

1

How to make mathematical statements (Numbers, equations and functions)

'In science there is only physics and stamp collecting'
Ernest Rutherford (1871–1937), the father of nuclear physics.

'I have hardly ever known a mathematician who was capable of reasoning'
Plato (428–348 BC), the father of all science.

One of the exciting challenges of quantitative ecology is to examine whether a set of observations that have been classified by name can be **ordered** along a continuum. Therefore, this chapter begins with a discussion of nominal and ordinal scales (Section 1.1). Although there is still a valuable role for **nominal classification** (see Chapter 12), the deceptively simple act of comparing two, apparently different, individuals, species or communities along one or more quantitative scales, propels us forward from natural history to modern ecology. This transition is mediated by **numbers** (Sections 1.2 and 1.17). **Symbols** (Section 1.3) are often used instead of numbers either to cope with ignorance or to make general statements. Mathematical **operators** (Sections 1.4 and 1.5) are used to connect different (known or unknown) quantities into algebraic **expressions**. **Algebra** is the set of rules dictating how these expressions may be manipulated (Sections 1.7–1.9). The two main scientific applications of mathematics are in formalising known facts or assertions as **equations** or **inequalities** (Sections 1.10–1.15) and expressing **relationships** between variables (Sections 1.18–1.25).

1.1. Qualitative and quantitative scales

Data are called **qualitative** if they cannot be compared using some measure of magnitude. For example, **nominal** observations can only be compared in a rudimentary way, by checking for 'sameness'. If they are not the same, one nominal observation cannot readily be said to be greater than another. In contrast, **quantitative** data can be ordered and the degree of dissimilarity between them can be evaluated objectively. This rudimentary taxonomy of data will be elaborated in Chapter 7. For now, it is sufficient to say that the distinction between quality and quantity is not always clear. Often, observations that appear to be nominal can be ordered by means of their attributes, as in Example 1.1.

Example 1.1: Habitat classifications



Fern frond

We can easily distinguish between marine and terrestrial habitats. In the marine environment there are polar, upwelling, shelf, open-ocean and coral habitat types. In the terrestrial environment, examples include the boreal, tundra, tropical, temperate, desert and montane habitat types. The definitions of these are generally vague but suffice for most applied purposes. However, studies in spatial ecology (Manly *et al.* 2002; Aarts *et al.* 2008) have increasingly found that it is more useful to describe the distribution of plants and animals in terms of individual habitat characteristics such as temperature and precipitation (measured on a quantitative scale) rather than using arbitrary – and occasionally anthropocentric – habitat types (Figure 1.1).

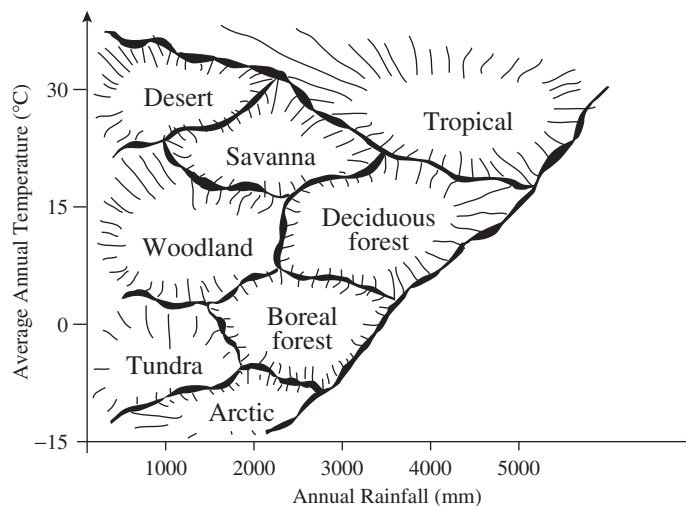


Figure 1.1: Habitat types are arbitrary subdivisions imposed on an environmental continuum.

1.1: Declaring nominal categories

To create a simple computerised taxonomic scheme involving the categories of Animals, Plants, Fungi, Protocista, Archaea and Monera, it is first necessary to tell R that these labels are to be treated as text, so that it doesn't expect a numeric value for them. This is done by enclosing the labels in quotation marks:

```
"An", "Pl", "Fn", "Pr", "Ar", "Mo"
```

The labels can be collected together using the **concatenation** command `c()`:

```
c("An", "Pl", "Fn", "Pr", "Ar", "Mo")
```

and the taxonomy is declared using the command `factor()` which says to R that a collection of specimens can be classified according to this scheme of labels (more on factors in Chapter 7):

```
factor(c("An", "Pl", "Fn", "Pr", "Ar", "Mo"))
```

so, to classify a collection of organisms according to kingdom, each specimen needs to be associated with one of the six categories in this factor.

1.2. Numbers

Numbers are certainly useful for counting, but not all measurable quantities can be counted. Thankfully, the different types of numbers used for measurement are both countable and few; all-in-all there are only five. Each type is a **set**, an imaginary container that may enclose (or be itself enclosed in) other sets (Figure 1.2).

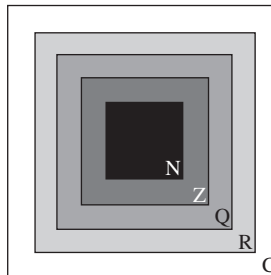


Figure 1.2: There are five types of numbers, usually represented as a hierarchy of nested sets.

The first set of numbers, both historically and in order of simplicity, are the **naturals** (collectively denoted by \mathbb{N}). These are the numbers 1, 2, 3, 4, etc., that you would use to count whole items, such as the number of animals in a population or the number of species in a community. If we use curly brackets to enclose the elements of a set, then we have

$$N = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots\} \quad (1.1)$$

The three dots at the end of the sequence imply an infinite continuation of the pattern already expressed by the preceding numbers.

The second set of numbers are the **integers**, collectively denoted by \mathbb{Z} . They are also known as the **signed numbers** because they are preceded by a minus or a plus

$$\mathbb{Z} = \{\dots, -4, -3, -2, -1, 0, +1, +2, +3, +4, \dots\} \quad (1.2)$$

Zero represents the absence of any magnitude and the plus signs of the positive numbers are usually implied,

$$\mathbb{Z} = \{\dots, -4, -3, -2, -1, 0, \underbrace{1, 2, 3, 4, \dots}_{\mathbb{N}}\} \quad (1.3)$$

Compare (1.1) with (1.3) and note that the set of naturals is a **subset** of the integers (i.e. \mathbb{N} is contained in \mathbb{Z}). In mathematical notation, this is written $\mathbb{N} \subset \mathbb{Z}$.

The third set of numbers are the **rationals**, denoted by \mathbb{Q} . They are the numbers produced from the **ratio** or division of any two integers n, m , assuming m is not zero. In mathematical notation:

$$\mathbb{Q} = \left\{ \frac{n}{m} \forall n, m \in \mathbb{Z}, m \neq 0 \right\} \quad (1.4)$$

Try not to panic when you see an expression like this. Mathematical notation is admittedly unfriendly but it makes up for it by being both precise and brief. Often, even the most intimidating expressions have a plain-language translation. In Equation (1.4) the symbols \forall and \in are mathematical shorthand, meaning 'for every' and 'belonging to', respectively. So, the whole expression says: The rationals are the numbers that can be obtained by dividing two integers n over m , excluding the value zero for the denominator m .

Example 1.2

$$6/3 = 2, \quad 1/2 = 0.5, \quad 7/4 = 1.75 \text{ and } -10/2 = -5 \text{ are rational numbers}$$

All integers can be produced as the ratio of other integers so that all integers are also rationals. However, not all noninteger numbers can be produced as ratios of integers. This surplus set of numbers are, quite appropriately, termed the **irrationals**. We will encounter examples later on (e.g. square root of 2) but, for now, it is useful to note that irrational numbers have an infinity of nonrepeating decimals. The combined set of rationals and irrationals gives us the fourth type of numbers called **real numbers**. The set of reals (denoted by \mathbb{R}) is used when we need to measure continuous quantities, such as length, density or mass.

The fifth and final type are known as the **complex numbers**, denoted by \mathbb{C} . A more detailed presentation of complex numbers is left until Section 1.17, but it is worth noting here that the set of complex numbers is a **superset** of the reals. So, we can represent Figure 1.2 in mathematical terms by:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \quad (1.5)$$

Example 1.3: Observations of spatial abundance

There is a correspondence between the different types of ecological measurement and the sets of numbers in Equation (1.5). Consider the measurements that might typically be used to describe the distribution of a plant population along a linear study site, such as a stretch of river, which is 1 km long and has been subdivided into ten segments (Figure 1.3). The easiest description is in terms of **occupancy**, the presence or absence of the species from

any particular segment. Although occupancy can be thought of as a qualitative trait, it is readily made quantitative by attributing the value 0 to absence and the value 1 to presence (Figure 1.3(a)). If, in addition, there are data on the number of occurrences in each segment, then the plant distribution can be described by a series of counts which take their values from the set of non-negative integers, $\{0, \mathbb{N}\}$ (Figure 1.3(b)). These count data can be readily converted to densities by dividing each count by the size of the segment – in this case 100 m (Figure 1.3(c)). Standardised density (or relative abundance) can be obtained by dividing the count in each segment by the total number of observations. This conveys the proportion of the total count occurring in any given segment (Figure 1.3(d)). Both density and relative abundance are rational numbers. Finally, we may want to compare the distribution of the species with that of some environmental covariate that could be used as a proxy at other unsurveyed sites (a covariate is a quantity that is closely related to the measure of interest). For example, there may be a gradient in soil pH along the study site (pH measurements are real numbers). A look at Figure 1.3(e) suggests that the plant has a preference for soil pH around 6.

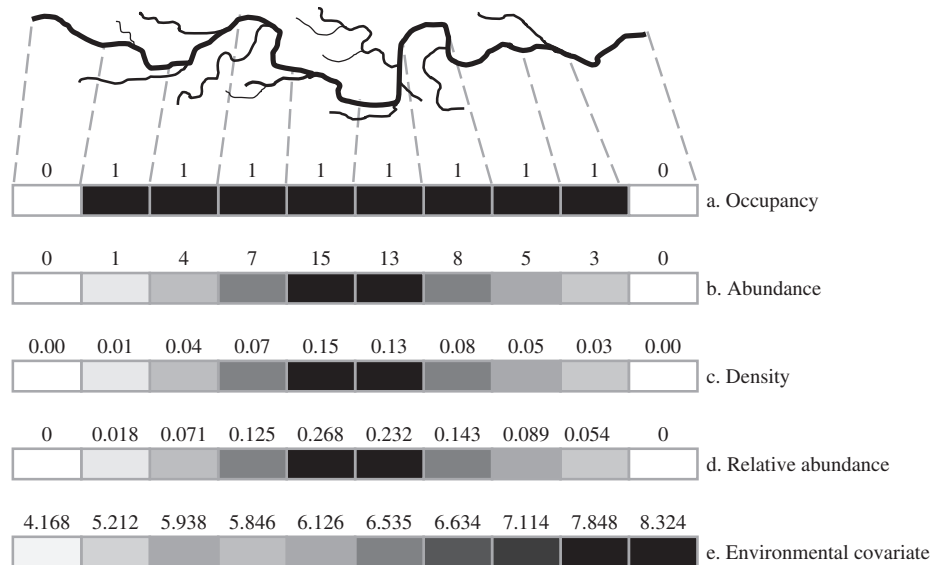


Figure 1.3: Measurements belonging to different sets of numbers naturally occur in ecology. In quantifying the spatial distribution of a species, we may use (a) occupancy (nominal data), (b) counts of abundance (non-negative integers), (c) density, (d) relative abundance (both rational numbers) or (e) environmental covariates, such as soil pH (real numbers).

1.2: Declaring simple sets of numbers

As we saw in R.1.1, a set is declared by the concatenation command `c()`. The set of the first nine natural numbers can be declared as `c(1, 2, 3, 4, 5, 6, 7, 8, 9)`. A quicker alternative is to specify these as a range using a colon `c(1:9)`. These two types of declaration can be used in combination. Here are three different ways of declaring the set of the first ten non-negative integers:

```
c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9) or c(0, 1:9) or c(0:9)
```

1.3. Symbols

A time-honoured mathematical trick is to use symbols to represent unknown quantities. This is done for two reasons. The first is practical: an unknown quantity of interest needs a name (say, x , y or z) while its value is being deduced. As we will see in Sections 1.9–1.13, this kind of speculation involves formulating mathematical statements (equations or functions) that combine known facts with unknown quantities. One might then try to manipulate (i.e. solve or plot) these statements to find out more about the unknown quantities. The second reason is more important: one of the primary endeavours of all scientists is to establish **generality** from their results. In most cases, you won't want to limit your mathematical arguments to specific numbers and symbols facilitate this task. You can fix some of the properties of a symbol and let the others loose. You can, for example, stipulate that x is any integer number. Alternatively, you may specify a narrower range for it, or leave it completely unspecified. Hence, when biological first principles dictate limitations for the values of symbols (e.g. biological populations cannot be negative), *you can avoid investigating biologically meaningless scenarios*.

You can treat a symbol just like a number and operate on it using the rules of algebra (Section 1.8). However, remember that once you've defined your symbol (say, as an integer) you can't treat it as anything that defies that definition (e.g., by talking about its fifth decimal). Of course, the end result of a calculation using symbols may also be partly symbolic (as a result of partial ignorance or the desire for generality).

When forming a mathematical statement, it is important to distinguish between **variables** (quantities that change over time, space or in response to other inputs) and **constants** (quantities that don't change). This distinction is purely pragmatic, and may vary between different systems or applications.

Example 1.4: Population size and carrying capacity



Figure 1.4: Highly-packed colonial breeders. If available space is the only factor limiting the density of these birds, then we can interpret peak densities as a colony's carrying capacity (© Sheilla Russell).

In population models, the size of a population is represented by a variable, say P . If the population size is restricted by limited resources, then the environment can only sustain a limited number of individuals (Figure 1.4). This is called the environment's **carrying capacity** (K). If the environment does not change, carrying capacity should remain constant. So, although P and K are both symbols representing population size, the first is a variable and the second a constant. More complicated models may acknowledge the fact that carrying capacity is subject to environmental change, hence treating it as a variable.

1.3: Naming conventions

When introducing a new symbol in an R session, make sure that you will understand the meaning of the symbol when you look again at your code at a later date. This can be done by using entire words rather than single letters, which also ensures an infinite supply of names, certainly more than the combined Latin and Greek alphabets used in maths notation. However, bear in mind that you will be typing some symbols many times, so try to be sensible with the length of names.

1.4. Logical operations

Logical operators can be used to express equality ($=$), inequality (\neq), directed inequality ($>$, $<$) or inequality with the possibility of equality (\geq , \leq). These enable us to make comparisons, or to make assertions about numbers and symbols.

