5, Test B: –1, Test C: 1.25, Test D: 0.2

3 Computing (1.216), Chemistry (0.278), Mathematics (–0.385), Electronics (–0.616)

4 English, Mathematics, Science, Social Studies.

5 Jill: '*Well I got 1*'

Jill: '*The mean was zero.*'

Jill: '*Oh he got –0.25 .*'

Exercise 4C PAGE 90

1 0.6915 **2** 0.9088 **3** 0.8849 **4** 0.5793 **5** 1.73 **6** 37.64

7 7.54 **8** 21.25 **9** 0.2266 **10** 0.6377 **11** 54.56

12 **a** 0.5828 **b** –0.6433 **c** 1.2265 **d** –0.7388

13 **a** 19.5 **b** 21.9 **c** 18.7 **d** 23.1

14 **a** 0.68 **b** 0.95 **c** 0.997 **d** 0.95 **e** 0.997 **f** 0.34

g 0.84 **h** 0.16 **i** 0.84 **j** 0.16

15 **a** 99.7% **b** 16% **c** 13.5%

16 **a** 16% **b** 2.5%

17 **a** 0.3085 **b** 0.0062

18 **a** 0.5 **b** 0.34

19 **a** 0.3085 **b** 0.2902 **c** 0.0228

20 **a** approximately 11 **b** approximately 11 **c** approximately 39

21 0.0548

22 **a** 415 **b** 217 **c** 55.5 (nearest half mark) and 68.5 (nearest half mark).

23 To nearest 0.5 cm: 158.5 cm, 191.5 cm

24 A/B: 78, B/C: 68, C/D: 55, D/F: 47

25 **a** 0.842 standard deviations **b** 44.2

26 **a** 0.1587 **b** 7:38 a.m. **c** 7:33 a.m.

27 **a** 2 years **b** 7 years **c** 91 years **d** 0.783

28 **a** approximately 40 **b** 0.236

29 **a** 0.036 **b** 0.242 **c** 0.31

30 **a** 0.2054 **b** 0.0066

31 **a** 505 **b** 255

32 **a** approximately 2.3% **b** 4.6

Exercise 4D PAGE 96

1 For normally distributed data we would expect approximately 68% of the data points to be within one standard deviation of the mean. The fact that 80% meet this criteria would indicate that it could be unwise to model the data set as being normally distributed.

2 For $X \sim N(4.37, 2.52^2)$, $P(X > 8) = 0.0749$, i.e. approximately 7.5% of the measurements are above 8. The data set mentioned in this question has more than 14% of the measurements greater than 8 thus indicating that to assume the data set to be normally distributed could be unwise.

3 For $X \sim N(8.9, 3.1^2)$, $P(5.8 < X < 12.0) = 0.6827$, i.e. approximately 68% of the measurements are within one standard deviation of the mean. We have just 31% within one standard deviation thus indicating that to assume the data set to be normally distributed could be unwise.

4 The given information shows that the data distribution is not symmetrical. For a normal distribution we would expect about 34% of measurements between the mean and one standard deviation below the mean and 34% between the mean and one standard deviation above the mean. We have 26% and 44% respectively. Thus to assume the data set to be normally distributed could be unwise.

5 The given information shows that the measurements are not symmetrically distributed.

Also, for $X \sim N(\bar{x}, \sigma^2)$, $P(X > \bar{x} + 2\sigma) = 0.022\,75$, i.e. approximately 4 measurements would be more than two standard deviations above the mean and approximately 4 would be more than two standard deviations below the mean. We have ten above and none below. Thus to assume the data set to be normally distributed could be unwise.

6 **a** Yes

b Depends what 'sufficiently' reliable means. Without knowing what the probabilities are to be used for we cannot judge if they are likely to be 'sufficiently' reliable. With 520 measurements and the scientist's 3% rule the approximation to normal is likely to be good but her rule does not allow for any unusual grouping within each interval that might make the approximation questionable. (Having the 520 measurements she could carry out further checks for this aspect.)

