# 2 The key mathematical tools

## Aim of this chapter

**The aim of this chapter is to explain the basic concepts involved in tackling quantitative problems. Much of it is probably already familiar to you, but it is worthwhile to go through it again to consolidate your understanding of the topics. If you wish to go into more depth, the book by Cornish-Bowden (1999) should be consulted.**

## 2.1 Estimation of the results of calculations

### KEY CONCEPTS

- Breaking down calculations involving multiplication and/or division into a series of simple steps
- Making estimates as a check on the results obtained using a calculator

The Dalton (Da) is equal to 1 atomic mass unit (approximately the mass of a hydrogen atom, but more exactly 1/12 of the mass of the $^{12}$C isotope of carbon ($1.66 \times 10^{-24}$ g)); see Chapter 3, Appendix 3.1.

Whenever you face a numerical problem, do you automatically reach for the electronic calculator? Although these wonderful devices are almost universally used for performing calculations, it is a very good idea to develop the habit of trying to estimate the result of a calculation in advance. This gives you a check that you have actually used the calculator correctly.

For example, you might wish to relate the number of amino acids in the polypeptide chain of a protein to its molecular mass. In practice, it has been found that on average each amino acid contributes about 110 Da (Daltons) to the mass of the protein. Thus, if the number of amino acids in the chain were (for example) 260, then the molecular mass would be 260 × 110 Da. Rather than using the calculator, we can easily estimate the result. To a first approximation we merely multiply by 100 to give 26 000 Da, or 26 kDa (kiloDaltons). We can then add 10% of this value, to give 26 + 2.6 kDa, i.e. 28.6 kDa.

The prefixes of units are described in more detail in Chapter 3, section 3.3. k (note small k) is an abbreviation for kilo, i.e. ‘1000 times’; thus 1 kilogram (kg) = 1000 grams (g).

We might wish to proceed in the reverse direction, for example if a protein is of molecular mass 40 kDa, how many amino acids are in the chain? Division of the mass (in Da) by 100 gives 40 000/100 = 400 amino acids; we could then take 10% of this value (40) away to give 360 amino acids (a more accurate answer is 364 amino acids). The importance of being able to relate the molecular mass and the number of amino acids in a protein is explained in Chapter 3, section 3.1.

As a further example of this type of approach it is worth noting that the average contribution of each nucleotide to the mass of a nucleic acid is 330 Da (0.33 kDa). It should thus be relatively easy to see that the mass in kDa can be obtained by dividing the number of bases by 3 (i.e. a synthetic oligonucleotide 80 bases in length has a mass of about 27 kDa).

WORKED EXAMPLE ✓

The genome of the bacterium *Escherichia coli* is a circular DNA molecule with 3.4 million base pairs. What is the molecular mass of this DNA?

**STRATEGY**

This is a relatively simple application of the rule stated above and can be solved without use of a calculator.

**SOLUTION**

The number of bases is $6.8 \times 10^6$; division of this by 3 gives the mass in kDa. The mass is therefore about $2.3 \times 10^6$ kDa; this could also be expressed as $2.3 \times 10^3$ MDa or 2.3 GDa.

The definition of a molar (abbreviated M) solution is given in Chapter 3, section 3.4.1. A 1 M solution contains 1 mole (abbreviated mol; equal to the gram formula weight) of the solute in 1 litre (L) of solution.

WORKED EXAMPLE ✓

The molar concentration of a solution can be obtained by dividing the concentration of the solute expressed in terms of mg $\text{mL}^{-1}$ (equivalent to g $\text{L}^{-1}$) by the molecular mass in Da (see Chapter 3, section 3.4.1). Estimate the molarity of a 3.5 mg $\text{mL}^{-1}$ solution of bovine serum albumin, whose molecular mass is 66 000 Da (66 kDa).

**STRATEGY**

This is an example of estimating the result of a division by a large number; again it is good practice to do this without a calculator.

**SOLUTION**

The molarity = 3.5/66 000 M (M is the abbreviation for molar). Multiply the top and bottom of this division sum by $10^{-5}$ to bring the denominator to a small number (in the region of 1). Hence the molarity = $3.5 \times 10^{-5}/0.66$ M. Since 0.66 goes into 3.5 about five times, the molarity can be estimated as about $5 \times 10^{-5}$ M (50 µM). A more accurate answer is $5.3 \times 10^{-5}$M (53 µM).

When multiplying numbers with powers of 10, add the powers together, e.g. $(2 \times 10^5) \times (3 \times 10^3) = 6 \times 10^{5+3}$, i.e. $6 \times 10^8$. When dividing, the powers are subtracted, e.g. $(8 \times 10^{23})/(4 \times 10^{15}) = 2 \times 10^{23-15}$, i.e. $2 \times 10^8$.