Example 1.5

The following are true: $1 = 1$, $2 < 3$, $4 \neq -4$

The following are false: $1 \neq 1$, $2 > 3$, $4 = -4$

The following are assertions: $x = 1$, $x \geq y$

1.4: Assigning values to symbols

The general operator for making assignments in R is the arrow \rightarrow or \leftarrow . The value 2 can be assigned to the symbol x , in any one of two ways: $x \leftarrow 2$ or $2 \rightarrow x$. If both sides of the assignment are symbolic, then the information flows according to the arrow, assuming the symbol at the base of the arrow already has a value. For example:

```
> x<-2
> y<-x
> y
[1] 2
```

1.5. Algebraic operations

Addition, subtraction, multiplication and division are the fundamental algebraic operations that bind together numbers and symbols to create mathematical expressions, such as $12 + 45$, x/y , $x - 5$.

The result of addition is called a **sum**. The number zero is also known as the **identity operator** in addition and subtraction because it leaves the result unaffected ($a \pm 0 = a$). The result of subtraction is called a **difference** and is sometimes denoted by an upper case Greek delta prefix (Δ). Sometimes, it is convenient to interpret subtraction as the addition of a negative number because, numerically, the result is exactly the same.

Example 1.6: Size matters in male garter snakes

Garter snake, genus *Thamnophis*

Mating success in male garter snakes appears to depend on body length (Shine *et al.*, 2000). The longest (l_{Long}) and shortest (l_{Short}) adult males recorded in a population were, respectively, 50 and 45 cm. Therefore, the difference between a successful and an unsuccessful male is $\Delta l = l_{Long} - l_{Short} = 50 \text{ cm} - 45 \text{ cm} = 5 \text{ cm}$. Although body length can only increase with time, an individual's weight w may go either up or down. We can express the overall change in w during a 28-day period as the sum of weekly measurements of weight change (gain or loss). For example:

$$\Delta w_{Total} = \Delta w_1 + \Delta w_2 + \Delta w_3 + \Delta w_4 = 5 + 7 + (-4) + (-2) = 6\text{g} \quad (1.6)$$

The result of multiplication is called a **product** and the result of division is called a **ratio**. Remember the various incarnations of multiplication. If a and b are defined as numbers then, in mathematical notation, $a \times b$, $a \cdot b$, ab are all the same. The **identity operator** in multiplication is 1. Multiplication is allowed between any two numbers, but division by zero is not. Products and ratios can be combined in all possible ways (Table 1.1):

Example 1.7**Table 1.1**

Operation	Example	Simplifies to
Product of products	$(1 \times 2) \times (3 \times 4)$	$1 \times 2 \times 3 \times 4$
Product of ratios	$\frac{1}{2} \times \frac{3}{4}$	$\frac{1 \times 3}{2 \times 4}$
Ratio of products	$\frac{1 \times 3}{2 \times 4}$	$\frac{3}{8}$
Ratio of ratios	$\frac{\frac{1}{2}}{\frac{3}{4}}$	$\frac{1 \times 4}{2 \times 3}$

Furthermore, ratios with the same denominator can be added. Any two ratios can be added by first forcing them to have the same denominator.

Example 1.8

$$\frac{2}{3} + \frac{1}{5} = \left(\frac{2}{3} \times \frac{5}{5}\right) + \left(\frac{1}{5} \times \frac{3}{3}\right) = \frac{10}{15} + \frac{3}{15} = \frac{10+3}{15} = \frac{13}{15} \quad (1.7)$$

Multiplication gives rise to powers. Hence, multiplying the same number (a) a total of n times gives

$$\underbrace{a \times a \times a \times a \dots}_{n \text{ times}} = a^n \quad (1.8)$$

In this notation, a is called the **base** and n the **exponent**. An exponent of 1 returns the base and is therefore the **identity operator for powers**.

$$a^1 = a \quad (1.9)$$

Multiplication of powers with the same base translates to addition of their exponents. Here's why:

$$a^n a^m = \underbrace{(a \times a \times a \dots)}_{n \text{ times}} \times \underbrace{(a \times a \times a \dots)}_{m \text{ times}} = \underbrace{a \times a \times a \dots}_{n+m \text{ times}} = a^{n+m} \quad (1.10)$$

Powers with negative exponents have the following interpretation:

$$a^{-n} = \frac{1}{a^n} \quad (1.11)$$

To see why this is useful, consider the following example.

Example 1.9

$$\frac{a^3}{a^2} = \frac{a \times a \times a}{a \times a} = \frac{a}{a} \times \frac{a}{a} \times a = 1 \times 1 \times a = a \quad (1.12)$$

or, alternatively

$$\frac{a^3}{a^2} = a^3 a^{-2} = a^{3-2} = a \quad (1.13)$$

An exponent of 0 indicates that the base should not even appear once and yields 1.

$$a^0 = 1 \quad (1.14)$$

To see why, consider an exponent of zero as the result of two equal and opposite exponents (say 1 and -1). This gives

$$a^0 = a^{1-1} = a^1 \times a^{-1} = \frac{a}{a} = 1 \quad (1.15)$$

Therefore, Equation (1.14) does not hold for $a = 0$, because $0/0$ is not defined.

Roots of numbers are a by-product of powers. The n th root of a number b (written $\sqrt[n]{b}$) can be thought of as the number a that was raised to the power n to give b . The alternative notation for roots uses fractional powers. In short,

$$\text{If } a^n = b \text{ then } \sqrt[n]{b} = a \text{ and } b^{\frac{1}{n}} = a \quad (1.16)$$

Example 1.10

The shorthand for the square root of four is $\sqrt{4}$. This can also be written $\sqrt[2]{4}$ or $4^{\frac{1}{2}}$. Note that 4 can result from a power of 2 in one of two ways, either 2^2 or $(-2)^2$. Hence, mathematically, the square root of 4 is written ± 2 , because it can be either 2 or -2 . This applies to all even

roots. In some cases, we may be able to use biological first principles to specify the result further. For example, negative numbers are not plausible values for variables like body length or population size.

Finally, a particularly useful operator, the **absolute value**, is used to turn any real number into a positive number with the same magnitude. Formally,

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{otherwise} \end{cases} \quad (1.17)$$

Example 1.11

If $x = -3$, then $|x| = |-3| = 3$. If x is an unknown negative number, then $|x| = -x$, which, perversely enough, is a positive number. So, given an expression like $-x$, don't be misled into thinking of it as negative.

There is an agreed order to how the different parts of a complicated algebraic expression should be calculated. The priority of operations is as follows: powers, divisions and multiplications, additions and subtractions. Brackets can be used to override this order as required. Also, the notations for absolute values ($| \ |$) and roots ($\sqrt{\quad}$) operate as brackets, signifying the extent of their application within the expression.

Example 1.12

The expression $\frac{1}{2} \times 4^2 + 1$ has the value 9. Alternative priorities are possible and can be specified using brackets: $\frac{1}{2} \times (4^2 + 1) = 8.5$ or $(\frac{1}{2} \times 4)^2 + 1 = 5$.

The expression $\sqrt{1+2+1+5}$ has the value 3. The scope of the square root can be modified by the length of the overhanging line: $\sqrt{1+2+1}+5 = 7$. A clearer way of writing these two expressions uses brackets and fractional powers: $(1+2+1+5)^{\frac{1}{2}} = 3$, $(1+2+1)^{\frac{1}{2}}+5 = 7$.

Finally, the expression $|-1-1|$ is not the same as $|-1|-1$.



1.5: Encoding algebraic expressions

Algebraic operations are denoted as follows in R:

Addition and subtraction	+ -
Multiplication and division	* /
Power	^
Square root	sqrt ()
Absolute value	abs ()

For example, the mathematical expression $\sqrt{5^2 - 2^2} / (2 \times 3)$ would be encoded as `sqrt(5^2-2^2)/(2*3)`. Note that, in R, as in any other computer language, only parentheses () can be used for specifying priority in algebraic expressions. Other types of bracketing, such as [] or {}, are reserved for different purposes.

1.6. Manipulating numbers

Here is a trick question: if the number 1.30719 is printed at the end of a calculation, do you interpret it as rational or irrational? Strictly speaking, the rational number 1.30719 can be generated as the ratio of two integers ($200/153$) but an irrational number beginning 1.30719... and having infinite digits may also be curtailed to 1.30719. This is an example of **numerical approximation**, also known as **rounding off**. The number of **decimal places** in a numerical result is the number of digits after the decimal point. Hence, 1.30719 has five decimal places. We can round it off to four decimal places to get 1.3072. Similarly, rounding off to an integer gives 1. Rounding off the more ambiguous case of 1.5 can be treated in several different ways. We may, for example, decide to round off to the nearest integer up, to get the result 2. However, this introduces a consistent tendency to obtain numerically larger results. Alternatively, we may toss a coin every time we need to make a decision, but doing this mentally is likely not to be random. A more systematic approach (implemented by R) is to round off to the nearest even number.

Example 1.13

The rational numbers 1.5, 2.5, -1.5 , -2.5 , 0.5 can be rounded off to the integers 2, 2, -2 , -2 , 0.

Rounding off a number leads to some loss of information, known as **rounding error**. The accumulation of even small rounding errors can lead to inaccurate results, a process that can afflict both manual and computerised calculations (Chartier, 2005).

The information content in a particular number can be characterised by its **significant figures**. All nonzero digits in a number are significant. In contrast, 'padding' a number with zeroes doesn't add to the information content of a reported result. If an instrument gives you measurements to one decimal point, reporting a measurement of 1.3 as 1.3000 adds no information. Similarly, multiplying the reported unit of measurement doesn't add any new information to it (1.3 cm is the same as 0.013 m). So, **leading zeroes** are considered nonsignificant and so are **trailing zeroes** before the decimal point (1.3 m = 130 cm). If a number is reported with trailing zeroes after the decimal point, this is taken to indicate the accuracy of the measurement and these zeroes are considered significant.

Example 1.14

The numbers 123, 1.23, 0.123 and 0.0123 all have three significant figures. A reported value of 123.0 has four significant figures.

To simplify the reported results when the number of decimal digits is much greater than the significant figures we can use a combination of significant figures and **orders of magnitude**, expressed as powers of ten.

Example 1.15

The numbers 123, 12.3, 0.123 and 0.0123 can, respectively, be written as 1.23×10^2 , 1.23×10^1 , 1.23×10^{-1} and 1.23×10^{-2} . Calculations with orders of magnitude use the properties of powers. For example, it is rather tricky to see without a calculator that $(0.003 \times 300)/0.0000009$

is equal to a million. But the calculation is greatly simplified using orders of magnitude:

$$\begin{aligned} \frac{0.003 \times 300}{0.0000009} &= \frac{3/1000 + 3 \times 100}{9/10,000,000} = \frac{(3 \times 10^{-3}) \times (3 \times 10^2)}{9 \times 10^{-7}} \\ &= \left(\frac{3 \times 3}{9}\right) \times \left(\frac{10^{-3}10^2}{10^{-7}}\right) = 1 \times 10^{-3+2+7} = 10^6 \end{aligned} \quad (1.18)$$

1.6: Orders of magnitude and rounding off

R decides automatically when it needs to report numbers using orders of magnitude. It does so by using the 'e notation'. Typing `0.000123` gives the output `1.23e-5`, which stands for 1.23×10^{-5} .

Several commands can be used to deal with decimal digits in subtly different ways. The command `round(x, digits = n)` will round the number `x` to `n` decimal digits. The command `floor(x)` rounds the number `x` to the nearest integer down (effectively chopping off all the decimal digits) and the command `ceiling(x)` rounds the number `x` to the nearest integer up. The command `signif(x, digits = n)` gives `x` to `n` significant digits. For example, `round(1.2345, 3)` gives `1.234` but `signif(1.2345, 3)` gives `1.23`. Finally, the command `trunc(x)` rounds both positive and negative numbers towards zero. The difference between `trunc(x)` and `floor(x)` can be illustrated by the following examples:

```
> floor(1.3)
[1] 1
> trunc(1.3)
[1] 1
> floor(-1.3)
[1] -2
> trunc(-1.3)
[1] -1
```

1.7. Manipulating units

Many pencil-and-paper calculations are accompanied by measurement units. Most of us think of units as a nuisance along the way to a numerical result, and it's very tempting to drop them and get on with the numerical calculation. This can backfire in two ways: first, in our haste to number-crunch, we may forget to 'homogenise' the units of measurement to the same unit set. A calculation that simultaneously involves centimetres, metres and inches will yield nonsense. Second, units are a means of verifying the physical relevance of the answer. If a calculation about daily energetic costs yields a result in joules instead of joules per day, it is just possible that the numerical result is also wrong. Units can be treated as symbols and manipulated algebraically alongside the numerical part of the calculation. Example 1.19 will provide an illustration of how this works.

1.8. Manipulating expressions

The presence of symbols in a mathematical expression may make it impossible to reduce it to a single numerical result. Similarly, if we are after a general answer to a problem, we

may prefer to express it symbolically. The objective is then to simplify these expressions as much as possible, by moving things around or collecting similar things together. Sometimes, an expression may first need to be expanded before it can be simplified. To carry out these manipulations it is useful to brush up on the basic rules of algebra. The first type of rule deals with the order of terms within a sum or product. The **commutative** rule states that, when adding or multiplying any two terms a and b , the order of the terms is reversible

$$\begin{aligned} a + b &= b + a && \text{(commutative addition)} \\ ab &= ba && \text{(commutative multiplication)} \end{aligned} \tag{1.19}$$

The **associative** rule says that, when adding or multiplying any three terms a, b and c , the priority with which the operation is performed is not important

$$\begin{aligned} (a + b) + c &= a + (b + c) = (a + c) + b && \text{(associative addition)} \\ (ab)c &= a(bc) = b(ac) && \text{(associative multiplication)} \end{aligned} \tag{1.20}$$

The second set of rules deals with collecting similar things together in sums and products. The sum of identical symbols can be expressed as a multiple of the same symbol. This readily extends to sums of identical expressions or powers:

Example 1.16

$x + x + x = 3x$	Identical symbols
$3x + 8x = 11x$	Multiples of identical symbols
$3x^2 + 8x^2 = 11x^2$	Multiples of identical powers of symbols
$3(a + b) + 8(a + b) = 11(a + b)$	Multiples of identical expressions
$3x^2 + 8x^2 - 3x - 8x = 11x^2 - 11x$	Sum involving different exponents cannot be simplified

Simplifying products is achieved by collecting powers of the same base.