7 If the data is normally distributed, from the symmetry of such a distribution, we should have the mean close to the median, which is the case for this data, and the lower quartile should be approximately as far below the mean as the upper quartile is above it.

Again this is the case for this data, $69.2 - 44.7\ (= 24.5) \approx 44.7 - 19.3\ (= 25.4)$.

For $X \sim N(0, 1^2)$ solving $P(X \le x) = 0.25$ gives $x = -0.67$

Hence, in a normal distribution the lower quartile is 0.67 standard deviations below the mean. For our data the lower quartile is 0.96 standard deviations below the mean.

Similarly, in a normal distribution the upper quartile is 0.67 standard deviations above the mean. For our data the upper quartile is 0.93 standard deviations above the mean.

The figures suggest that the data is not normally distributed.

8 **a** By binomial 0.8697
By normal 0.8692

b By binomial 0.0563
By normal 0.0564

c By binomial 0.0944
By normal, i.e. $P(20.5 < X < 25.5)$ with $X \sim N(30, 12)$, 0.0939

Miscellaneous exercise four PAGE 98

1 $\frac{1}{x}$ **2** $\frac{10}{x}$ **3** $\frac{-1 + \ln x}{(\ln x)^2}$ **4** $\frac{6x}{x^2 + 1}$ **5** $\frac{2(x+2)}{(x+1)(x-1)}$ **6** $\frac{1}{x \ln 5}$

7 0.0668 **8** 0.6827 **9** 1.32 **10** 18.25 **11** $\frac{3}{e}$ **12** 2

13 0.0038

14 **a** -0.202 **b** -1.126 **c** 0.332 **d** -0.228

15 **a** Approximately 17 days **b** Approximately 237 days **c** Approximately 197 days.

16 $a = 0.3, k = 0.1, \mathrm{E}(X) = 1.4, \mathrm{VAR}(X) = \frac{17}{75}$.

17 **a** 0.5 **b** $\frac{\sqrt{2}}{2}$

18 **a** $\mathrm{P}(x) \approx 1000x - 25\,000 + 20\,000\log_e\left(1 - \frac{x}{100}\right), x < 100.$

b Extract 80 kg per 5 tonne batch for a maximum profit of approximately \$22 800.

19 **a** $\frac{1}{6}$ **b** $\frac{5}{6}$ **c** $\frac{11}{18}$ **d** 0

20 **a** 0.0304 **b** 0.116

Exercise 5A PAGE 109

1 **a** Likely to introduce bias. **b** Likely to introduce bias. **c** Not likely to introduce bias.

d Not likely to introduce bias. (Not that is likely to influence car colour anyway.)

e Not likely to introduce bias. (Not that is likely to influence height anyway, unless perhaps the school has something like a special basketball intake.)

2 Two under twenty, three in their twenties, four in their thirties and one of 40 or over.

4 19 year 8s, 19 year 9s, 18 year 10s, 12 year 11s and 12 year 12s.

5 Approximately 190. **6** Approximately 760. **7** Approximately 3200.

8 Either by considering the symmetry of the situation, or from the calculation shown below, the long term mean, or expected value, for rolling a fair normal die is 3.5.

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

With one mean, 4.08 (= 3.5 + 0.58), noticeably further from 3.5 than the other (3.42 = 3.5 − 0.08) we would expect the one closer to the expected value to be the one involving the 150 rolls. This suggests that Christine, with her 150 rolls of the die, would have the mean of 3.42 and Shane, with his 12 rolls of the die, would have the mean of 4.08.

9 Either by considering the symmetry of the situation, or from the calculation shown below, the long term mean, or expected value, for the sum of the two numbers obtained by rolling two normal fair dice is 7.

$$2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7.$$

With one mean, 6.42 (= 7 − 0.58) noticeably further from 7 than the other (7.17 = 7 + 0.17) we would expect the one closer to the expected value to be the one involving the 150 rolls. This suggests that Horace, with his 150 rolls of the two dice, would have the mean of 7.17 and Portia, with her 12 rolls of the two dice, would have the mean of 6.42

10 Approximately 6500 people did not complete a census form.

11 Ask someone to read and comment about your explanation and you read and comment about their explanation.

Have they explained why the process works or have they just listed the steps involved?