A rather different sort of problem would involve an estimation of the number of heart beats in a human lifetime. You would have to make some assumptions about a typical lifespan (say 80 years) and heart rate (say 70 beats per min).

This would give the number of beats in a lifetime as:

Number of minutes in a lifespan = $80 \times 365 \times 24 \times 60$

Hence, number of beats = $80 \times 365 \times 24 \times 60 \times 70$

We can estimate this by taking out the powers of 10 from each term to leave small numbers that can be easily multiplied together:

Number of beats = $8 \times 3.65 \times 2.4 \times 6 \times 7 \times 10^6$

In this estimation, we have taken 1 power of 10 from 80, 2 from 365, 1 from 24, 1 from 60, and 1 from 70, making 6 powers of 10 overall.

Now we estimate the multiples in pairs (i.e. 8 × 3.65 is about 30; 2.4 × 6 is about 15):

$$\begin{aligned}\text{Number of beats} &\approx 30 \times 15 \times 7 \times 10^6\\ &\approx 30 \times 100 \times 10^6\\ &\approx 3 \times 10^9 \text{ (i.e. three thousand million or three billion)}\end{aligned}$$

If the calculation involved a division, we would go through the same procedure separately for the numerator and the denominator before estimating the final result of the division.

Note the convenient way of representing very large or very small numbers is by use of powers of 10. For example, it is much easier to write that 1 nanometre (nm) = $10^{-9}$ metre (m), rather than 0.000000001 m, or that the speed of light is $2.997 \times 10^8$ m $s^{-1}$ rather than 299 700 000 m $s^{-1}$.

In the heartbeat example, if we were to feed the values into a calculator, we would obtain the result that there were $2.94336 \times 10^9$ heartbeats in a lifespan. However, to state this as the answer would in this case give a completely false impression of the accuracy of the estimate. The assumptions of a lifespan of 80 years and a heart rate of 70 beats per minute are likely to be at best only reasonable approximations, and hence we should be wary about stating anything other than that there are likely to be about 3 thousand million or 3 billion heart beats in a human lifespan.

## ? SELF TEST

**Check that you have mastered the key concepts at the start of the section by attempting the following questions without using a calculator then use the calculator to check your answers.**

**ST 2.1** The molecular mass of the trypsin inhibitor protein from soya bean is 21 kDa. How many amino acids does it contain?

**ST 2.2** The protein hormone insulin contains 51 amino acids. Estimate its molecular mass.

**ST 2.3** Human mitochondrial DNA contains about 16 000 base pairs. Estimate its molecular mass.

**ST 2.4** The human body is estimated to contain $2.5 \times 10^{13}$ red blood cells, each of which contains $2.8 \times 10^8$ molecules of haemoglobin. Each molecule of haemoglobin has four binding sites for oxygen. How many molecules of oxygen can be bound by the haemoglobin in the body? If 1 mole of oxygen contains $6.02 \times 10^{23}$ molecules, how many moles of oxygen does this correspond to?

Answers

ST 2.1 The number of amino acids is estimated as 190 amino acids.

ST 2.2 The molecular mass is estimated as 5600 Da or 5.6 kDa.

ST 2.3 The molecular mass is estimated as 11 000 kDa or 11 MDa.

ST 2.4 The number of molecules is estimated as $3 \times 10^{22}$; the number of moles is estimated as 0.05.

In ST 2.1 and ST 2.2, remember each amino acid contributes about 110 Da to the mass. Assume 100 Da and then make the small (10%) adjustment.

In ST 2.3, remember each base contributes about 0.33 kDa, i.e. three bases contribute 1 kDa. The DNA consists of two strands (base pairs), so we multiply our answer for each strand by 2 to obtain the overall molecular mass.

ST 2.4 is a good example of the manipulation of powers of 10.

## 2.2 Significant figures

KEY CONCEPTS

- Expressing the value of a quantity to the stated number of significant figures
- Understanding the degree of precision appropriate for the experimental approaches employed

The number of significant figures in the quoted value of a quantity is the number of figures ignoring leading or trailing zeroes, ignoring the position of the decimal point; it provides a measure of the confidence with which that value is known. Thus, if the molecular mass of a protein is quoted as 30 kDa (i.e. 30 000 Da), this represents only 1 significant figure; we would be confident that the mass were between 25 and 35 kDa. A different technique might yield the answer to 2 significant figures, e.g. 34 kDa. The technique of mass spectrometry might give an answer of 34.503 kDa; this would represent 5 significant figures. When values are rounded off, 0–4 are rounded down, 5–9 are rounded up. Thus, to 1 significant figure, 34 would be expressed as 30; 35 would be expressed as 40.