Example 1.17

$xxx = x^3$	Identical base
$x^2x^3 = x^5$	Powers of the same base
$x^{\frac{1}{2}}x^3 = x^{\frac{1}{2}+3} = x^{\frac{1+6}{2}} = x^{\frac{7}{2}}$	Powers of the same base
$(a + b)^2(a + b)^3 = (a + b)^5$	Powers of the same base
a^2b^3	Product of powers with different bases cannot be simplified

Sometimes, sums within brackets need to be multiplied out if this eventually leads to a simplification of the overall expression. This is done by the **distributive** rule which expresses the product of sums as the sum of pairwise products,

$$a(b + c) = ab + ac \tag{1.21}$$

Example 1.18

The expression $3x(x+1)(x-2) + 4x^3 + x(2x+1)$ can be simplified as follows:

$$3x(x^2 - 2x + x - 2) + 4x^3 + (2x^2 + x) = \text{(Distributive rule)}$$

$$3x^3 - 6x^2 + 3x^2 - 6x + 4x^3 + 2x^2 + x = \text{(Distributive rule)}$$

$$3x^3 + 4x^3 - 6x^2 + 3x^2 + 2x^2 - 6x + x = \text{(Commutative rule for sums)}$$

$$7x^3 - x^2 - 5x \quad \text{(Collecting similar terms)}$$

We can employ the same set of rules for manipulating numbers (Section 1.6), units (Section 1.7) and expressions (this section) to find the answer in the following example:

Example 1.19: Energy acquisition in voles

Vole, genus Microtus

A species of vole spends 70% of its time foraging and the remainder of its time resting. While resting, its metabolic rate (in joules per day) is $E_r = 30000 \text{ J} \cdot \text{d}^{-1}$. This increases by 20% during foraging. Females defend territories of about 100 m^2 .

Each cm^2 in the territory produces 0.0005 g of grass that could be entirely consumed by the vole if it spent 100% of its time foraging. The vole gains $\varepsilon = 40.6 \text{ cal}$ from a single gram of grass. We want to calculate the daily rate of energetic gain (or loss) of the animal.

This is a complicated calculation, made particularly messy by the incongruence between the units employed to report physical quantities. For example, area is reported both in m^2 and cm^2 . The first task is to sketch a symbolic solution and simplify it by algebraic manipulation. Daily change in energy is the difference between foraging intake (E_f) and metabolic expenditure (E_c),

$$\Delta E = E_f - E_c \quad (1.22)$$

First, we deal with metabolic costs: If p is the proportion of its time that a vole spends foraging, then its complement, $1 - p$, is the proportion of time that it spends resting. If E_a and E_r are the active and resting metabolic rates, energetic cost can be written as

$$E_c = pE_a + (1 - p)E_r \quad (1.23)$$

This can be simplified somewhat by using the fact that $E_a = 1.2E_r$ (because active metabolic rate exceeds resting metabolic rate by 20%).

$$E_c = p(1.2E_r) + (1 - p)E_r = (1.2p + 1 - p)E_r = (1 + 0.2p)E_r \quad (1.24)$$

Second, we deal with foraging intake: the territory of the vole has area A and each unit of area produces m units of food daily. The vole exploits its territory efficiently, so it consumes a mass mA for every complete day spent foraging. The effective energetic value of this to the animal is

$$E_f = \varepsilon mpA \quad (1.25)$$

Placing Equations (1.24) and (1.25) back into Equation (1.22) we get the symbolic answer

$$\Delta E = \varepsilon mpA - (1 + 0.2p)E_r \quad (1.26)$$

We must now put some numbers into this in the same units (I will use g for mass, d for time, cm^2 for area and J for energy). Some conversions are required: $1 \text{ cal} = 4.184 \text{ J}$, $1 \text{ m}^2 = (100 \text{ cm})^2 = 10^4 \text{ cm}^2$. We can use these to obtain a complete list of required values in a consistent system of units (Table 1.2).

Table 1.2

Quantity	Unconverted	Converted
p	70%	0.7
E_r	$3000 \text{ J} \cdot \text{d}^{-1}$	$3 \times 10^4 \text{ J} \cdot \text{d}^{-1}$
A	100 m^2	10^6 cm^2
m	$0.0005 \text{ g} \cdot \text{d}^{-1} \cdot \text{cm}^{-2}$	$5 \times 10^{-4} \text{ g} \cdot \text{d}^{-1} \cdot \text{cm}^{-2}$
ε	$40.6 \text{ cal} \cdot \text{g}^{-1}$	$1.7 \times 10^2 \text{ J} \cdot \text{g}^{-1}$

We are now ready to do the calculation in Equation (1.26):

$$\Delta E = (1.7 \times 10^2 \text{ J} \cdot \text{g}^{-1})(5 \times 10^{-4} \text{ g} \cdot \text{d}^{-1} \cdot \text{cm}^{-2})(0.7)(10^6 \text{ cm}^2) - 1.14(3 \times 10^4 \text{ J} \cdot \text{d}^{-1}) \quad (1.27)$$

For best book-keeping, we can separate the numbers, orders of magnitude and units and simplify as follows:

$$\begin{aligned} \Delta E &= (1.7 \times 5 \times 0.7)(10^2 10^{-4} 10^6)(\text{J} \cdot \text{g}^{-1} \cdot \text{g} \cdot \text{d}^{-1} \cdot \text{cm}^{-2} \cdot \text{cm}^2) - (1.14 \times 3)10^4(\text{J} \cdot \text{d}^{-1}) \\ &= 5.95 \times 10^4 \text{ J} \cdot \text{d}^{-1} - 3.42 \times 10^4 \text{ J} \cdot \text{d}^{-1} \\ &= 2.53 \times 10^4 \text{ J} \cdot \text{d}^{-1} \end{aligned} \quad (1.28)$$

The subtraction leading to the final result was only possible because the two quantities had the same order of magnitude (10^4) and units ($\text{J} \cdot \text{d}^{-1}$). By way of verification, note that the final result is an energetic rate correctly expressed in $\text{J} \cdot \text{d}^{-1}$.

1.7: Symbolic manipulations

R is primarily intended for data manipulation and consequently has only limited capabilities for symbolic mathematics (but see Chapter 4). Proprietary packages, such as MAPLE and MATHEMATICA, specialise in symbolic calculations and are well worth investigating if you are planning to do complicated algebra.

1.9. Polynomials

Polynomials are a class of algebraic expressions that occur naturally in many ecological problems (Example 1.21). Their generality is matched by their ease of use. Polynomials are the best-behaved mathematical expressions, partly due to their repetitive structure. By way of introduction, consider how you would summarise the following set of four expressions in words

$$\{3x, -5x, 0.33421x, 1298x\} \quad (1.29)$$

You may have noticed that each expression is the product of a real number and the same symbol. Mathematically, these algebraic expressions have the form $a_i x$, where i and a_i have the values shown in Table 1.3.

Here, the symbol a is used to represent a constant number. The subscript i locates a_i in the set and declares that it is not necessarily the same as the others. So, $a_i x$ conveys the *pattern* of the set rather than its detail, and it works just as well with four thousand terms as with four.

Table 1.3

$i =$	$a_i =$
1	3
2	-5
3	0.33421
4	1298

This notation can now be applied to something more complicated. A **polynomial** is an algebraic expression of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \quad (1.30)$$

where the a s are indexed constants and n is a non-negative integer called the **order of the polynomial**. Equation (1.30) contains some clever tricks of notation. It is a sum of $n + 1$ terms. To avoid writing them all out, we interrupt listing them once we have conveyed the pattern, write three dots to indicate that the sequence continues, and then pick it up again to show where the sequence ends. Can you see that all the terms in Equation (1.30) are of exactly the same form? Remember, that $x^1 = x$ and $x^0 = 1$ (Equations (1.9) and (1.14)). So, Equation (1.30) can be written

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x^1 + a_0 x^0 \quad (1.31)$$

where each term of the sum is the product between a constant (known as a **coefficient**) and the variable x raised to a non-negative, integer power.

In each term of the polynomial, the subscript of the coefficient and the superscript of the variable are the same but, while the subscript is a book-keeping convention to distinguish between coefficients, the superscript is an exponent.

Example 1.20

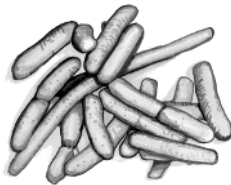
Let's have a look at some specimen polynomials. First, one that looks like Equation (1.30)

$$2x^5 + 4x^4 + 6x^3 + 8x^2 + 3x + 5 \quad (1.32)$$

The highest power is 5, so this is a polynomial of order 5. The coefficients don't have to be positive or integers. The following is also a polynomial of order 5 in which the coefficients of the fourth and second order terms are -4 and -1 .

$$2x^5 - 4x^4 + 2.6x^3 - x^2 + 0.001x + 5 \quad (1.33)$$

The coefficients may also be zero, so that an expression like $4x^5 - 6$ is also a fifth order polynomial.

Example 1.21: The law of mass action in epidemiology

Bacterial, airborne pathogen

A theoretical abstraction that has contributed greatly to the study of epidemics is the law of **mass action**. It asserts (quite unrealistically) that all individuals in a population are identical and interact randomly. The second of these assumptions is known as **perfect**

mixing because it implies that spatial proximity does not influence transmission between any two individuals in a population. To investigate the implications of these assumptions, consider a small population of ten animals, four of which (black discs in Figure 1.5(a)) carry an infectious disease. Because the population is so small, every individual comes into contact with every other individual with approximately the same frequency (say, three times a day). As a result, each infected individual is equally likely to pass the infection to each of the susceptible individuals. The daily number of transmission opportunities between a particular carrier and a susceptible individual can readily be tabulated (Figure 1.5(b)).

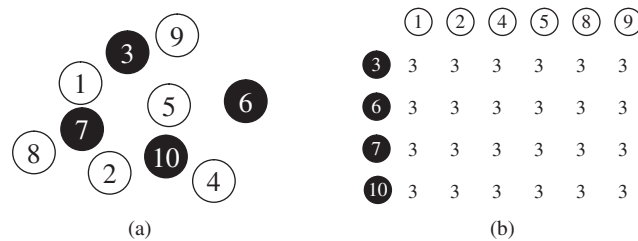


Figure 1.5: (a) A population of ten individuals in which four (shown in black) carry the disease; (b) the number of opportunities of transmission in one day between any two individuals.

The total number of interactions in Figure 1.5(b) is $3 \times 4 \times 6$. If only one out of a thousand interactions leads to transmission, then the number of expected transmissions is $\frac{1}{1000} \times 3 \times 4 \times 6 = 0.072$.

More generally, if K is total population size, N is the number of carriers, b is the rate of interactions between two individuals and c is the proportion of interactions that lead to transmission, then the rate of transmission (γ) is

$$\gamma = cbN(K - N) \quad (1.34)$$

You can read more on the law of mass action in Heesterbeek (2005). For now, note that if we multiply out the right-hand side of Equation (1.34) we get a second order polynomial in the variable N , the number of infected animals

$$-cbN^2 + cbKN \quad (1.35)$$

You can check that this is a typical polynomial of the form $a_2N^2 + a_1N + a_0$ by setting $a_2 = -cb$, $a_1 = cbK$ and $a_0 = 0$.

1.8: Set indexing and element extraction

We saw in this section how indexing the members of a set makes it easier to identify them individually. Consider the set of numbers A , defined as

```
A <- c(2, 4, 1, 6, 0, 5)
```

The documentation for R sometimes refers to such sets as **vectors** (more on vectors in Chapter 6). The n th element of A , in R syntax, is $A[n]$. In the example above, if you type $A[1]$, R will return 2, typing $A[3]$ will give you 1. In addition to this, there are several other ways of using indexing to get specific subsets of a set, as shown in Table 1.4.

Table 1.4

<i>If you want. . .</i>	<i>. . .the R syntax is</i>	<i>Example input</i>	<i>Result</i>
The n th element of A	$A[n]$	$A[5]$	0
All but the n th element	$A[-n]$	$A[-5]$	2 4 1 6 5
The first n elements	$A[1:n]$	$A[1:5]$	2 4 1 6 0
The last n elements	$A[-(1:n)]$	$A[-(1:3)]$	6 0 5
Specific elements from A	$A[c()]$	$A[c(1, 3)]$	2 1
Elements above a value a	$A[A>a]$	$A[A>2]$	4 6 5
Elements between a and b	$A[A>a \ \& \ A<b]$	$A[A>2 \ \& \ A<6]$	4 5
Elements between a and b , inclusive	$A[A>= a \ \& \ A<= b]$	$A[A>= 2 \ \& \ A<= 6]$	2 4 6 5

1.10. Equations

An equation consists of two parts, separated by an equality sign. It states that these two expressions have the same value. Verifying this may be just a matter of numerical calculation. Such equations are called **trivial**.

Example 1.22

Here's a trivial equation

$$1 + 1 = 2 \quad (1.36)$$

The following is also a trivial equation, because it contains no symbols,

$$\frac{\sqrt{\frac{1295.78}{1987.43}} \times (9.32^2)^2}{9569.86} = 0.637 \quad (1.37)$$

Nontrivial equations come in two flavours: an **identity** is valid for all values of its variables (a refresher of basic algebraic identities can be found in Section A1.3 in the appendix). A **conditional equation** is valid only when the variables take specific values (called the **solutions** or **roots** of the equation).

Example 1.23

By trying out different values for x you can convince yourself that the following is an identity

$$(x + 4)^2 = x^2 + 8x + 16 \quad (1.38)$$

whereas

$$x^2 + 8x + 16 = 36 \quad (1.39)$$

is a conditional equation because it is only valid when $x = 2$ or $x = -10$.

Conditional equations may be solved; that is, we can try to find the value(s) of x that satisfy the equation. In a simple equation, this can be done by trial and error but, in general, it is best

achieved by systematically isolating the variable on one side of the equality sign and all the known elements (numbers or symbols with known or assumed known values) on the other. Note, however, that some equations cannot readily be solved. For example, there is no real number that can satisfy the equation $1 + x^2 = 0$.

At this point, I owe you an apology for a particular misuse of terminology: cross-references to numbered mathematical expressions are incorrectly labelled as 'equations', throughout the book. Hence, references like 'Equation (1.30)' are commonplace, even though that particular expression may not strictly be an equation as defined in this section. Unfortunately, for the sake of brevity and uniformity, almost all books and journals commit this intentional error.

1.11. First order polynomial equations

A polynomial can be used to construct an equation, which can always be written as

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0 \quad (1.40)$$

But polynomials don't always appear in this form and you may need to carry out some manipulation before they can be identified.

Example 1.24

The following expression is a third order polynomial equation. Can you see why?

$$\frac{3}{2} + x + \frac{1}{x} = x(3x + 2) - 1 \quad (1.41)$$

To get a first order polynomial equation from the general definition in Equation (1.40) we set the higher-order coefficients to zero (mathematically, $a_i = 0, \forall i > 1$)

$$a_1 x + a_0 = 0 \quad (1.42)$$

The equation is solved by isolating x . To maintain the equality sign, exactly the same things must be done to both sides of it. Subtracting a_0 from both sides

$$a_1 x = -a_0 \quad (1.43)$$

and then dividing both sides by a_1

$$x = -\frac{a_0}{a_1} \quad (1.44)$$

gives the general solution of a first order polynomial equation. It is general because it gives the root of any equation of the form of Equation (1.42) for any values of a_0 and a_1 .

Example 1.25: Population size and composition



Sexual dimorphism in redstarts (*Phoenicurus phoenicurus*)

In a population of tree-nesting birds there are two males for every three females. During the breeding season nests are highly cryptic, but it is known that all females breed and each produces five eggs, of which only three survive to become

fledglings. Although an accurate, independent estimate of population size can be obtained after the breeding season, by that time it's no longer possible to distinguish juveniles from adults. Nevertheless, we want to know the number of males in a population of 70 animals counted just after the breeding season.