Have they considered possible sources of errors, for example:

Will the tagged turtles remix with the rest of the population?

Will ‘once caught turtles’ be more prone to recapture?

Will the initial capture involve a random selection from the lake? What if just one region of the lake is used for capture and recapture? etc.

Counting seals PAGE 111

Explain how we could 'randomly select' the squares to be photographed and suggest how many squares should be selected.

Number the squares and use a random process to select at least 30, let us say 50, squares from the ones entirely covering land, as per the plan. This could be done using a random number generator set to generate integers from 1 to 900 and record the numbers, ignoring any repeated numbers and any that represented squares that were not 'all land' until 50 different 'all land' squares were chosen.

Will your random selection guarantee that the sample is an accurate representation of the population of seals on the island at the time the photographs were taken? Explain.

We cannot guarantee that the sample is an accurate representation of the population but if the land only squares are representative of the whole island then the number of squares in our sample allows us to be reasonably confident that the results from the sample can be used to give a reasonable population estimate.

How could the 'seal counts' from the selected photographs be used to estimate the seal population on the island at the time the photographs were taken?

Using our random squares to determine an average number of seals per 10 000 m^2 we multiply this by 734 to estimate the number on the 7 340 000 m^2 (= 7.34 km^2) island.

Suggest any improvements that could be made to the plan?

The squares that contain some water and some land will be near the ocean and, given the somewhat laboured nature of a seal's movement on land, many seals may choose to stay near the water's edge. Hence these squares, with their proximity to the water, could well be where many seals choose to rest, choosing not to struggle further inland. Thus whilst it is perhaps wise to treat these 'part water part land' squares differently, dismissing them from the sample altogether may not be the best option. Better to sample these water's edge squares too and then include the data from them in some proportional way.

Miscellaneous exercise five PAGE 123

1 5

2 $\log_e\left(\dfrac{P}{9}\right) - 1$ **a** 1.996 **b** 4.991 **c** 2

3 **a** $p+q$ **b** $p-q$ **c** $2p+3q$ **d** $0.5p$ **e** $\dfrac{p}{\log e}$ **f** $\dfrac{2q}{1-\log 2}$

4 $n = 150$, $p = 0.4$, $\text{P}(X \le 50) = 0.056$ (correct to 3 dp).

5 $1 + \ln(5x)$ **6** $\dfrac{2\log_e x}{x}$ **7** $x + 2x\ln x$ **8** $\dfrac{2(3+\ln x)}{x}$ **9** $\dfrac{2(x-1)}{x^2}$ **10** $-\dfrac{1}{x(\ln x)^2}$

11 $f''(x) = 5x + 6x\ln x$

12 **a** 1.25 **b** 2

13 **a** $k = 0.25$ **b** $\dfrac{11}{6}$ **c** $\dfrac{11}{36}$ **d** $\dfrac{\sqrt{11}}{6}$

e $$\text{P}(X \le x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ -\dfrac{1}{8}x^2 + x - \dfrac{7}{8} & \text{for} \quad 1 \le x \le 3 \\ 1 & \text{for} \quad x > 3. \end{cases}$$

(Placement of the 'equals part of the inequality' could vary from that shown here.)

Or, without actually performing the integration, this could be written:

$$\text{P}(X \le x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ \int_1^x 0.25(4-t)\,dt & \text{for} \quad 1 \le x \le 3 \\ 1 & \text{for} \quad x > 3. \end{cases}$$

14 First recapture suggests 3768 birds of this species in the area.

Second recapture suggests 3143 birds of this species in the area.

Estimated number of birds of this species in the area: Approximately 3500.

A possible problem with using capture-recapture techniques on migratory birds would occur if the swampy area was just a short stay location during the migration. If this was the case some of the birds tagged in the first batch may have moved on by the time of the recapture. The proportion of tagged ones in the second, and subsequent captures may not then reflect the proportion tagged in the whole population. However if it were known that almost all of the birds stayed in the swampy area for a reasonable amount of time, and that the capturing and recapturing could all be carried out during this time, this problem would be avoided.