Try to develop the skill of quoting the results to the appropriate number of significant figures. This shows that you have understood the basis of the calculation or measurement.

It is extremely important to quote the results of calculations to the appropriate number of significant figures. For example, the molecular mass of a protein can be determined by SDS-PAGE (see Chapter 8, section 8.2.1), in which the mobility of the protein on electrophoresis is compared with standard proteins of known molecular mass. If the mobility is measured to 2 significant figures (e.g. the distance travelled by a band on a gel was 5.2 cm), then the molecular mass should not be quoted to more than 2 significant figures, e.g. 35 kDa, even if the calculator display gives an answer of 34.631782 kDa.

It is very tempting to think that because a calculator gives, for example, 8 places of decimals it must somehow be accurate and authoritative. This is not the case! Of course, it is good practice to carry as much precision as possible forwards during calculations, so long as proper rounding off is performed at the end.

SELF TEST ?

**Check that you have mastered the key concepts at the start of this section by attempting the following question.**

**ST 2.5** A calculator gives the result of a calculation as 4623.708. Express this result to 1, 2, 3, and 4 significant figures.

Answer

ST 2.5 The results are 5000, 4600, 4620, and 4624, respectively.

## 2.3 Logarithms

### KEY CONCEPTS

- Understanding what is meant by the logarithm of a number
- Deriving the values of log $x$, ln $x$, $10^x$, $e^x$ for a given value of $x$
- Understanding the importance of logarithms in analysing biological systems

The logarithm (abbreviated log) of a number $n$ is the power to which the reference base number (usually 10) must be raised to give $n$.

Thus, $10^2 = 100$, so $\log 100 = 2$; similarly $\log 100\,000 = 5$.

The log does not have to be an integer, thus $10^{2.7634} = 580$, so $\log 580 = 2.7634$.

From the rules regarding powers during multiplication and division:

$$10^a \times 10^b = 10^{a+b}, \text{ so } \log(a \times b) = \log a + \log b$$

$$10^a / 10^b = 10^{a-b}, \text{ so } \log(a/b) = \log a - \log b$$

$$a^2 = a \times a, \text{ so } \log(a^2) = \log a + \log a = 2 \log a; \text{ in general } \log(a^n) = n \log a$$

Leonhard Euler was an 18th-century Swiss mathematician who made major contributions to many areas of mathematics, even after he became totally blind. Natural logarithms (to base *e*) arise in the branch of mathematics known as integral calculus (the area under the curve $y = 1/x$ from $x = e$ to $x = 1$ equals 1).

An alternative reference base number for logarithms is the Euler number, *e* (equal to 2.71828 . . . ). Logarithms to base *e* are known as natural logarithms and generally denoted by ln (though you will see $\log_e$ used in some books).

Now ln 10 = 2.303, so in general $\ln x = 2.303 \log x$.

If you enter various numbers into your calculator and use the log and ln keys you should be able to get a feel for the behaviour of logarithms. You will discover the following key properties of logarithms.

! KEY INFO

### KEY PROPERTIES OF LOGARITHMS

- The log of 1 = 0 (this is because $10^0 = 1$)
- The log of a number between 0 and 1 is negative (an example of this is pH, see section 2.3.1)
- The log of a number greater than 1 is positive
- Negative numbers do not have logarithms; if you try to find the log of –4 for example, you will get an error message.

You should also learn to use the anti-logarithm or inverse logarithm keys ($10^x$ and $e^x$ for logarithms to base 10 and *e*, respectively). If you know the logarithm of a number, you can use these keys to evaluate the number. For example, if $x = 4.702$, $10^x = 50\,350.1$, and $e^x = 110.17$.

Because logarithms are expressions of the power to which a number is raised, we can use the log $x$ and $10^x$ functions on the calculator to work out squares, square roots, cubes, and cube roots, etc. of numbers. For example, if we wished to work out the cube root of 983, we take the log of 983 (2.9926), divide this by 3 (0.9975), and then take $10^x$ of this number (9.943). To work out 7.52 cubed ($7.52^3$), take the log of 7.52, multiply it by 3, and then take $10^x$ of the result (to give 425.26). We can also use this approach to deal with non-integral powers of numbers. For example, $7.52^{0.28}$ can be shown to be equal to 1.759. The self-test question **ST 2.8** at the end of this section provides an application of this type of calculation.

The use of logarithms makes it possible to compress what can be a huge numerical range. Some applications of logarithms in biology are described in sections 2.3.1–2.3.6.