Let's denote males by M , females by F , juveniles by J and population size by P . If we assume that no deaths of adults occur during the breeding season, then all four of these quantities refer to a point in time just after the breeding season so that

$$M + F + J = P \quad (1.45)$$

The number of unknown quantities can be reduced by using the information on sex ratio and chick-rearing success. Specifically, we can use the following two facts

$$J = 3F \text{ and } \frac{M}{F} = \frac{2}{3} \quad (1.46)$$

to reduce the four symbols in Equation (1.45) to two. There are three ways to do this:

$$M + 1.5M + 3(1.5M) = P \quad (1.47)$$

$$\frac{2}{3}F + F + 3F = P \quad (1.48)$$

$$\frac{2}{3} \left(\frac{1}{3}J \right) + \frac{1}{3}J + J = P \quad (1.49)$$

Since we are interested in the number of males, Equation (1.47) is the more appropriate. We can simplify it further and solve it for M

$$\begin{aligned} M(1 + 1.5 + 4.5) &= P \\ M &= \frac{P}{7} = 10 \end{aligned} \quad (1.50)$$

1.12. Proportionality and scaling: a special kind of first order polynomial equation

A scale drawing is one that maintains the proportions but not the size of its subject. Decreasing or increasing a picture while maintaining its relative proportions is called **scaling** (Figure 1.6). You will probably have encountered scales on maps, where an inscription of, say, 1:100 000 indicates that a single unit of length on the map corresponds to 10^5 units of length on the ground. Formally, a scaling operation is given by a first order polynomial equation with $a_0 = 0$. If we use y to denote the size of the scaled object and x for the size of the real object, then

$$y = ax \quad (1.51)$$

where a is called the **proportionality constant**. If the proportionality constant is greater than 1, then the scaling operation is a **magnification**, if it is smaller than 1 it is called a **contraction**. If we don't know the value of a , but we know that y and x are proportional, we state this using the notation $y \propto x$.

Any two dimensions (x_1, x_2) of an object (e.g. height and width in Figure 1.6) are related to the scaled dimensions by the same constant $y_1 = ax_1$ and $y_2 = ax_2$. Scaling maintains the

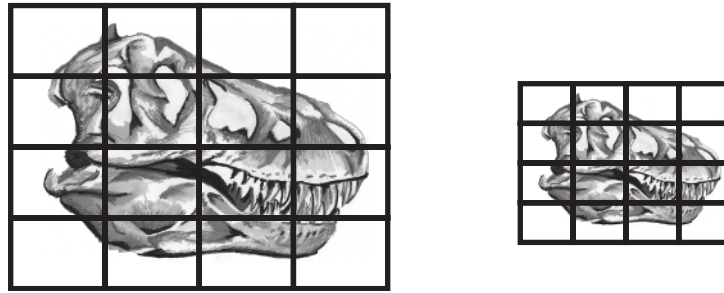


Figure 1.6: The second drawing is smaller than the first but it maintains the same ratio of height over width.

relationship between any two of the object’s dimensions so that

$$\frac{y_1}{y_2} = \frac{x_1}{x_2} \tag{1.52}$$

Equation (1.52) is an alternative way of stating proportionality. A number of scaling problems can be solved using either Equation (1.51) or Equation (1.52).

Example 1.26: Simple mark-recapture



Ringling is a traditional mark-recapture field technique for bird species

Mark-recapture methods are used to estimate the size of animal populations and vital rates, such as survival. Simple mark-recapture assumes that the population has a constant size. This implies that there is no migration into or out of the population (in which case the population is said to be **closed**), and that the study happens quickly enough to exclude the possibility of new births and deaths. Suppose we make two visits to the population. In the first, a number of animals n are captured and marked. In the second, we capture M animals, and find that m of these had already been marked (i.e. they are **recaptures**). If animals mix freely and the event of capture does not affect the chance of recapture, then we can use proportionality to estimate the (unknown) size of the population (N): the proportion of marked animals in the population should be the same as the proportion of recaptures in the second sample

$$\frac{n}{N} = \frac{m}{M} \tag{1.53}$$

We can now solve this for N

$$N = \frac{M}{m}n \tag{1.54}$$

The proportionality constant linking N and n is M/m . This constant cannot be less than 1, because the number of recaptures cannot exceed the total number of marked animals. If the proportionality constant is close to 1, this implies that almost all of the animals in the second sample were found to be marked or, equivalently, that almost the entire population was caught in the first visit.

Example 1.27: Converting density to population size

Regular quadrats, used to sample plant diversity

It is usually difficult to make a complete inspection (a **census**) of a field site to count the number of occurrences of a species. In plant ecology, this is approached by randomly placing quadrats of known size (a) at the site and counting the number of occurrences of the species within each. To estimate the population size N in a field of area A , we may count the number of occurrences n in a total of m quadrats, each of size a . Assuming that the quadrats are representative of the field, then the density of plants in the quadrat ($n/(ma)$) should be the same as the overall density of plants in the field (N/A), giving the estimate

$$N = \frac{A}{ma} n \quad (1.55)$$

1.13. Second and higher order polynomial equations

As a starting point for this discussion consider the following example:

Example 1.28

$$(3x - 2)(x - 4) = 0 \quad (1.56)$$

Setting $x = 4$ gives $(12 - 2)(4 - 4) = 0$, hence 4 is a solution (or root) of this equation. There is also a second root: $2/3$.

Equation (1.56) has the general form

$$(s_1x - r_1)(s_2x - r_2) = 0 \quad (1.57)$$

where s_1, s_2, r_1 and r_2 are constants. Look at the two parts of the product on the left-hand side. They are first order polynomials. A product of zero can only arise if either or both of the following equations are true:

$$s_1x - r_1 = 0 \text{ or } s_2x - r_2 = 0 \quad (1.58)$$

The two roots of Equation (1.58), and hence the solutions of Equation (1.57), are

$$x = \frac{r_1}{s_1} \text{ and } x = \frac{r_2}{s_2} \quad (1.59)$$

However, if the left-hand side of Equation (1.57) is multiplied out prior to solving, it is

$$\begin{aligned} (s_1x - r_1)(s_2x - r_2) &= s_1s_2x^2 - s_1r_2x - s_2r_1x + r_1r_2 \\ &= s_1s_2x^2 + (-s_1r_2 - s_2r_1)x + r_1r_2 \end{aligned} \quad (1.60)$$

The constant components of the result can be renamed as follows:

$$a_2 = s_1s_2, a_1 = -s_1r_2 - s_2r_1, a_0 = r_1r_2 \quad (1.61)$$

Using these definitions we can rewrite Equation (1.57) as

$$a_2x^2 + a_1x + a_0 = 0 \quad (1.62)$$

This is a second order polynomial equation, which also goes by the name **quadratic equation**. Although it is the same as Equation (1.57), it is impossible to tell at a glance what its roots are. Unfortunately, second order polynomial equations appear in the form of Equation (1.62) more often than in the form of Equation (1.57) and we need a reliable way of solving them. The roots (x_1 and x_2) of a quadratic equation are given by the **quadratic formula**

$$x_1, x_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2a_0}}{2a_2} \quad (1.63)$$

The plus-or-minus sign (\pm) means that there are a maximum of two roots obtained by respectively adding and subtracting the entire expression $\sqrt{a_1^2 - 4a_2a_0}$ in the numerator. The expression within the square root is called the **discriminant**. It has a special name because it gives information about how many roots the quadratic equation has. If the discriminant is positive, its square root can be added and subtracted from a_1 to give two roots. If the discriminant is zero, it makes no difference in the numerator and the equation has a single root. Finally, if the discriminant is negative, the two roots are complex numbers (see Section 1.17).

Example 1.29

The roots of the quadratic equation $2x^2 - 11x - 21 = 0$ are found as follows

$$\frac{-(-11) \pm \sqrt{11^2 - 4 \cdot 2 \cdot (-21)}}{2 \cdot 2} = \frac{11 \pm \sqrt{289}}{4} = \begin{cases} 7 \\ -3/2 \end{cases} \quad (1.64)$$

Example 1.30: Estimating the number of infected animals from the rate of infection



In a perfectly mixed population of size $K = 1080$ animals, a total of $\gamma = 3$ new cases of an infectious disease appeared today. The daily rate of encounters between animals is $b = 0.03$ and a proportion $c = 0.001$ of encounters lead to transmission. The disease is not lethal but individuals do not recover within the timescale of a single disease outbreak. We would like to estimate the total number of infected animals (N) in the population. To do this, we can rearrange the model in Equation (1.34) as follows:

$$(cb)N^2 - (cbK)N + \gamma = 0 \quad (1.65)$$

Placing some of the constants in parentheses highlights the fact that this is a second order polynomial equation. It can have a maximum of two solutions (say, N_1, N_2) that can be found by applying the quadratic formula in Equation (1.63)

$$N_1, N_2 = \frac{cbK \pm \sqrt{(-cbK)^2 - 4cb\gamma}}{2cb} \quad (1.66)$$

Replacing the symbols with numbers and doing the arithmetic,

$$\begin{aligned} N_1, N_2 &= \frac{0.001 \times 0.03 \times 1080 \pm \sqrt{(0.001 \times 0.03 \times 1080)^2 - 4 \times 0.001 \times 0.03 \times 3}}{2 \times 0.001 \times 0.03} \\ &= \frac{3.24 \times 10^{-2} \pm 2.63 \times 10^{-2}}{0.006 \times 10^{-2}} = \frac{3.24 \pm 2.63}{0.006} \cong \begin{cases} 978 \\ 102 \end{cases} \quad (1.67) \end{aligned}$$

The answers have been rounded off to the nearest integer animal and so the symbol \cong expresses approximate equality. The appearance of two answers that are so different (relative to the total population size of 1080) may seem biologically unrealistic. However, both values are feasible because the (relatively low) rate of three new cases per day can occur either when there are only a few infected individuals (giving $N_1 = 102$) or a few uninfected individuals (giving $N_2 = 978$).


Higher order polynomial equations are easy to write, but not so easy to solve. Analytic, general solutions such as those presented in Equations (1.44) and (1.63) exist only for equations up to fourth order. However, it may be possible to simplify the equation into lower-order components.

Example 1.31

Strictly speaking, $a_6x^6 + a_5x^5 + a_4x^4 = 0$ is a sixth order equation but, since the terms below fourth order are missing, it can be written as the product of two lower-order components

$$a_6x^6 + a_5x^5 + a_4x^4 = x^4(a_6x^2 + a_5x + a_4) \quad (1.68)$$

When such simplifications are not possible, the easiest way to find the roots is **numerical approximation**, a group of computationally efficient methods for finding numerical answers by trial-and-error. In doing this, it is useful to know that the maximum number of solutions of a polynomial equation is equal to the order of the equation. R can handle equations up to order 49 in this way.

 **1.9: Numerical solution of higher order polynomial equations**

The command for numerically solving polynomial equations is `polyroot()` which requires a list of the polynomial's coefficients in increasing order. To find the roots of the polynomial equation $2x^3 + 19x^2 - 150x - 100 = 0$ type

```
> polyroot(c(-100, -150, 19, 2))
[1] -0.6210104+0i 5.5716520+0i -14.4506415-0i
```

These are the three solutions of the third order equation (namely, -0.621 , 5.572 and -14.451). The $\pm 0i$ parts are of no relevance to the answer because they are effectively zero. Their significance should become clearer in Section 1.17.

1.14. Systems of polynomial equations

The equations we have looked at so far contain a single variable but this does not have to be so.

Example 1.32

The equation $3x - 4y + 8 = 0$ contains two variables, x and y .

To get numerical results for multiple variables we need just as many *unique* equations. Complementing the equation in Example 1.32 with the equation $1.5x - 2y + 4 = 0$ does not lead to two unique equations because the second equation is simply a scaling of the first. When

more than one variable are entangled in more than one equation, we call these equations **simultaneous** or **coupled**. A list of coupled equations forms a **system of equations**.

Example 1.33

A system of two coupled equations involving the variables x and y is

$$\begin{aligned} 3x - 4y + 8 &= 0 \\ x + 2y - 4 &= 0 \end{aligned} \tag{1.69}$$

Solving systems of polynomial equations relies on the same ideas used for single, self-contained equations, although more work is required. One approach, called **solution by substitution** is to pick any equation from the system and solve it with respect to one of the variables it contains. The solution is phrased in terms of the remaining variables. The next step is to choose another equation from the system and replace the first variable with the expression that was obtained in the previous step. This procedure is repeated until an expression is obtained containing only one variable. Then, if all has gone well, a numerical answer for that variable can be obtained. Finally, these steps need to be retraced backwards, replacing variables with their values in the process.

Example 1.34

The system in Example 1.33 can be solved as follows:

Step 1: Take the first equation and solve it for one of the two variables, say x , expressing it in terms of y .

$$x = \frac{4y - 8}{3} \tag{1.70}$$

Step 2: Substitute Equation (1.70) into the second equation of the system

$$\frac{4y - 8}{3} + 2y - 4 = 0 \tag{1.71}$$

x has now vanished and we are left with one equation containing only y .

Step 3: Solve the second equation for the remaining variable

$$\begin{aligned} 4y - 8 + 6y - 12 &= 0 \\ 10y - 20 &= 0 \\ y &= 2 \end{aligned} \tag{1.72}$$

Step 4: Retrace the steps replacing variables by values. Substituting y by 2 in Equation (1.70) gives $x = 0$.

The complexity of a system of equations is potentially unlimited. We could, for example, have a lot more variables (and equations), or the equations themselves could be more complicated. In Chapter 6, we will discuss more efficient ways of solving large systems. For now, here is a biological example:

Example 1.35: Deriving population structure from data on population size

The size of a particular bird population has been accurately estimated to be 18 050 just before the breeding season and 20 130 at some time after it. We know that 10% of males and 5% of females died in the interval between the two counts. It is also known that each female produced 0.3 juveniles that survived (as adults) until the post-breeding count. We would like to calculate the number of males and females at the beginning of the breeding season. Using intuitive notation we have the system

$$\begin{aligned} M + F &= 18050 \\ 0.9M + 0.95F + 0.3F &= 20130 \end{aligned} \quad (1.73)$$

This can be solved to give $M = 6950, F = 11100$

1.15. Inequalities

Often, the available biological information can only be specified up to a range of values rather than a countable set of solutions, leading to **inequalities**.

Example 1.36

Table 1.5 lists some verbal statements and the corresponding inequalities.

Table 1.5

<i>Statement</i>	<i>Corresponding inequalities</i>
x is greater than 3	$x > 3$
x is at least 3	$x \geq 3$
x is no larger than 3 and no smaller than 1	$1 < x < 3$
x is outside the range 1 to 3	$x < 1, x > 3$

The fourth statement requires two separate inequalities. The comma between them implies that either $x < 1$ or $x > 3$ will be true for any given value taken by x . An expression like $1 > x > 3$ is not meaningful because both inequalities cannot be true for any given value of x .