15 Only the last four parts, **i**, **j**, **k** and **l**, are true for all $p > 0$ and $q > 0$.

16 0.6205

17 **a** 45.9 **b** 59.2 **c** 48.2 **d** 43.3

18 **a** 28 is 0.385 standard deviations from the mean (below).

b The mean is 30.21 (2 dp).

Exercise 6A PAGE 142

1 First survey has $\hat{p} = \frac{49}{123} \approx 0.398$ Second survey has $\hat{p} = \frac{761}{2348} \approx 0.324$

The second value for $\hat{p}$ is likely to be the better estimate of p, the population proportion, due to it involving a larger sample.

2 Assuming the samples are reasonably representative of the shoppers using this supermarket (which might not be the case given all of the samples were on just one day), and with each sample involving the same number of people, we can estimate the population proportion by finding the mean of the sample proportions. This gives an estimate of 0.72.

3 **a** By adding the numbers of samples $(1 + 4 + 6 + 7 + 2 + 7 + 2 + 2 + 4)$ we obtain the total number of samples as 35.

b By finding the mean of the sample proportions we obtain an estimate of the population proportion of 0.799.

4 **a** 0.225 **b** 0.25 **c** Mean of $\hat{p}$ is 0.25, standard deviation is $\sqrt{\frac{0.25(1-0.25)}{320}} \approx 0.024$.

5 **a** $\frac{13}{18}$ **b** 0.67 **c** Mean of $\hat{p}$ is $\frac{13}{18}$, standard deviation is $\sqrt{\frac{\frac{13}{18}(1-\frac{13}{18})}{100}} \approx 0.0448$

d Our value for $\hat{p}$ is approximately 1.166 standard deviations below the mean value, p.

6 **a** The population proportion is 0.84 (or 84%).

b The sample proportion is 0.6125 (or 61.25%).

c With $p = 0.84$ and $n = 240$ we would expect $\hat{p}$ to be normally distributed with mean 0.84 and standard deviation 0.024. A $\hat{p}$ value of 0.6125 is more than 9 standard deviations below the mean! Such an extraordinary result suggests that the sample was not truly representative of the population as a whole with regard to household internet access.

7 **a** For $p = 0.1$ and a sample size of 1000 we would expect the sample proportions of left handers to be normally distributed with mean 0.1 and standard deviation 0.0095 $(= \sqrt{\frac{0.1 \times 0.9}{1000}})$. Our sample proportion of 0.112 is just 1.26 standard deviations above the mean. Thus, whilst it is not clear whether the classification of the school students as being 'left handed' was the same as the 'ranging from moderate through strongly left handed' classification the paper mentions, the proportion of left handers in the sample seems consistent with what we might expect for a sample size of 1000 and population proportion 0.1.

b Not knowing the number of participants in the 1981 World Championship foil competition means that we do not know the sample size. However, even not knowing the sample size, the proportion of 0.35 is so much higher than the population proportion of 0.1. The participants in the 1981 World Championship foil competition do not form a representative sample of the left handedness in the general population.

8 **a** The sample proportion is $\frac{461}{1247} \approx 0.37$

b An estimate of the standard deviation of the sample proportions is $\sqrt{\dfrac{\frac{461}{1247}\,(1 - \frac{461}{1247})}{1247}} \approx 0.0137.$

9 **a** The sample proportion is $\frac{143}{248} \approx 0.577.$

b An estimate of the standard deviation of the sample proportions is $\sqrt{\dfrac{\frac{143}{248}\,(1 - \frac{143}{248})}{248}} \approx 0.0314.$

10 **a** Simply finding the mean of the sample proportions takes no account of the fact that the proportions for larger samples should give a better estimate than those with smaller samples. We need to attach more importance to the sample proportions coming from the larger samples.

b If we use the given information to determine the number in each sample with high blood pressure, then express the total number with high blood pressure as a proportion of the total number surveyed, this would give a better estimate of the population proportion.