## 2.3.1 Acid-base behaviour and the pH scale

Acidity is quantitatively defined by the concentration of protons ($H^+$ ions) present in a solution. The $[H^+]$ varies enormously in living systems. Thus, after a meal, the $[H^+]$ in the stomach is typically about 0.03 M (30 mM), whereas in the duodenum it is around 0.00000001 M ($1 \times 10^{-8}$ M or 10 nM). Inside the lysosome (a subcellular organelle concerned with degradation of macromolecules), the $[H^+]$ is usually 0.00003 M ($3 \times 10^{-5}$ M or 30 μM).

In order to handle this huge range of numbers, the pH scale is used as a measure of acidity. pH is defined by eqn. 2.1:

$$\mathrm{pH} = -\log\,[\mathrm{H}^+] \qquad (2.1)$$

The stomach contains a very strong solution of hydrochloric acid (HCl). This would degrade the stomach wall if it was not protected by a layer of mucus. Ulcers arise if this mucus layer is damaged, e.g. by aspirin or other drugs, excessive alcohol, smoking, etc.

Thus, in the stomach the $\mathrm{pH} = -\log(0.03) = -(-1.52) = 1.52$
In the duodenum, the $\mathrm{pH} = -\log(1 \times 10^{-8}) = -(-8) = 8$
In the lysosome, the $\mathrm{pH} = -\log(3 \times 10^{-5}) = -(-4.52) = 4.52$

As we shall see in Chapter 3, section 3.7, an analogous system is used to denote the strengths of acids, employing the term $\mathrm{p}K_a$ (equal to $-\log K_a$, where $K_a$ is the dissociation constant of the acid).

Another illustration of logarithmic scales to show a very large range of concentrations is the formation plot, used to depict the binding of a drug to a receptor, for example (see Chapter 4, section 4.3.1).

## 2.3.2 Variation of reaction rates with temperature

The rates of reactions increase dramatically with temperature as a greater proportion of the reactants possess the energy necessary to surmount the activation energy barrier for reaction to occur. The equation derived by Arrhenius (eqn. 2.2;

Svante Arrhenius was a Swedish physical chemist who made important contributions in the 19th and 20th centuries to the theory of reaction rates and to understanding the behaviour of ionic solutions.

see Chapter 4, section 4.2.3) to describe the variation of the rate constant of the reaction ($k$) with absolute temperature ($T$) is:

$$k = Ae^{-Ea/RT} \qquad 2.2$$

where $A$ is the pre-exponential factor, $E_a$ is the activation energy for the reaction, and $R$ is the gas constant (8.31 J $K^{-1}$ $mol^{-1}$).

We shall see in section 2.5.2 how eqn. 2.2 can be transformed to plot data conveniently.

### 2.3.3 First-order processes and bacterial growth

The decay of radioactive isotopes or the decrease in the concentration of drugs in the blood plasma normally follow first-order kinetics, according to eqn. 2.3:

$$[A]_t = [A]_0e^{-kt} \qquad 2.3$$

where $[A]_t$ and $[A]_0$ are the concentrations at time $t$ and at zero time, respectively, and $k$ is the rate constant for the reaction. We shall see in section 2.5.2 how eqn. 2.3 can be transformed to plot data conveniently.

Bacteria with a plentiful supply of nutrients will grow in an exponential (or logarithmic) fashion; thus if, say, the generation time (the time for cell growth and division to provide to daughter cells) were 30 min, and we start with 100 cells in a culture, then after 30 min there will be 200 cells, after 60 min, 400 cells, and after 10 h, $1.0486 \times 10^8$ (i.e. $2^{20} \times 100$) cells. At this rate, after 20 h there would be no less than $1.0995 \times 10^{14}$ (i.e. $2^{40} \times 100$) cells! Of course, the culture will eventually run out of nutrients and the numbers will level off. The period of rapid growth is known as the log phase. A plot of the log of the number of cells against time in this phase can be used to determine the generation time ($t_{1/2}$) of the bacterial culture; the slope of this plot is equal to $(\log 2)/t_{1/2}$, i.e. $0.301/t_{1/2}$.

The rapid growth in the number of bacteria makes it very important to try to achieve 100% killing to combat disease. That is why it is advisable to complete a prescribed course of antibiotics, for example.

✓ WORKED EXAMPLE

The number of bacterial cells in a culture (where there is a plentiful supply of nutrients) increases from $1.2 \times 10^5$ to $5.8 \times 10^5$ over 120 min. What is the generation (doubling) time for the bacteria under these conditions?

**STRATEGY**

The two data points can be used to calculate the slope of the plot of log (number of cells) against time. This can be used to calculate $t_{1/2}$.

**SOLUTION**

The slope of the plot is 0.684/120 $min^{-1}$, i.e. 0.0057 $min^{-1}$. Thus, $0.301/t_{1/2} = 0.0057$, from which $t_{1/2} = 0.301/0.0057$ min = 52.8 min.