Example 1.37: Minimum energetic requirements in voles

The voles from Example 1.19 feed on two types of plant during part of the year, each with a different energetic value, ε_1 and ε_2 (in some appropriate unit, say $\text{J}\cdot\text{g}^{-1}$). The amounts (say, x and y) consumed by any given vole on any given day may vary but, in total, each vole requires $E\text{J}\cdot\text{day}^{-1}$ to survive. Voles that satisfy their energetic requirements are described by the following inequality

$$x\varepsilon_1 + y\varepsilon_2 \geq E \quad (1.74)$$

An inequality such as Equation (1.74) may be solved with respect to a single variable. As with equations, this is achieved by using the basic rules of algebra (see Section 1.8) to rearrange the terms on either side of the inequality and applying the same algebraic operations to both sides to move things from one side to the other. For inequalities, the second of these two actions requires particular care.

Example 1.38

Table 1.6 considers different algebraic operations applied to both sides of the inequality $4 < 5$.

Table 1.6

<i>Operation</i>	<i>Outcome</i>
+1	$5 < 6$
-1	$3 < 4$
$\times 1$	$4 < 5$
$\times(-1)$	$-4 > -5$

Notice that the first three operations maintain the direction of the inequality but multiplication (or division) by a negative number reverses it.

In Section 1.5 we saw how absolute values and even powers can turn negative numbers into positive ones. Therefore, inequalities involving their variable in even powers or absolute values contain some ambiguity that needs to be expressed when solving for the variable.

Example 1.39

Table 7.1 provides the interpretation of four inequalities involving powers and absolute values.

Table 1.7

<i>If we are given</i>	<i>... .then this implies</i>
$x^2 < 9$	$-3 < x < 3$
$x^2 > 9$	$x < -3$ or $x > 3$
$ x < 3$	$-3 < x < 3$
$ x > 3$	$x < -3$ or $x > 3$

Example 1.40

The inequality $x^2 - 4x - 1 > (x - 3)x$ can be solved for x as follows:

- | | |
|----------------------------------|---------------------------|
| ① Expand the right-hand side | $x^2 - 4x - 1 > x^2 - 3x$ |
| ② Subtract x^2 from both sides | $-4x - 1 > -3x$ |
| ③ Add $3x$ to both sides | $-x - 1 > 0$ |
| ④ Add 1 to both sides | $-x > 1$ |
| ⑤ Multiply both sides by -1 | $x < -1$ |

(remembering that this reverses the inequality).

R 1.10: Comparisons and the logical TRUE and FALSE

It is often useful to be able to test for equality or inequality between two quantities. For example, consider the following assignments (the semicolon allows us to put multiple statements in one line):

```
a<-2; b<- -4; c<-1; d<-8
p<-(a+b)/(c-d)
q<-(a-b)/(c+d)
```

It is not immediately clear if p is equal to, smaller than or greater than q . We can query R to tell us using any one of six logical operators $=, !=, >, >=, <, <=$. The response can either be **TRUE** or **FALSE** with numerical values 1 and 0, respectively. Table 1.8 shows some examples of queries and their responses for the above example.

Table 1.8

<i>The query...</i>	<i>...translates to...</i>	<i>...and R responds with...</i>
Is p the same as q ?	$p==q$	FALSE
Is p different from q ?	$p!=q$	TRUE
Is p greater than q ?	$p>q$	FALSE
Is p smaller than or equal to q ?	$p<=q$	TRUE

1.16. Coordinate systems

Comparison between two real numbers m and n can have one of three outcomes ($m > n, m = n, m < n$). Such comparisons enable us to order numbers.

Example 1.41

The small set of integers $S = \{5, 3, 9, 4, -4, 1, 7, 10, -2\}$ can be rewritten in ordered form as $S = \{-4, -2, 1, 3, 4, 5, 7, 9, 10\}$.

The entire ordered set of real numbers can't be represented in the form of Example 1.41 because there is an infinity of real numbers before, after and between any two real numbers. We therefore need a different way to visualise such noncountable sets. If there are infinite real numbers between any two real numbers, then we need something that is so small that an infinity of it would fit in the gap. We envisage such a dimensionless object and call it a **point**. We arrange an infinity of points along the **line of real numbers** to visualise the ordered set of real numbers (Figure 1.7). We think of it as having an origin (a point corresponding to the number 0) and extending indefinitely to the left and to the right. Viewing numbers on this line is more informative than ranking them because we can also show how far apart they are from each other.

Example 1.42

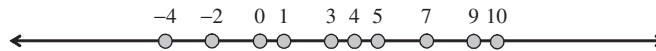


Figure 1.7: Any subset of the real numbers can be arranged along a line.

Such one-to-one correspondence between numbers and points along a line (or **axis**) creates a **coordinate system on the line**. However, we often want to characterise things according to more than one of their attributes. We may, for example, want to characterise a species of animal according to both its life-span and its metabolic rate. We therefore need to visualise pairs of numbers on a **coordinate system on the plane**. Mathematicians most frequently use the Cartesian system. This consists of two axes intersecting at right angles at their origin (O). The same unit is used to place numbers along both axes of a Cartesian coordinate system. Just like coordinate systems on the line, systems on the plane are visualisation tools. Each pair of numbers must be ordered so that we know which one is to be mapped on which axis. Conventionally, the first number in the ordered pair is used as a coordinate for the horizontal axis and the second for the vertical. To represent a point on a planar system of coordinates, you need to mark the two numbers on their corresponding axes and from there draw two lines parallel to the other axes. The representation of the ordered pair is the point of intersection of these two lines.

Example 1.43: Non-Cartesian map projections

The ordered pair (2,4) can be plotted on the Cartesian system of coordinates (Figure 1.8(a)) but the approach to plotting would be the same even if the coordinate system was not Cartesian; for example, if the axes were not perpendicular and the unit of length was different along each axis (Figure 1.8(b),1.8(c))

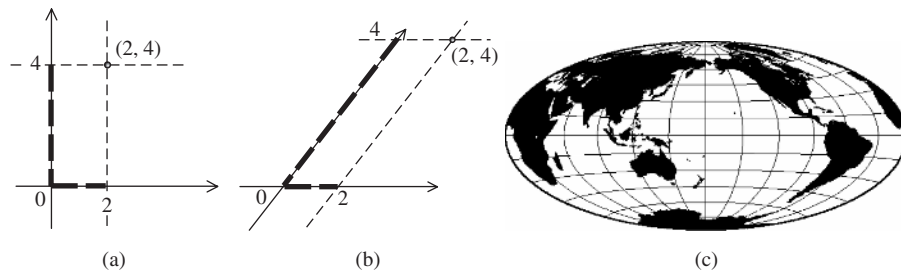


Figure 1.8: The same ordered pair of coordinates in a Cartesian (a) and non-Cartesian system (b). Projections of the globe onto two-dimensional maps (c) are a familiar example of non-Cartesian coordinates.

1.11: Basic plotting

As part of the larger body of literature motivated by the 'rate of living' hypothesis (Pearl, 1928), a study was undertaken to look at the relationship between mass-specific, daily energy expenditure and body mass in small mammals (Speakman *et al.*, 2002). I will ignore which indexes were used for measuring the two variables (to avoid specifying units), but simply seek to visualise the data. The body mass and energy expenditure of 12 species were

```
> bodM<-c(3.9, 3.1, 2.0, 8.3, 5.8, 5.1, 3.7, 6.6, 3.2, 4.3, 2.9, 7.4)
> enExp<-c(-1.1, -0.5, 2.5, -2.2, -0.7, -0.1, 0.4, -0.4, 0.8, 0.7,
0.4, -1.5)
> plot(bodM, enExp, xlab="Body Mass", ylab="Energy expenditure")
```

Lines 1 and 2 assign the data to the symbols `bodM` and `enExp`. The order of the numbers matters because these are paired measurements for each species. The command `plot` places the variables `bodM` and `enExp` on a system of coordinates. The axes are labelled as instructed by the options `xlab` and `ylab` (remember that options within a command are assigned using `=`, not the arrow `<-`). The graph is shown in Figure 1.9.



Figure 1.9: Graphical output of the data on body mass and energy expenditure.

1.17. Complex numbers

Although mathematics is renowned as a system of rigorous and formal rules, progress has very often come about by acts of rule-breaking. Complex numbers are a good example. Their introduction was motivated by the need to make sense of quadratic equations with a negative discriminant (see Section 1.13). The elementary maths approach at this point is to report no solution. However, the lack of real roots does not necessarily imply a lack of useful information. To work around this problem, a new number i is introduced. This is called an **imaginary** number and is defined as

$$i^2 = -1 \text{ or, equivalently } i = \sqrt{-1} \quad (1.75)$$

Example 1.44

The square root of any negative number can now be ‘calculated’. For example, $\sqrt{-9} = \sqrt{-1 \times 9} = \sqrt{-1} \times \sqrt{9} = 3i$.

Mixing real with imaginary numbers gives rise to the set of complex numbers, already mentioned in Section 1.2. A complex number z is always of the form

$$z = a + bi \quad (1.76)$$

where both a and b are real numbers. This section presents the ground rules for how to work with complex numbers graphically and algebraically.

Unlike any other type of number, complex numbers cannot readily be compared. So, we cannot say whether $3 + 2i$ is greater or smaller than $2 + 3i$ (despite the fact that both of them are numbers and they are clearly not the same). Instead, we can visualise them in a two-dimensional system of coordinates by plotting the real part on one axis and the imaginary part on another (Figure 1.10).

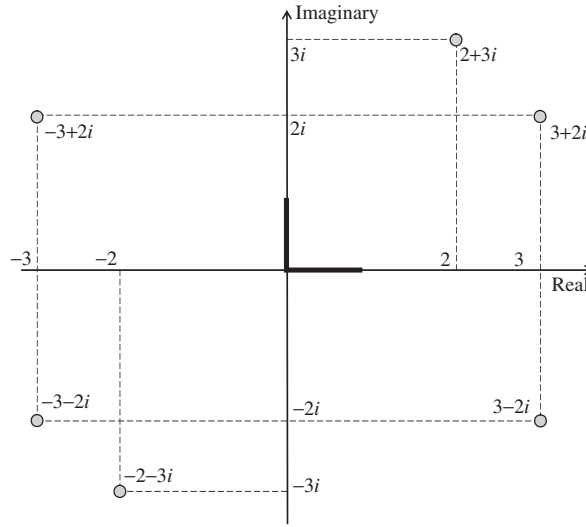


Figure 1.10: The complex plane, with the real component plotted along the horizontal axis and the imaginary part on the vertical axis. In this example there are two conjugate pairs, $3 \pm 2i$ and $-3 \pm 2i$.

In the complex plane, numbers that are symmetric around the horizontal axis (the line of real numbers) are called **complex conjugates**. Such numbers are usually highlighted with an over-bar notation. So, the conjugate of any complex number $z = a + bi$ is $\bar{z} = a - bi$.

For any two complex numbers, say $z_1 = a_1 + b_1i$ and $z_2 = a_2 + b_2i$, addition and subtraction are done separately for the real and imaginary parts

$$\begin{aligned} z_1 + z_2 &= (a_1 + a_2) + (b_1 + b_2)i \\ z_1 - z_2 &= (a_1 - a_2) + (b_1 - b_2)i \end{aligned} \tag{1.77}$$

Multiplication is equally common-sense if we think of the imaginary number i as just any other symbol

$$\begin{aligned} z_1 z_2 &= (a_1 + b_1i)(a_2 + b_2i) \\ &= a_1 a_2 + a_1 b_2 i + a_2 b_1 i + b_1 b_2 i^2 \\ &= a_1 a_2 + (a_1 b_2 + a_2 b_1)i - b_1 b_2 \end{aligned} \tag{1.78}$$

Note that the last step in Equation (1.78) uses the definition of i in Equation (1.75) to get $b_1 b_2 i^2 = -b_1 b_2$. A particularly important version of Equation (1.78) arises when the two complex numbers being multiplied are conjugate,

$$\begin{aligned} z \bar{z} &= (a + bi)(a - bi) \\ &= a^2 + abi - abi - b^2 i^2 \\ &= a^2 + b^2 \end{aligned} \tag{1.79}$$

So, in general, the product of two complex conjugates is a real number.

Division of a complex number by a noncomplex number is a straightforward division of the real and imaginary parts

$$\frac{z}{c} = \frac{a + bi}{c} = \underbrace{\left(\frac{a}{c}\right)}_{\text{Real}} + \underbrace{\left(\frac{b}{c}\right)}_{\text{Imaginary}} i \tag{1.80}$$

However, division of two complex numbers is a bit tougher. Feel free to avoid the following derivation if you are not in the mood for mental gymnastics, and skip to the result in Equation (1.87).

We can think of division between two complex numbers as the multiplication of one with the inverse of the other

$$\frac{z_1}{z_2} = z_1 \frac{1}{z_2} \quad (1.81)$$

Unfortunately, we don't really know how to calculate the inverse of a complex number. We therefore start by assuming that it exists, and try to calculate it. In general, if $z = a + bi$, then its inverse may also be a complex number, say $1/z = a' + b'i$. To find a' and b' we begin by stating a requirement: for one number to be the inverse of the other, their product must equal 1

$$z \frac{1}{z} = 1 \quad (1.82)$$

We can expand this to see where it leads

$$\begin{aligned} (a + bi)(a' + b'i) &= 1 \\ (aa' - bb') + (ab' + ba')i &= 1 \end{aligned} \quad (1.83)$$

This says that the complex number with real part $(aa' - bb')$ and imaginary part $(ab' + ba')$ must equal 1. Since 1 is a real number, this statement can only be true if the coefficient $(ab' + ba')$ of the complex part is zero, implying

$$\begin{aligned} ab' + ba' &= 0 \\ aa' - bb' &= 1 \end{aligned} \quad (1.84)$$

The only two unknown quantities in Equations (1.84) are a' and b' . So, we have a system of two equations in two unknowns that can be solved (see Section 1.14) to get

$$a' = \frac{a}{a^2 + b^2}, \quad b' = -\frac{b}{a^2 + b^2} \quad (1.85)$$

Putting everything together, the inverse of a complex number is

$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} \quad (1.86)$$

Which, in turn, yields a rule for dividing two complex numbers z_1 and z_2

$$\frac{z_1}{z_2} = \frac{z_1\bar{z}_2}{z_2\bar{z}_2} \quad (1.87)$$

Note that the denominator on the right-hand side ($z_2\bar{z}_2$) is the product of complex conjugates and therefore, according to Equation (1.79), just a real number.