A better estimate for the population proportion would be $\frac{193}{756} \approx 0.2553.$

(Note: Simply averaging the sample proportions gives 0.3275.)

Sample	Number in sample	Sample proportion having high blood pressure	Number in sample having high blood pressure
1	8	0.5	4
2	10	0.1	1
3	50	0.28	14
4	25	0.24	6
5	10	0.2	2
6	80	0.2375	19
7	56	0.286	16
8	10	0.1	1
9	50	0.2	10
10	180	0.261	47
11	20	0.35	7
12	10	0.1	1
13	8	0.375	3
14	25	0.2	5
15	8	0.5	4
16	150	0.2	30
17	20	0.3	6
18	25	0.32	8
19	10	0.8	8
20	1	1	1
Total	756		193

11 For $n = 200$ and $p = 0.\,1$ we would expect the sample proportions to be well modelled by a normal distribution with mean 0.1 and standard deviation $\sqrt{\dfrac{0.1(1 - 0.1)}{200}} \approx 0.021\,21.$

With 35 seeds (or more) from a sample of 200 unable to germinate the sample proportion is 0.175 (or more). This is more than 3.5 standard deviations above the mean of the distribution, $0.175 \approx 0.1 + 3.54 \times 0.02\,12$, and is therefore extremely unlikely.

12 We expect the distribution of sample proportions to approximate to a normal distribution if $np \geq 10$ and $n(1-p) \geq 10$.

This is the case with Graph One with

$$\begin{aligned} np &= 50 \times 0.5 \\ &= 25 \end{aligned} \quad \text{and} \quad \begin{aligned} n(1-p) &= 50 \times 0.5 \\ &= 25 \end{aligned}$$

However, with Graph Two,

$$\begin{aligned} np &= 50 \times 0.05 \\ &= 2.5 \end{aligned}$$

Hence it should not be a surprise that Graph One better approximates the shape of the normal distribution.

13 0.45 and 0.55.

14 Even if the politician's claim is correct, in a sample of 200 people we would not expect to necessarily find that the proportion of people voting for the politician would exactly match the 52% claimed but we would expect the sample percentage to be quite close. Indeed we would expect the sample proportion to come from a normally distributed random variable of mean 0.52 and standard deviation 0.035. For such a distribution the sample proportion of 81 out of 200 (= 0.405) is approximately 3.3 standard deviations below the mean. This would be very unusual. Hence, if the sample of 200 people fairly represented the voting behaviour of the people who intended to vote in the election for the seat of Dasha we would have to question the politician's claim that he would get 52% of the vote.

15 There is a 90% chance that in a sample of 800 cars produced by this company the sample proportion that are blue will be between 21.5% and 26.5%.

Exercise 6B PAGE 154

Note
- The accuracy stated here is to allow you to check your answers. In practice values could well be rounded more heavily, dependent upon the situation and what the statistics are to be used for. For example the 90% confidence interval in question 2 may well be quoted as 0.83 to 0.87, i.e. 0.85 ± 0.02, and the interpretation could well refer to 83% and 87%.
- Confidence intervals given here have been rounded using the usual regime for rounding, or according to the level of rounding stipulated in the question. However, if in a real situation it were crucial that when stating a 95% confidence interval we were not claiming to be greater than 95% confident, or perhaps not claiming to be less than 95% confident, the rounding regime would have to be more carefully applied.

1 The 95% confidence interval is 0.3476 to 0.4024. (I.e. 0.375 ± 0.0274).

Were we to repeat such sampling we could expect 95% of the 95% confidence intervals so formed to contain the population proportion. Hence, with 95% confidence we estimate that between 34.76% and 40.24% of the people living in Australia are in favour of the idea of introducing compulsory national service.

2 The 90% confidence interval is 0.8292 to 0.8708. (I.e. 0.85 ± 0.0208).

Were we to repeat such sampling we could expect 90% of the 90% confidence intervals so formed to contain the population proportion. Hence, with 90% confidence we estimate that between 82.92% and 87.08% of the people living in Australia who had recently contacted their bank for online help were either satisfied or very satisfied with the service they received.