### 2.3.4 Molecular mass calibration graphs

Molecular masses of proteins are often estimated by the techniques of gel filtration and SDS-PAGE (see Chapter 8, sections 8.2.1 and 8.2.3). The former method is usually carried out under conditions where a protein retains the three-dimensional structure required for activity (i.e. native conditions) and therefore can be used to estimate the mass of the intact protein. The latter is performed under denaturing conditions and almost invariably will yield the mass of the constituent polypeptide chains of the protein. In both cases, the behaviour of protein being analysed is compared with those of standard proteins of known molecular mass, and calibration graphs are constructed. These are log molecular mass vs. elution volume (gel filtration) and log molecular mass vs. mobility (SDS-PAGE).

SDS-PAGE is an abbreviation for sodium dodecylsulphate-polyacrylamide gel electrophoresis. It is a technique that measures the mobility of a protein in an electric field in the presence of SDS which is a detergent. It can be used to give a good estimate of the molecular mass of a protein as well as the degree of purity of a protein preparation.

### 2.3.5 Spectrophotometry

As we shall see in Chapter 3, section 3.6, in many cases measurement of the absorption of light by a solution provides a convenient way of determining its concentration. The quantity measured is known as the absorbance ($A$), which is defined by the equation $A = \log (I_0/I_t)$ where $I_0$ and $I_t$ are the intensities of incident and transmitted light. The logarithmic nature of this relationship has important practical consequences for the accurate determination of concentrations (see Chapter 3, section 3.6).

### 2.3.6 Energy changes and equilibrium constants of reactions

The standard free energy change in a reaction ($\Delta G^0$) is related to the equilibrium constant for the reaction ($K_{eq}$) by eqn. 2.4:

$$\Delta G^0 = -RT \ln K_{eq} \tag{2.4}$$

where $R$ is the gas constant (8.31 J K$^{-1}$ mol$^{-1}$) and $T$ is the absolute temperature.

The nature of this equation means that the value of $K_{eq}$ will change logarithmically with changes in $\Delta G^0$; at 310 K (37°C) each change of about 5.9 kJ mol$^{-1}$ will lead to a 10-fold change in the value of $K_{eq}$. This point is discussed further in Chapter 4, section 4.1.

The free energy change of a reaction under standard state conditions ($\Delta G^0$) is discussed further in Chapter 4, section 4.1. The Greek letter Δ (capital delta) is used to mean 'the change of'; $G$ is the symbol for free energy (denoted as $G$ in honour of Josiah Willard Gibbs, an American 19th-century physical chemist. The superscript zero indicates that the change in free energy is under standard state conditions (see chapter 4, section 4.1).

SELF TEST

**Check that you have mastered the key concepts at the start of this section by attempting the following questions.**

Use a calculator to perform the following calculations.

**ST 2.6** Find the values of log $x$ and ln $x$ for the following values of $x$: 0.018, 0.632, 1.589, 29.97, 8713

**ST 2.7** If log $x$ and ln $x$ have the values −3.72, −1.59, 0.033, 1.15, 4.858, what are the values of $x$?

**ST 2.8** The resting heart rate ($H$ in beats min$^{-1}$) for mammals has been found to vary with body mass ($m$ in kg) according to an empirical relationship $H = 202/(m^{0.25})$. Use this equation to estimate the heart rate for the following animals: elephant (6000 kg), white rhinoceros (2500 kg), lion (220 kg), human (75 kg), domestic cat (5 kg), rat (0.5 kg).

Answers

ST 2.6 The values of log $x$ are −1.745, −0.199, 0.201, 1.477, 3.940, respectively; the values of ln $x$ are −4.017, −0.459, 0.463, 3.400, 9.073, respectively.

ST 2.7 The values of $x$ are: (log $x$) $1.905 \times 10^{-4}$, 0.0257, 1.079, 14.125, 72 111 respectively; the values of $x$ are: (ln $x$) 0.0242, 0.204, 1.034, 3.158, 128.77, respectively.

ST 2.8 The heart rates (beats min$^{-1}$) are: elephant, 23; rhinoceros, 29; lion, 53; human, 69; cat, 135; rat 241.

ST 2.8 provides an application of the use of logarithms to evaluate powers of numbers. Using more extreme examples, we could estimate that the heart rate of a blue whale (100 000 kg) is 11 beats min$^{-1}$ and that of a small shrew (0.003 kg, i.e. 3 g) is 863 beats min$^{-1}$. These values are in line with measured values for these parameters. It is worth noting that the heart of a blue whale is the size of a modest saloon car and the aorta is large enough for an adult human to crawl along!