Example 1.45

Here are the four operations applied to the numbers $z_1 = -2 + i$ and $z_2 = 3 - 2i$:

$$z_1 + z_2 = (-2 + 3) + (1 - 2)i = 1 - i$$

$$z_1 - z_2 = (-2 - 3) + (1 + 2)i = -5 + 3i$$

$$z_1 z_2 = -6 + 4i + 3i + 2 = -4 + 7i$$

$$\frac{z_1}{z_2} = \frac{(-2 + i)(3 + 2i)}{(3)^2 + (-2)^2} = \frac{-6 - 4i + 3i - 2}{13} = -\frac{8}{13} - \frac{1}{13}i$$

1.12: Complex numbers in R

If you type something like $(-2)^{0.5}$ or `sqrt(-2)`, R will return the value NaN meaning 'not a number'. So, despite the fact that some of the commands in R will deal with complex numbers (see R1.9), R is generally reluctant to perform complex algebra. Nevertheless, given a complex number z , you can extract its real and imaginary parts by typing `Re(z)` and `Im(z)`.

1.18. Relations and functions

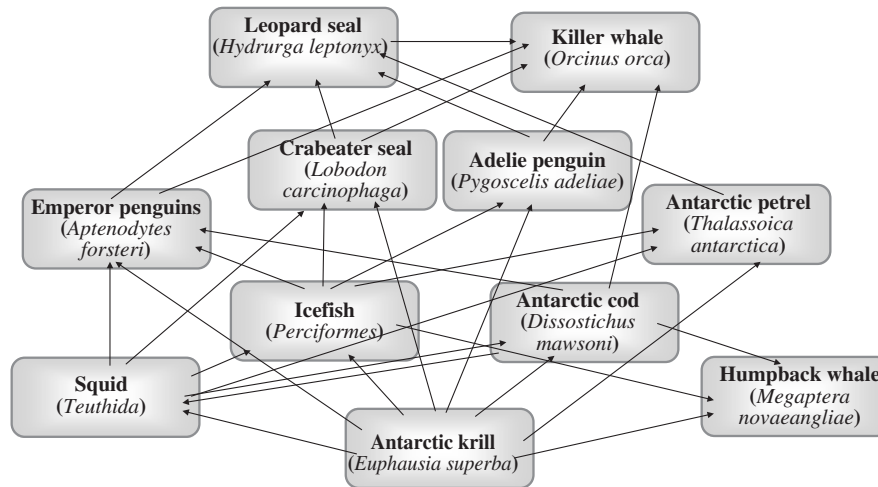
Consider two sets A and B containing n and m members respectively. We can create as many as $n \times m$ pairs from these two sets, implying a total of $n \times m$ possible associations. Any collection of such associations is called a **relation**.

Example 1.46: Food webs



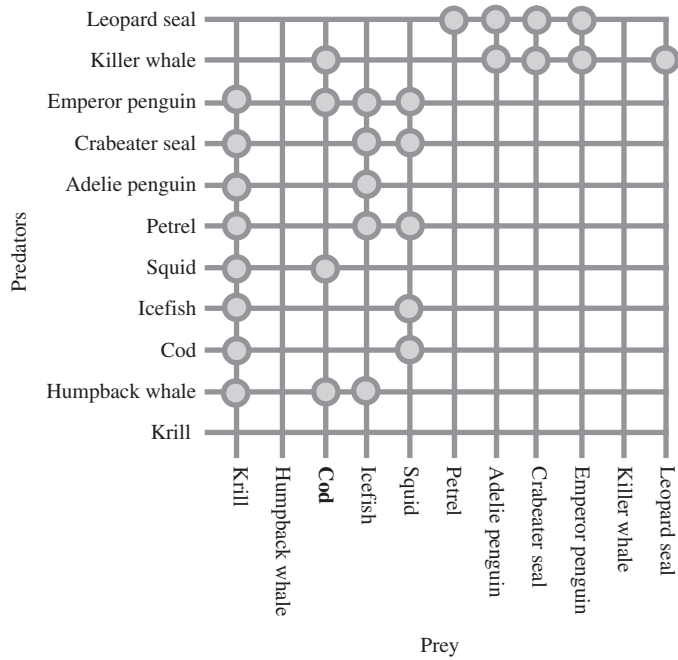
Adélie penguin (Pygoscelis adeliae)

Trophic interactions in ecological communities relate predators to their prey. If A is the set of predator species and B the set of prey species, we may be interested in the linkages between members of A and B . In most ecological communities, some species may act as both predators and prey. Therefore, with no loss of generality, we may assume that the sets A and B have the same membership. However, generally not all predators are also prey and not all prey are also predators. The food web can be illustrated as a network of interconnected nodes (Figure 1.11(a)), or as a regular grid of all conceivable associations (Figure 1.11(b)).



(a)

Figure 1.11: (a) A food-web representation of part of an Antarctic marine community and (b) the corresponding graph of associations.



(b)

Figure 1.11: continued

In Figure 1.11(b), it is clear that a species can eat more than one prey and be eaten by more than one predator. Other types of relations may present simpler associations. One particularly simple case is when every member of the set A is associated with exactly one member of the set B ; such a relation is called a **function**.

Example 1.47: Mating systems in animals



The set of reproductively active males and females in a sexually reproducing species can be associated in any one of four mating schemes (monogamous, polygynous, polyandrous, promiscuous). Figure 1.12 shows examples of associations between females and males. Although all of them are relations, only two are functions, because in the other two more than one arrow leaves certain members of the set of females. If we were associating males with females, then the situation would change: the polygynous system would cease to be a function and the polyandrous system would become one.

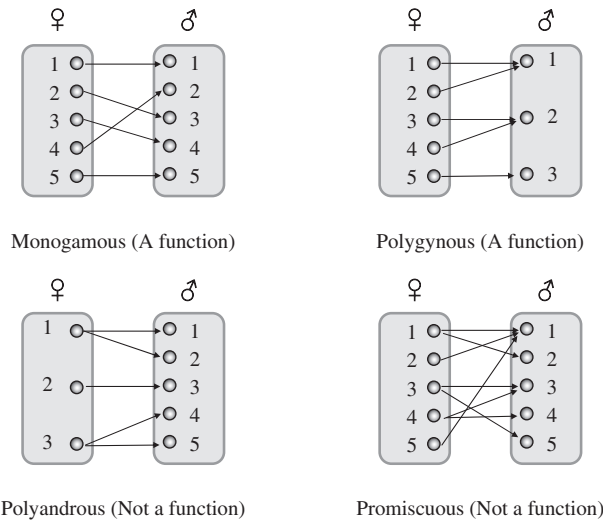


Figure 1.12: Four types of animal mating system represented as associations between females and males.

Algebraic operations can be used on numbers and symbols to create algebraic expressions. The variables contained in these expressions can be as widely or as narrowly defined as we require. We envisage a set of values that such a variable can take and call this the **domain** of the variable.

Example 1.48

Consider the expression $5x + 7$ and let the set $A = \{1, 2, 3, 4\}$ be the domain of the variable x . Now imagine that we pick any value from the domain of the variable x (say, 1) and calculate our algebraic expression (to get 12). If we repeat this process for all the other values, we notice that the expression $5x + 7$ transforms the set A to a new set $B = \{12, 17, 22, 27\}$.

In mathematical terminology we say that the expression in Example 1.48 **maps** the set A onto the set B . The term ‘map’ originates from cartography where people would identify points in real space as points on a map. The set generated by operating on the function’s domain is called the function’s **range**.

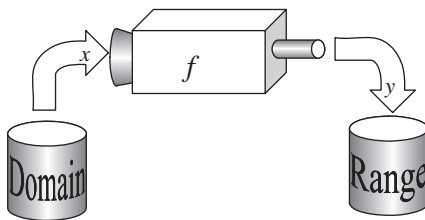


Figure 1.13: A function f operates on an input x , selected from the function’s domain. This operation produces an output y that belongs to the function’s range.

Like most things in mathematics, functions can be given symbolic names (conventionally letters like f or g). The input of the function is called an **independent variable**. It is convenient to think of the function as **operating** on the independent variable (Figure 1.13). For a function f with independent variable x , the operation is denoted by $f(x)$. The result of the operation $f(x)$ generates an output, say y . The function's output is also a variable whose value depends on the choice of the input value x , so it is called the **dependent variable**.

1.13: Defining computer functions

In computing, the term 'function' has a somewhat broader definition than in mathematics, although computer functions can be used to encode mathematical functions. In general, a computer function is a self-consistent piece of code that performs a particular task. Just like a mathematical function, it usually requires some input, upon which it operates to produce an output. However, unlike mathematical functions, the output generated by a particular input does not have to be unique. Four components are needed to fully specify a computer function:

- ① its name
- ② a temporary name for its input (to be used internally by the function);
- ③ a description of how the function performs its task (sometimes called its 'main body');
- ④ a command to return the function's output.

The syntax looks a bit like this:

```
function.name<-function(input)
{
  # insert function's main body here
  return(output)
}
```

The main body of the function may comprise several lines of code. As we will see below, the ability to replace several lines by a `function.name` is the greatest advantage of 'functional' computing. To illustrate, consider the simple task described in Example 1.48. The easiest way to perform this is by first entering a value for the input and then operating on it directly:

```
> x<-4
> y<-5*x+7
> y
[1] 27
```

For such a simple task, this approach is perfectly acceptable. However, more complex tasks, which will need to be performed several times for different inputs, are best encoded as functions. To achieve the above task functionally, type the following in your text editor and then paste it into R:

```
f<-function(x)
{
  y<-5*x + 7
  return(y)
}
```

No output is produced because no specific input has been provided. R knows that it should expect an input (here called x) and it knows what to do with that input once received, but it is waiting for the function to be 'called' by typing its name and specifying the input (e.g. 4):

```
> f(4)
[1] 27
```

The entire mapping of Example 1.48 can be generated as

```
> f(1:4)
[1] 12 17 22 27
```

1.19. The graph of a function

The input x and output y of a function are an ordered pair of coordinates that can be visualised on the plane. Every number from the function's domain generates a new ordered pair and hence a new point on the plane. Sweeping through the entire domain yields a collection of points called the **graph of the function**.

Example 1.49: Two aspects of vole energetics



In Example 1.19, an equation was derived for energy acquisition in voles in terms of physical quantities such as territory size (A) and the proportion of a day spent foraging (p)

$$\Delta E = \epsilon mpA - (1 + 0.2p)E_r \tag{1.88}$$

To see how changes in these quantities affect ΔE we can interpret Equation (1.91) as a function with dependent variable ΔE and independent variables A and p . Calculating the value of ΔE for 20 example values of A and p gives two graphs (Figure 1.14) that hint at linear responses.

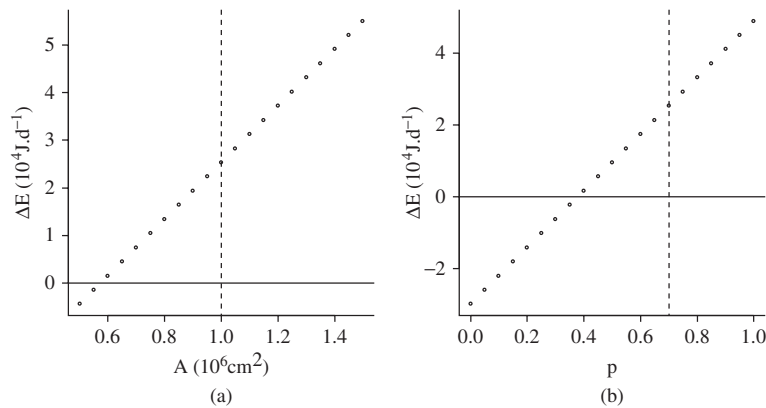


Figure 1.14: Plotting combinations of the independent and dependent variables can help visualise how changes in one affect the other. The above is a graphical representation of the common-sense fact that energy gain will increase if the vole exploits a greater home range with the same amount of effort (a), or if it increases its daily foraging effort but maintains the same home range (b). Note that, at low values of the independent variables, the vole will suffer net energy losses (signified by negative values of the dependent variable). The dashed lines indicate the fixed values of the independent variables used in Example 1.19.



1.14: Plotting the graphs of functions

R offers an overwhelming variety of options for customising graphs. These allow the production of publication-quality figures but require considerable input from the user. However, plotting a simple function isn't too challenging. The bare minimum required is to generate a set of paired data for the independent and dependent variables, plot them and label the axes in a meaningful way. Here is an example for the function $f(x) = 5x + 7$, using the integers between -10 and 10 as the domain.

```
x<-seq(-10,10,by=1)
y<-5*x+7
plot(x,y)
```

The first line introduces the new command `seq()` that is here used to produce a sequence of values from -10 to 10 in steps of 1 . The second line takes this entire list (now called `x`) and applies to it the function. The third line plots the data. You will notice that R automatically uses the names of the two variables to label the axes. If you want to specify your own labels you can do it as follows:

```
plot(x,y, xlab="My own x label", ylab="My own y label")
```

There is a subtle difference between the true graph of a function and that generated by a computer because many functions have infinite or noncountable domains. Computers always plot functions using only a finite number of values from a function's domain. The resulting output is therefore only a discrete approximation of the true graph. To give the impression of continuity, the dots can be joined up by asking for a line plot via the `type` command:

```
plot(x,y, type="l")
```

However, it is strictly not certain that the graph follows a straight line between any two successive dots. Much of this uncertainty can be dispelled by knowing roughly what the graph of a function should look like. This is the topic of the next section.

1.20. First order polynomial functions

Polynomials can be used to construct functions of the general form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \quad (1.89)$$

The simplest polynomial function is

$$f(x) = b \quad (1.90)$$

for $x \in \mathbb{R}$ and for some constant b . The graph of such a constant function is a straight line parallel to the x -axis. The line meets the y -axis at b . Technically, Equation (1.90) is a function of x whose domain includes all the real numbers. However, its output and range are limited to a single value: b .

Example 1.50: Population stability in a time series

Let $N(t)$ describe the number of lizards in an island population as a function of time (t , measured in days). We know that at time $t = 3.2$, the population $N(3.2)$ was 503 and that during the interval $3.2 \leq t \leq 9.2$, no lizards were born and none died. We can express this mathematically using the constant function: $N(t) = 503, \forall t \in [3.2, 9.2]$.

The second simplest polynomial function, the **identity function**, maps its domain exactly onto its range.

$$f(x) = x \quad (1.91)$$

Its graph is a straight line that goes through the origin (the point $0,0$) because $f(0) = 0$, and it forms a 45° angle with the x -axis. We can interpret such a function as having no impact on its input.

Example 1.51: Population stability and population change

If we use a unit of one day to measure time, then we may be interested in plotting how today's population, $N(\text{today})$, determines tomorrow's population, $N(\text{tomorrow})$. We can simplify this notation by using the symbols N_t and N_{t+1} . Future populations depend on current populations so we can envisage a function that describes this transition for any current population size. The independent and dependent variables of such a function are the population size

at time t and the population size at $t + 1$

$$N_{t+1} = f(N_t) \quad (1.92)$$

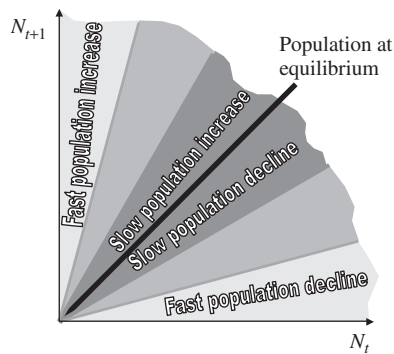


Figure 1.15: A population remains constant if its size is the same from any time instant to the next ($N_{t+1} = N_t$). This situation is represented by the identity function, the graph of which (labelled 'Population at equilibrium') splits the plane into two parts. The lower part corresponds to population decline ($N_{t+1} < N_t$) and the upper part to population increase ($N_{t+1} > N_t$).