3 The 99% confidence interval is 0.6904 to 0.8296. (I.e. 0.76 ± 0.0696).

Were we to repeat such sampling we could expect 99% of the 99% confidence intervals so formed to contain the population proportion. Hence, with 99% confidence we estimate that between 69.04% and 82.96% of the people regularly playing the particular sport agreed that the recent rule changes were a good idea.

4 70% ± 3%.

5 43.2% to 46.8%.

Were we to repeat such sampling we could expect 90% of the 90% confidence intervals so formed to contain the population proportion. Hence, with 90% confidence we estimate that between 43.2% and 46.8% of the people of the particular nation involved wanted to see changes to the current daylight saving rules.

For the particular community the sample proportion is 70% which is a long way outside of the 90% confidence interval. If the original sample was fairly representative of the population as a whole the 70% figure would suggest that the particular community returning the 70% proportion was not typical of the national opinion. Perhaps local considerations made this community much more inclined to want to see changes to the daylight saving rules.

6 Discuss and compare your statement with the statements made by others in your class.

7 0.0218

8 The 99% confidence interval.

9 **a** The sample proportion of acceptable components is 0.82

b An example of the sort of statement that could be made:

With 90% confidence we estimate that around the time the sample was taken, between 77.5% and 86.5% (i.e. 82% ± 4.5%) of the components made by this machine were of an acceptable standard.

c An example of the sort of statement that could be made:

With 99% confidence we estimate that around the time the sample was taken between 75% and 89% (i.e. 82% ± 7%) of the components made by this machine were of an acceptable standard.

10 Rounding up to the next integer gives a sample size of 228.

11 Rounding up to the next integer gives a sample size of 752.

12 The sample size should be 350 or greater.

13 We can be 95% confident that of all Australian males between the ages of 20 and 30, between 72% and 80% are taller than their father.

To be 99% confident our interval would need to be larger.

To be more confident that p will lie in our interval the interval needs to be larger.

14 We can be 90% confident that the proportion of Australians having the particular attribute lies between 18% and 30%.

The sample size was 137 of whom 33 possessed the particular attribute.

15 **a** 0.35, or 35%.

b 0.0218

c 0.3073 to 0.3927, i.e. 0.35 ± 0.0427

Were we to repeat such sampling we could expect 95% of the 95% confidence intervals so formed to contain the population proportion. Hence we can be 95% confident that between 30.7% and 39.3% of year 12 Australian school students would say they intend to proceed to University the following year.

d 972 or greater.

Miscellaneous exercise six PAGE 156

1 **a** $x = 2$ **b** $x = 8$ **c** $x = 100$ **d** $x = 1$

2 **a** $\frac{3\sqrt{x}}{2}$ **b** $20x^4 + \frac{1}{x}$ **c** $\frac{7}{x}$ **d** $\frac{15x^2 - 6}{5x^3 - 6x}$

3 **a** $\frac{5}{5x - 1}$ **b** $\frac{4x^3}{x^4 + 1}$ **c** $\frac{2x}{x^2 - 1}$

4 $y = 3 - \frac{x}{e}$

5 **a** 0.25 **b** 0.2875 **c** Mean of $\hat{p}$ is 0.25, standard deviation is $\sqrt{\frac{0.25(1-0.25)}{160}} \approx 0.0342$.

d Our value for $\hat{p}$ is approximately 1.1 standard deviations above p.

6 **a** 0.125 **b** 0.75 **c** 0.75 **d** 0.8

7 0.9179 **8** $f''(x) = -\frac{5}{x^2}$, $f(x) = x + 5\ln x + 4$

9 **a** $\ln 5$ **b** $\frac{\ln 2}{\ln 5}$

10 30

11 **a** 0.003 **b** 0.006

12 **a** 0.76 **b** $\mu = 33.9, \sigma = 7.2$ **c** 0.95

13 Compare your answer with those of others in your class.

14 63.5%

15 Discuss your results with those of others in your class.

16 Were we to repeat such sampling we could expect 95% of the 95% confidence intervals so formed to contain the population proportion. Hence we can be 95% confident that between 17% and 21% of adult Australians were, at the time of the survey, 'mobile only'.