## 2.4 Reciprocals

KEY CONCEPTS

- Understanding what is meant by the reciprocal of a number
- Using reciprocals to evaluate a number of important parameters such as $V_{\max}$, $K_m$, $K_d$, and $E_a$ from appropriate graphs

The reciprocal of a number is 1 divided by that number; thus the reciprocal of 8 is 0.125 and the reciprocal of 0.02 is 50. Use the $1/x$ button on the calculator to calculate reciprocals and to explore this function.

Calculations of reciprocals are required in a number of situations, for example:

- In the Lineweaver–Burk plot of enzyme kinetic data and the subsequent calculations of the parameters $K_m$ and $V_{\max}$ (see Chapter 4, section 4.4)
- Calculating the $K_m$ or $K_d$ from a the slope of an Eadie–Hofstee or a Scatchard plot, respectively (see Chapter 4, section 4.4)
- Interconverting association and dissociation constants for binding processes (see Chapter 4, section 4.3.1)
- In the Arrhenius plot where the $x$-axis of the plot is $1/T$ ($T$ is the absolute temperature) (see Chapter 4, section 4.2.3).

SELF TEST ?

**Check that you have mastered the key concepts at the start of this section by attempting the following questions.**

Use a calculator to perform the following calculations

**ST 2.9** From a graph, $1/V_{max}$ is found to be 0.0235 min μM$^{-1}$. What is the value of $V_{max}$?

**ST 2.10** From the same graph, $-1/K_m$ is found to be $-0.0065$ μM$^{-1}$. What is the value of $K_m$?

**ST 2.11** The value of $K_a$ for a binding process is $4.53 \times 10^4$ M$^{-1}$. What is the value of $K_d$, given that $K_a = 1/K_d$?

Answers

ST 2.9 The value of $V_{max}$ is 42.6 μM min$^{-1}$.

ST 2.10 The value of $K_m$ is 154 μM; note that the minus signs on each side of the equation cancel out, so that $K_m$ is a positive number.

ST 2.11 The value of $K_d$ is $2.208 \times 10^{-5}$ M, or 22.08 μM.

In ST 2.9–2.11 note that when taking reciprocals, the units are also inverted.

In ST 2.9–2.11 note that the prefix μ (micro: small Greek letter mu) means ‘$10^{-6}$ times’, i.e. 1 μg = $10^{-6}$g.

In ST 2.10 note that $K_m$ is effectively a concentration (see Chapter 4, section 4.3.3), so it must be a positive number.

## 2.5 Testing hypotheses

KEY CONCEPTS

- Understanding the equation $y = mx + c$ for a straight line graph, and being able to derive the slope and intercept of this graph
- Rearranging simple equations into the form $y = mx + c$

Biochemistry and related subjects, e.g. molecular cell biology, aim to provide explanations of the behaviour of biological systems based on physical laws. The aim is to produce a hypothesis or model that can be tested against experimental data. An important aspect of the process is to derive an equation and then test the experimental data against this equation, usually by means of an appropriate plot. Once a model is verified, the equation can be used to predict the outcome of an experiment under a new set of conditions. If the data do not support the model, it may well be necessary to change it to accommodate the data. This section will deal with the way in which we analyse data so as to confirm that they obey proposed models. Section 2.6 will give a brief outline of some important statistical concepts, which allow us to assign the degree of confidence with which we can make such statements.

### 2.5.1 Dependent and independent variables

In a graph, the convention is that the $x$-axis (abscissa) is used to plot the variable that the experimenter varies (e.g. time, concentration of substrate, etc.). This is the *independent variable*.

Guidelines for plotting graphs are given in Chapter 11, section 11.2.9. Most of the points made also apply to graphs generated by computers.

The $y$-axis (ordinate) is used to plot the quantity that is then observed (e.g. concentration of product formed, rate of reaction, etc.). This is the *dependent variable.*

A very important relationship between $y$ and $x$ is given by the equation of a straight line (eqn. 2.5):

$$y = mx + c \tag{2.5}$$

where $m$ is the slope (gradient) of the line and $c$ is the intercept of the line on the $x$-axis.

It is very important when plotting data to make sure that the points on the graph actually correspond to the numerical values of the data points. This is a particular problem with certain computer-based graphics programs such as Excel, which will not automatically plot data points with the correct uniform scale on the $x$-axis; the advice is to look carefully at the plot and see whether it corresponds to what you intend. You should also be able to calculate the value of the slope (change in the value of $y$ divided by the change in value of $x$) and express it in the correct units. Finally, you should be able to look at an equation and recognize what terms could represent the $y$-axis and $x$-axis values, remembering that the slope must be a constant (or a combination of terms that are constant).

Some typical straight line plots which might be obtained are shown in Figs. 2.1–2.3.