In Chapter 3 we will learn how to construct mathematical functions that describe such transitions. For now, it is useful to be able to interpret the biological meaning of different regions of a graph of this function (Figure 1.15). It is particularly important to recognise that if $N_t = f(N_t)$ (compare this with Equation (1.91)), the input and output of the function are identical and population size remains constant.

Example 1.52: Visualising goodness-of-fit



Scientific rigour should not necessarily lead to boring science. Contrary to popular belief, good scientists are not averse to leaps of faith. They do, however, need to put these wild guesses to the test. The best way to challenge a hypothesis is to use it to make predictions. Mathematical modelling simply extends our scientific reach by enabling us to examine the quantitative consequences of ever-more-complicated hypotheses (so, in a sense, modelling leads to more interesting science...). To do this, we need to know how

well our predictions fit the data (not the other way around!). As you work your way through this book, you will see increasingly elaborate ways of doing this. However, when predictions involve a single variable, there is a simple graphical method using the identity function.

Consider a model that aims to predict the density of a plant species in different parts of its range, using several types of information (soil characteristics, climate characteristics, density of grazers, density of competitors, etc.). For each set of observed conditions, we can obtain a single prediction of plant density. For the same set of conditions, we can provide a corresponding measurement of observed density at the same scale. Inspecting the cloud of paired values in a plot can be immensely informative about the quality of a model (Figure 1.16).

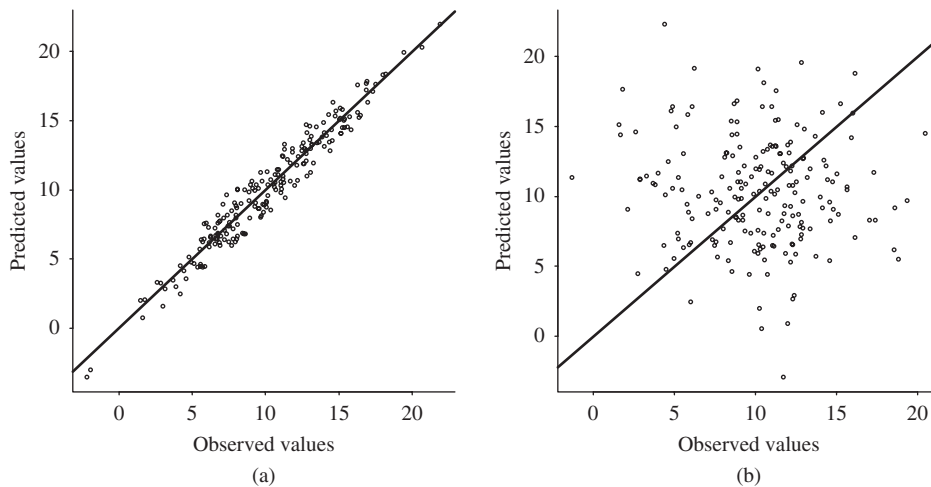


Figure 1.16: (a) The holy grail of every modeller who works with real data; (b) a more typical graph.

We can easily convert the identity function to a scaling relationship (see Section 1.12) by multiplying the right-hand side of Equation (1.91) by a constant a , to get

$$f(x) = ax \quad (1.93)$$

The graph of Equation (1.93) is a straight line that goes through the origin, but the angle it forms with the x -axis depends on the value of a . Adding another constant (b) to Equation (1.93) (i.e., combining Equations (1.90) and (1.93)) gives

$$f(x) = ax + b \quad (1.94)$$

This function also has a straight-line graph. Note that if we set $a = 0$ we recover Equation (1.90); setting $b = 0$ we recover the scaling function in Equation (1.93); and setting $a = 1$ and $b = 0$ we recover the identity function (1.91). In short, Equation (1.94) is the most general function with a linear graph, we will call it a **linear function**. Two important pieces of information about Equation (1.94) and its linear graph are that the constant a determines the **slope** (the inclination) of the line while the constant b is its **intercept** with the vertical (y) axis. More about slopes in Chapter 4.

1.15: Adding lines to plots

Graphs of linear functions can be generated using the techniques of R.1.14. However, you may want to add further lines to a plot, perhaps to indicate features such as goodness of fit, or reference axes. The command `abline()` can be used for this purpose in three ways:

```
abline(b, a)  draws a line of slope a and intercept b
abline(h=y)  draws a horizontal line that intercepts the y-axis at the value y
abline(v=x)  draws a vertical line that meets the x-axis at the value x
```

For example, the following code first plots the graph of the function $f(x) = 1.6x + 2$ (lines 1–3) and then adds the Cartesian axes (lines 4–5) and the graph of the identity function (line 6).

```
x<-seq(-10,10)
f<-1.6*x+2
plot(x,f,type="l",xlab="x",ylab="f(x)",xlim=c(-8,8),ylim=c(-8,8))
abline(v=0)    # y-axis
abline(h=0)    # x-axis
abline(0,1, lty=2) # Identity function
```

Options `xlim` and `ylim` in line 3 specify the ranges of values to be plotted in the x and y axes, and the option `type = "l"` indicates that a line graph is required. The option `lty = 2` in line 6 specifies a dashed line. Line styles can be specified by name: "solid", "dashed", "dotted", "dotdash", "longdash" and "twodash". The numerical codes 1, 2, 3, 4, 5 and 6 can also be used in place of these named options.

1.21. Higher order polynomial functions

The simplest second order polynomial function is

$$f(x) = x^2 \quad (1.95)$$

It is a bit harder to imagine what the graph of Equation (1.95) looks like, but trying out some values for x should help (Figure 1.17). This characteristic cup-like shape is called a **parabola** and it is common to the graphs of all second order polynomial functions.

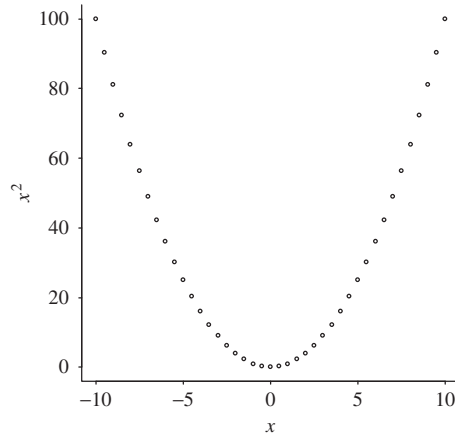


Figure 1.17: This graph was generated by trying out all values of x between -10 and 10 , at intervals of 0.5 .

Let's examine the effect on the graph of introducing a coefficient to the second order term:

$$f(x) = ax^2 \quad (1.96)$$

We can treat Equation (1.95) as a special case of Equation (1.96), obtained by setting $a = 1$. To explore the effect of other values of a we can use values smaller or greater than 1.

Example 1.53

Figure 1.18 shows the graphs of Equation (1.96) with six different values (± 0.5 , ± 1 and ± 2) for the parameter a . The absolute value of the coefficient regulates the spread of the parabola (the closer to zero it is, the wider the parabola) and its sign determines the parabola's orientation (a negative coefficient leads to a downward-pointing cup).

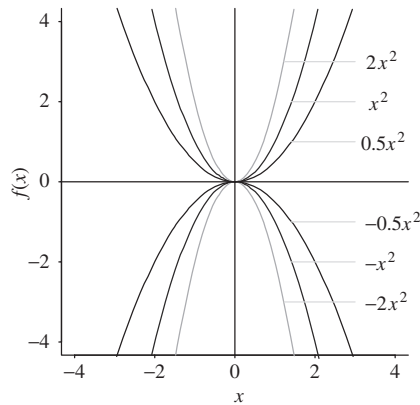


Figure 1.18: The graphs of six different versions of the function $f(x) = ax^2$.

The graph of the more general polynomials $f(x) = a_2x^2 + a_1x + a_0$ is still a parabola that intercepts the y -axis at a_0 . The term a_1x shifts the graph horizontally.

Example 1.54

To illustrate the effect of the sign of different polynomial coefficients, we look at an example $f(x) = a_2x^2 + a_1x + a_0$ with $|a_2| = 0.1$, $|a_1| = 0.4$ and $|a_0| = 1$ (Figure 1.19).

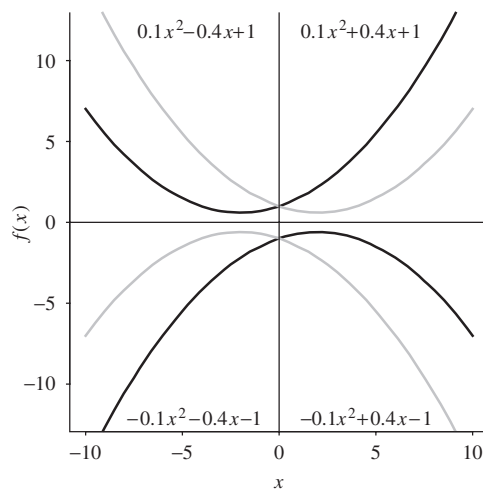


Figure 1.19: The effect of the signs of different polynomial terms.

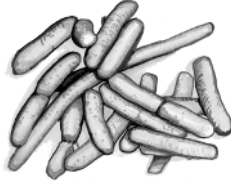
As with the simpler second order polynomials in Example 1.53, the sign of a_2 determines the orientation of the parabola in relation to the x -axis (if it's positive, the parabola is upward-pointing). The sign of a_1 determines the position of the parabola's peak (or trough) in relation to the y -axis (if it's positive, the peak/trough is to the left of the y -axis). Finally, the sign of a_0 determines whether the intercept will be above/below the x -axis.

1.22. The relationship between equations and functions

In Section 1.19 we saw how an equation could be turned into a function by allowing a constant to become a variable. The converse is also useful. To plot the graph of a function, we repeatedly fix its independent variable to a particular value and then calculate the dependent variable. Similarly, we can fix the dependent variable and try to solve the resulting equation for the independent variable. The most frequent use of this trick is to decide where the graph of a function will cross the x -axis. For any function $f(x)$ this is equivalent to solving the equation $f(x) = 0$.

Example 1.55

To find if and where the graphs of the three quadratic functions $f(x) = x^2 - 2$, $g(x) = x^2$ and $h(x) = x^2 + 2$ cross the x -axis, we set each function to zero and attempt to solve the resulting equation. From this, we find that $f(x)$ crosses the x -axis at $\pm\sqrt{2}$ (the equation has two solutions). The graph of $g(x)$ doesn't cross the x -axis but it touches it at 0 (the equation has one solution). The graph of $h(x)$ is an upward-pointing parabola with y -intercept equal to 2, so it can't cross the x -axis (indicated by the fact that the equation $x^2 + 2 = 0$ has no real solutions).

Example 1.56: Extent of an epidemic when the transmission rate exceeds a critical value

Consider a situation in which the rate of transmission of a disease in a population of K animals exceeds a critical value γ_c . We want to find what this implies for the total number N of infected individuals. Assuming perfect mixing, we can rewrite Equation (1.34) in Example 1.21 as an inequality:

$$cbN(K - N) > \gamma_c \quad (1.97)$$

where N are the infected animals. Since $cb > 0$, this can be manipulated into

$$-N^2 + KN - \frac{\gamma_c}{cb} > 0 \quad (1.98)$$

The graph of the function $f(N) = -N^2 + KN - \gamma_c/cb$ (Figure 1.20) can help visualise the range of values of N over which this might occur. In the particular example seen in Figure 1.20 we can identify that, for the given transmission threshold, the population size of infected individuals must lie in the range $N_1 < N < N_2$. These values can be found from the quadratic formula

$$N_1, N_2 = \frac{K \pm \sqrt{K^2 - 4\gamma_c/cb}}{2} \quad (1.99)$$

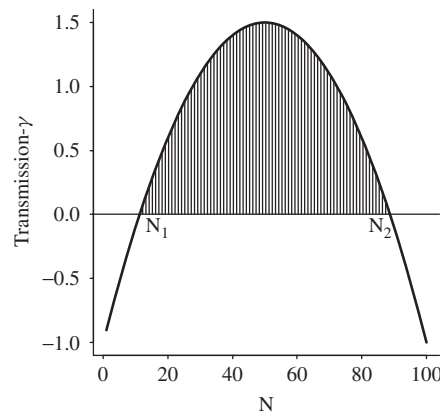


Figure 1.20: The graph of the function $f(N) = -N^2 + KN - \gamma_c/cb$, which gives the difference between actual transmission rate and the critical value of interest γ_c as a function of infected population size. In the shaded region the transmission rate exceeds the critical value γ_c . In this example, the parameter values are $K = 100$, $cb = 0.001$ and $\gamma_c = 1$.

1.23. Other useful functions

My initial focus on polynomials was entirely due to their flexibility and ease of use. However, they are not suited to describing certain ecological relationships. Chapter 2 discusses the fact that seasonal phenomena require a different class of (trigonometric) functions, and in Chapter 3, population growth gives rise to yet others (exponential and logarithmic). This section aims to broaden your concept of functions beyond polynomials by illustrating how we can use the basic algebraic operators to expand our toolbox of functions.

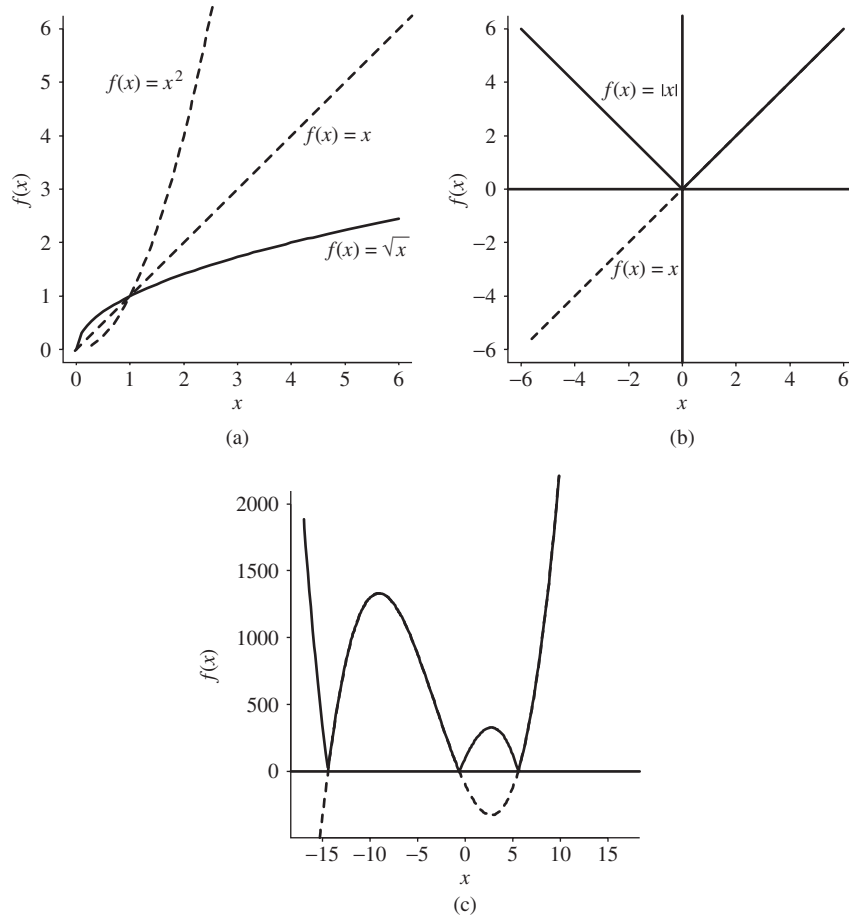
Example 1.57

Figure 1.21: (a) The graph of $f(x) = \sqrt{x}$ compared with the graphs of $f(x) = x$ and $f(x) = x^2$ (dashed curves); (b) the graph of $f(x) = |x|$; (c) the graph of $f(x) = |2x^3 + 19x^2 - 150x - 100|$, the dashed curve shows the segments of the original graph that have been reflected upwards by taking its absolute value.