The proportion for the Australians aged 65 or over, 0.04 or 4%, and for Australians in their twenties, 0.42 or 42%, are both a long way outside the 17% to 21% interval that we would expect the population proportion to lie, and would therefore expect other sample proportions to be quite close to. The figures suggest that samples considering only Australians over 65 or only Australians in their twenties would not be representative of the population as a whole, as we would probably expect when considering this 'mobile only' category. Compared to the populations as a whole, those in their twenties are more likely to be 'mobile only' and those over 65 are less likely to be 'mobile only'. For this 'mobile only' attribute, the '65 and over' sample and the 'in their twenties' sample are not representative of the population as a whole.

17 With a population proportion of 0.22 and sample size of 400 we would expect the sample proportions to be normally distributed with mean 0.22 and standard deviation of approximately 0.0207. Thus 20%, or 0.2 is approximately one standard deviation from the mean. The change could be explained as being reasonable variation in sample proportions.

If the survey had involved 4000 people we would now expect the sample proportion to be from a normal distribution with mean 0.22 and standard deviation 0.006 55. The sample proportion of 20%, or 0.2 is then just over 3 standard deviations from the mean – very unlikely to occur just on the basis of random variations and now much more likely to indicate a loss of popularity.

18 For a sample size of 100 and sample proportion of faulty batteries of 0.15 we would expect the sample proportions to be approximately normally distributed with a mean equal to the population proportion and a standard deviation of $\sqrt{\frac{0.15 \times 0.85}{100}}$, i.e. approximately 0.036. If the population proportion is as claimed, i.e. 0.15, then 0.17 is only 0.56 of a standard deviation above this, certainly possible for the stated population proportion. Thus the fact that a sample of 100 had 17 faulty batteries does not mean the claim that the population proportion of faulty batteries is 0.15 is necessarily false.

However there is the issue that the advertising might suggest to the customer that when they buy 100, as Joe did, they will be getting 85 that do work. Perhaps it should be suggested that the company tests the batteries and only sells working ones, or maybe that the advertising wording be changed so that customers realise that they could end up with less than (or maybe more than) 85% that work.

ISBN 9780170395205

INDEX

ISBN 9780170395205

ISBN 9780170395205

ISBN 9780170395205

Western Australia's leading senior mathematics series for over 25 years has had a makeover, introducing a new colourful page design.

Written by **Alan Sadler**, these textbooks cover the three new senior mathematics ATAR courses for WA and the Australian Curriculum.

This new series caters to students of Mathematics Applications, **Mathematics Methods** and Mathematics Specialist across Units 1 to 4.

Mathematics Methods Unit 1
ISBN: 9780170390330

Mathematics Methods Unit 2
ISBN: 9780170390408

Mathematics Methods Unit 3
ISBN: 9780170395137

This new format contains the same text as in the original format, so that both formats can be used in the classroom concurrently.

nelsonnet
www.nelsonnet.com.au

Features **Preliminary Work** at the start of the book and **Miscellaneous Exercises** at the end of each chapter.

Includes the **NelsonNetBook** version of this book with hotlinks to additional worksheets, plus student and teacher websites *(access conditions apply).*

For more information contact the relevant state Education Consultant at **www.nelsonsecondary.com.au** or visit **www.nelsonnet.com.au**

ALSO AVAILABLE

Mathematics Applications Unit 1 *Student Book*
ISBN: 9780170390194

Mathematics Applications Unit 2 *Student Book*
ISBN: 9780170390262

Mathematics Applications Unit 3 *Student Book*
ISBN: 9780170394994

Mathematics Applications Unit 4 *Student Book*
ISBN: 9780170395069

Mathematics Specialist Units 1 & 2 *Student Book*
ISBN: 9780170390477

Mathematics Specialist Units 3 & 4 *Student Book*
ISBN: 9780170395274

NELSON
CENGAGE Learning
For learning solutions, visit **cengage.com.au**

ISBN: 978-0170395205
9 780170 395205