**Fig. 2.1** Examples of straight-line graphs through the origin. (a) Concentration of product formed against time in a reaction (straight line through the origin); (b) the rate of reaction against the concentration of enzyme added, where there is no significant rate in the absence of enzyme. The equation for the line is $y = mx$, where $m$ is the gradient.

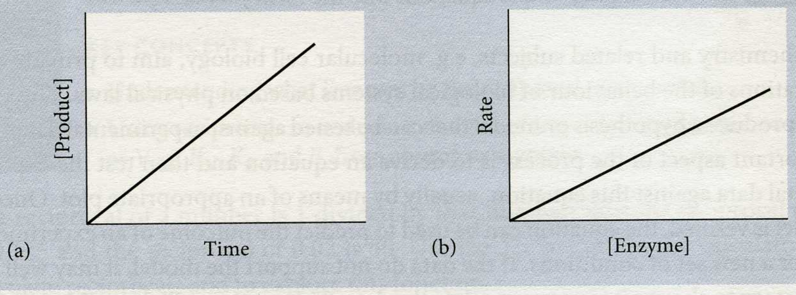
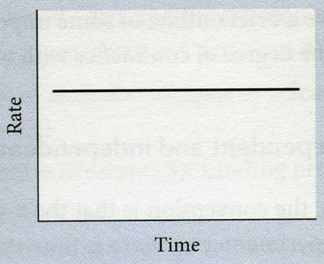


**Fig. 2.2** A straight-line graph of zero slope. Shortly after the start of the reaction, the rate of reaction is constant over the time period studied. The equation for the line is $y = c$, where $c$ is a constant, equal to the intercept on the $y$-axis.



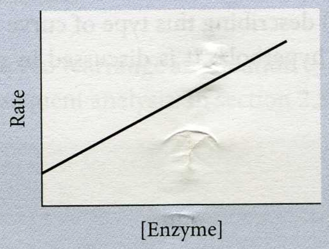


**Fig. 2.3** A straight-line graph with a non-zero intercept on the $y$-axis. The rate of product formation is plotted against the concentration of enzyme for the case where there is a significant blank rate in the absence of enzyme. The equation for the line is $y = mx + c$, where $m$ is the gradient and $c$ is the intercept on the $y$-axis.

Fig. 2.1(a), which depicts the concentration of product formed against time in a reaction, clearly shows a simple straight line relationship, with the equation $y = mx$ ($m$ = slope). This shows that product is being formed at a constant rate; at zero time, no product is present. We would normally expect to see a graph of this type if we plotted the rate of an enzyme-catalysed reaction against the concentration of enzyme added (Fig. 2.1(b)).

Fig. 2.2, which depicts the rate of product formation of the reaction in Fig. 2.1(a) against time, is a straight line of zero slope, i.e. the rate of the reaction is constant over the time period studied. The equation for this line is $y = c$ (i.e. $m = 0$, since there is no dependence on time).

At the instant the reaction starts, the rate is very low; it takes a finite period before the so-called steady-state rate is achieved. The plot in Fig. 2.2 assumes that this 'pre-steady-state' period is very short.

Fig. 2.3, which depicts the rate of a small number of enzyme-catalysed reactions against the concentration of enzyme added, shows a straight line relationship. However, in this case there is still a significant background rate of reaction when no enzyme is present. The equation is $y = mx + c$, where $m$ is the slope and $c$ is the intercept on the $y$-axis. The intercept would correspond to the background (or blank) rate of reaction.

Fig. 2.4, which depicts the rate of reaction of an enzyme-catalysed reaction against the concentration of substrate, is clearly not a straight line. The rate of the reaction shows saturation behaviour with respect to the concentration of

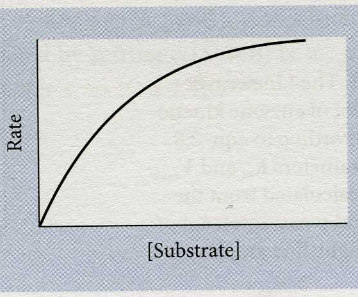


**Fig. 2.4** The dependence of rate on the concentration of substrate for an enzyme-catalysed reaction. The line is a rectangular hyperbola, described by eqn. 2.6.