Roots and absolute values (see Section 1.5) can be used to construct functions. The simplest ones are $f(x) = x^{1/a}$ and $f(x) = |x|$. You may remember that functions using roots are not real-valued if the expression under the root takes negative values. So, the function $f(x) = x^{1/a}$ must take (and also give) non-negative values. Its graph increases sharply and then slows down (Figure 1.21(a)), although it never levels off. The function shows an interesting symmetry with the function $f(x) = x^a$ around the 45° line (Figure 1.21(a)), and the two functions cross at $x = 1$. The graph of the absolute value of a function resembles the graph of the function itself (see Figure 1.21(b) and 1.21(c)). However, values that fall below the x -axis are reflected upwards in the graph of its absolute value.

One particular biological phenomenon that cannot easily be described using a polynomial is ‘plateauing’ or saturation behaviour. For example, there is usually an upper limit on the number of prey items that an individual predator can take per day, no matter how many prey are available to it. There are several different formulations that will achieve this effect; one of the least elaborate is given in Example 1.58.

Example 1.58

A useful function for modelling saturation is the **rectangular hyperbola**,

$$f(x) = \frac{ax}{b+x} \quad (1.100)$$

For most biological applications of saturation, the independent variable (e.g. prey density, enzyme or pollutant concentration) can only take non-negative values. The function starts from the origin ($f(0) = 0$) and increases with x . It approaches, but never attains, a value known as the function’s **asymptote** (a Greek word meaning ‘uncrossable’), equal to the constant a . Because the asymptote is never attained, the value of half-saturation ($f(x) = a/2$) is used to describe how fast the value of the function increases. Half-saturation occurs when $x = b$ (check this by calculating $f(b)$ in Equation (1.100)). The properties of Equation (1.100) are summarised graphically in Figure 1.22. The rectangular hyperbola appears throughout the biological literature, often with a different name. For example, in foraging ecology (see Example 4.14 in Chapter 4), it is called a **Holling type II, or hyperbolic, functional response**, in enzyme kinetics it is known as the **Michaelis–Menten equation** and in medical, behavioural and epidemiological applications, it is a type of **dose–response curve**.

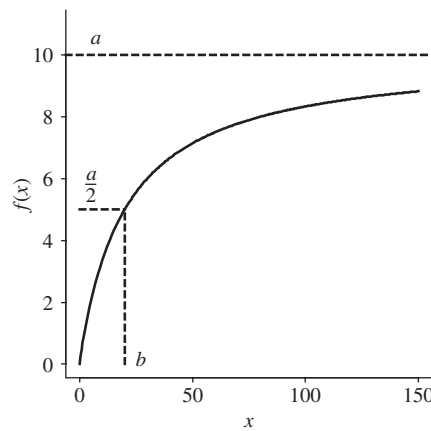


Figure 1.22: The rectangular hyperbola with $a = 10$ and $b = 20$.

1.24. Inverse functions

In Section 1.18, functions were described as processing machines $f(x) = y$. The operation of the function f on x gives the output y . However, given the output of a known function, is it possible to guess what the input was? The answer is yes, if each output in the function’s range can only be produced by using a single input x from the function’s domain. These are called **one-to-one** functions.

Example 1.59

We are given the function

$$f(x) = 2x - 4 \tag{1.101}$$

and three of its outputs: $y_1 = 2, y_2 = -4, y_3 = -1$. To find the corresponding inputs x_1, x_2, x_3 we must solve the general equation $2x - 4 = y$ for x ,

$$x = \frac{y + 4}{2} \tag{1.102}$$

If we treat this as a function of y and we denote it by f^{-1} we get

$$f^{-1}(y) = \frac{y + 4}{2} = x \tag{1.103}$$

Putting in the values of y_1, y_2, y_3 we get $x_1 = 3, x_2 = 0, x_3 = 1.5$. We have thus **inverted** the effect of the function f by introducing the new function f^{-1} (Figure 1.23).

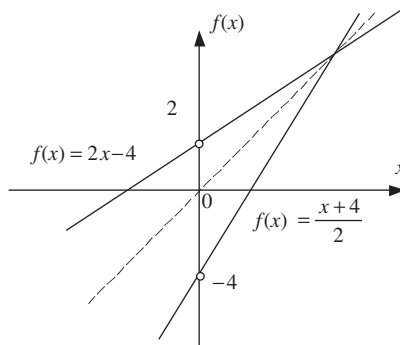


Figure 1.23: Graphs of a simple linear function and its inverse.

Such inverses can be defined for all one-to-one functions (Figure 1.24). Hence, Chapter 2 will introduce the inverses of trigonometric functions (within a constrained part of their range) and Chapter 3 will present the logarithm as the inverse of the exponential function.

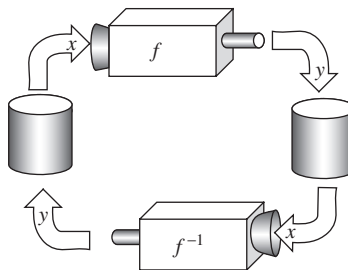


Figure 1.24: We can now complete Figure 1.12. What serves as the domain of f is the range of f^{-1} and vice versa.

1.25. Functions of more than one variable

A function need not be constrained to just one input. Most ecological phenomena are affected simultaneously by many influences and it would be very limiting if we could only examine them one at a time. The notation is similar to what we have encountered before: we need a name for the function (say, f) and names for its independent variables (we can use heuristic symbols as in Example 1.60 below, or indexing, x_1, x_2, \dots). The simplest function in two independent variables is linear

$$f(x_1, x_2) = a + bx_1 + cx_2 \quad (1.104)$$

The graphs of such functions are constrained by the three dimensions of physical space but functions of two independent variables can be readily represented in three axes. The graph of Equation (1.104) is a plane (Figure 1.25(a)). Nonlinear functions give more complicated surfaces (Figure 1.25(b)).

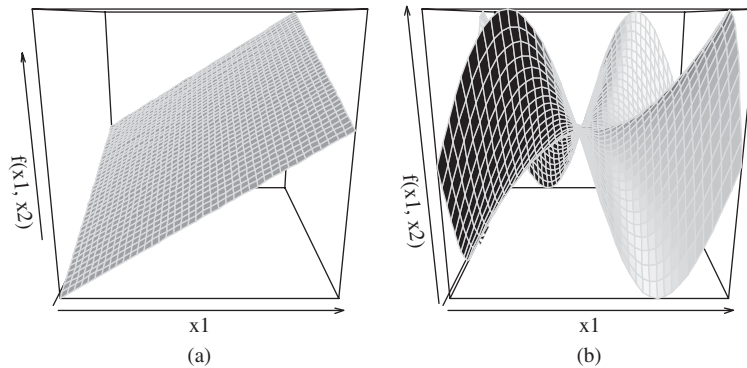


Figure 1.25: (a) The graph of the simple linear function $f(x_1, x_2) = -10 + 3x_1 + 5x_2$; (b) the graph of the nonlinear function $f(x_1, x_2) = x_1x_2(x_1^2 - x_2^2)/(x_1^2 + x_2^2)$.

Example 1.60: Two aspects of vole energetics



each yielding a value for ΔE under the model

$$\Delta E = \varepsilon mpA - (1 + 0.2p)E_r \quad (1.105)$$

An overview of the behaviour of Equation (1.105) can be obtained by plotting it (Figure 1.26). This reveals new aspects of the model's behaviour. Compared to Figure 1.13, which gave the impression of linearity, the plots in Figure 1.26 show how the two independent variables can interact to give higher energy gains for active animals with large territories.

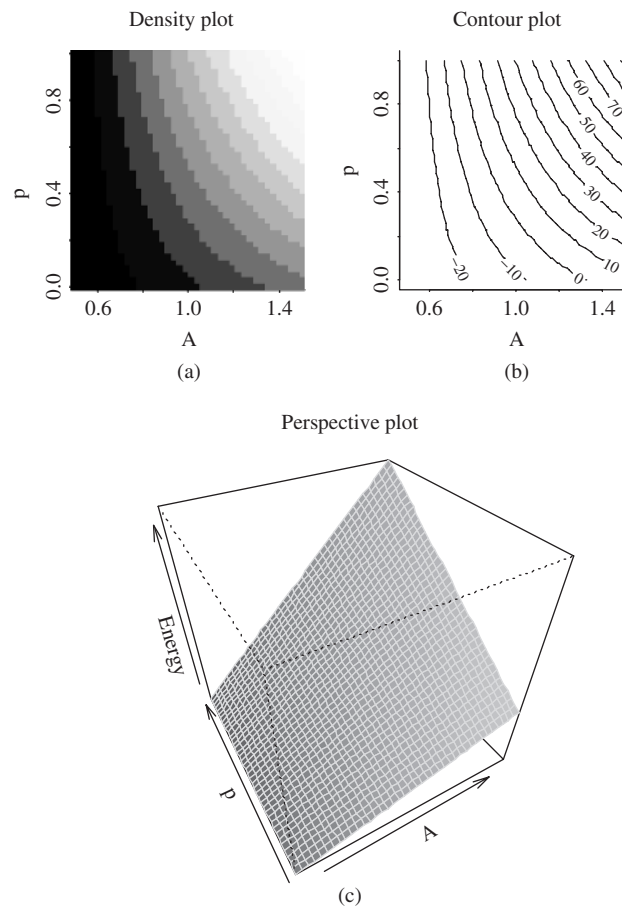


Figure 1.26: Three different ways to plot a function with two independent variables. The first two, (a) and (b), are plotted in two dimensions and use implicit representations of the third. In the case of the density plot (a), the third dimension is represented by shading or colour. A contour plot (b) uses isopleths for this purpose. A perspective plot (c) tries to emulate three dimensions.



1.16: Perspective, image and contour plots

Generating 3D plots requires some more preparation than 2D ones because the function of interest needs to be calculated for values on a grid. The first step, therefore, is to prepare the grid by specifying the plotting range and number of subdivisions to be used for each of the two independent variables:

```
#Plotting parameters
steps<-5 # Number of points in each dimension of the grid
x1r<-c(1,5) # Plotting range for 1st indep. variable
x2r<-c(1,5) # Plotting range for 2nd indep. variable
```

These are used to generate the labels for the two axes. The command `seq()` that we previously saw in R1.14, is here used with the option `len`, which specifies the total desired number of members in the sequence.

```
x1lab<-seq(x1r[1], x1r[2], len=steps) # Labels for x1
x2lab<-seq(x2r[1], x2r[2], len=steps) # Labels for x2
```

We now need to construct *all* pair-wise combinations of values of `x1` and `x2` on the grid. One way to achieve this is, again, by use of `seq()`

```
# Repeats the entire sequence for 1st variable a number of times
# equal to steps
x1<-rep(x1lab, steps)

# Repeats each value of the 2nd variable a number of times
# equal to steps
x2<-rep(x2lab, 1, each=steps)
```

In this example, `x1` and `x2` are:

```
> x1
[1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
> x2
[1] 1 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
```

We are now ready to calculate the function and then arrange the values in a tabular form using the command `matrix()` (more on matrices in Chapter 6).

```
fxy<- x1*x2*(x1^2-x2^2)/(x1^2+x2^2) # Type your own function here
f<-matrix(fxy,steps,steps,byrow=T) # Matrix of values for plotting
```

A perspective plot can now be generated with a single line

```
persp(x1lab,x2lab,f, xlab="x1", ylab="x2", zlab="f(x1,x2)")
```

A density plot (command `image()`) and a contour plot can be generated as follows:

```
image(x1lab,x2lab,f, xlab="x1", ylab="x2")
contour(x1lab,x2lab,f, xlab="x1", ylab="x2")
```

A combination between these two plots is the `filled.contour()` which colours the intervals between the contours. These four plots can be endlessly customised according to taste. The help files in R provide more information on graphing options.

Further reading

Most of the mathematical concepts covered in this chapter are also covered in Cann (2003, pp. 1–34), Foster (1998, pp. 3–48, 59–72), Harris *et al.* (2005, pp. 2–46) and Mackenzie (2005, pp. 1–9, 25–28, 46–53). The presentation is briefer with less relevance to ecology but the mathematical problems will provide additional practice. A slightly more old-fashioned, but very thoughtful, treatment of this chapter's topics is given by Batschelet (1979, pp. 1–14, 17–35, 59–109). Extensive coverage of scaling in ecology can be found in Schneider (2009).

References

- Aarts, G., Mackenzie, M.L., McConnell, B.J., Fedak, M.A. and Matthiopoulos, J. (2008) Estimating space use and environmental preference from wildlife telemetry data. *Ecography*, **31**, 140–160.
- Batschelet, E. (1979) *Introduction to Mathematics for Life Scientists*. Springer, Berlin.
- Cann, A.J. (2003) *Maths from Scratch for Biologists*. John Wiley & Sons, Ltd, Chichester.
- Chartier, T. (2005) Devastating roundoff error. *Math. Horizons*, **13**, 11.
- Foster, P.C. (1998) *Easy Mathematics for Biologists*. Harwood Academic Publishers, Amsterdam.
- Harris, M., Taylor, G. and Taylor, J. (2005) *CatchUp Maths & Stats For the Life and Medical Sciences*. Scion, Kent.
- Heesterbeek, H. (2005) The law of mass-action in epidemiology: A historical perspective. In *Ecological Paradigms Lost: Routes of Theory Change* (Eds K. Cuddington and B. Beisner). Elsevier, Amsterdam, pp 81–106.
- Mackenzie, A. (2005) *Instant Notes: Mathematics and Statistics for Life Scientists*. Taylor and Francis, New York.
- Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P. (2002) *Resource Selection by Animals*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 221 pp.
- Pearl, R. (1928) *The Rate of Living*. University of London Press, UK
- Schneider, D.C. (2009) *Quantitative ecology: Measurement, models and scaling*. Academic press. 432 pp.