Mathematicians prefer the use of the term 'limiting value', as it indicates that a limit is approached at high concentrations of substrate. The term 'maximum value' should be used when the value can decline as the substrate concentration is increased further. However, the term $V_{max}$ is very widely used by biochemists, so will be kept here.

substrate, i.e. it tends towards a maximum (more accurately, a limiting) value. The actual equation (eqn. 2.6) describing this type of curve is known mathematically as that for a rectangular hyperbola. It is discussed in more detail in Chapter 4, section 4.3.3.

$$v = \frac{V_{max}[S]}{K_m + [S]} \tag{2.6}$$

Equation 2.6 can be transformed in a number of ways to give a straight line relationship that would allow the validity of the equation to be tested. One of the most commonly used is the Lineweaver–Burk plot. By taking reciprocals of the terms on both sides of eqn. 2.6, we obtain eqn. 2.7:

$$\frac{1}{v} = \frac{K_m + [S]}{V_{max}[S]} \tag{2.7}$$

Dividing each term in the numerator of the right hand side of eqn. 2.7 by $V_{max}$ [S], we obtain eqn. 2.8:

$$\frac{1}{v} = \frac{K_m}{V_{max}}\left(\frac{1}{[S]}\right) + \frac{1}{V_{max}} \tag{2.8}$$

The Lineweaver–Burk plot is very commonly used to analyse enzyme kinetic data, but it should be remembered that the reciprocal nature of the axes makes it subject to a highly non-uniform distribution of errors (see Chapter 4, section 4.4).

Since $K_m$ and $V_{max}$ (and hence also $K_m/V_{max}$) are constants, it follows that eqn. 2.8 is of the form $y = mx + c$, where $y$ is $1/v$, $x$ is $1/[S]$, $m$ is $K_m/V_{max}$, and $c$ is $1/V_{max}$.

A plot of $1/v$ vs. $1/[S]$ is a straight line (Fig. 2.5), known as the Lineweaver–Burk plot, with the $y$-axis intercept $= 1/V_{max}$ (see Chapter 4, section 4.4).

Note that some plots you may obtain in the laboratory, e.g. the response of a dye-binding assay to the amount of protein added (see Chapter 6, section 6.1.1), do not necessarily conform to any simple theoretical equation and would be represented by smooth curves. Appropriate values can then be read off the calibration graphs.

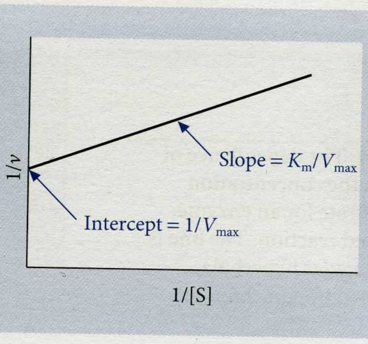


**Fig. 2.5** The Lineweaver–Burk plot of enzyme kinetic data, according to eqn. 2.8. The parameters $K_m$ and $V_{max}$ can be calculated from the slope and $y$-axis intercept of the straight-line graph.

## 2.5.2 Rearranging equations

In many cases, it is necessary to rearrange an equation to put it into a form appropriate for plotting and subsequent analysis. In section 2.5.1, we saw how eqn. 2.6

$$v = \frac{V_{\text{max}}[S]}{K_m + [S]} \qquad \text{2.6}$$

could be transformed (by taking reciprocals of both sides) into one (eqn. 2.8) which would give a straight line, i.e.

$$\frac{1}{v} = \frac{K_m}{V_{\text{max}}}\left(\frac{1}{[S]}\right) + \frac{1}{V_{\text{max}}} \qquad \text{2.8}$$

There are other ways in which eqn. 2.6 can be transformed so as to give the equation of the form $y = mx + c$ for a straight line. For example, by multiplying both sides of eqn. 2.6 by $(K_{\text{m}} + [S])$, we obtain:

$$vK_{\text{m}} + v[S] = V_{\text{max}}\,[S]$$

Rearranging terms:

$$vK_{\text{m}} = V_{\text{max}}\,[S] - v[S]$$

Dividing each term on both sides by $[S]$:

$$\frac{vK_m}{[S]} = V_{\text{max}} - v$$

Dividing each term on both sides by $K_{\text{m}}$, we obtain eqn.2.9:

$$\frac{v}{[S]} = \frac{V_{\text{max}}}{K_{\text{m}}} - \frac{v}{K_{\text{m}}} \qquad \text{2.9}$$

This is of the form $y = mx + c$, where $y$ is $v/[S]$, $x$ is $v$, $m$ is $-1/K_{\text{m}}$, and $c$ is $V_{\text{max}}/K_{\text{m}}$. Thus, a plot of $v/[S]$ vs. $v$ is a straight line of slope $-1/K_{\text{m}}$ and a $y$-axis intercept of $V_{\text{max}}/K_{\text{m}}$. The intercept on the $x$-axis (derived by setting $v/[S] = 0$) is $V_{\text{max}}$. This is known as the Eadie–Hofstee plot (see Chapter 4, section 4.4).

The Eadie–Hofstee plot is a better way of analysing enzyme kinetic data than the Lineweaver–Burk plot because the distribution of errors is more uniform (see Chapter 4, section 4.4).

### ✓ WORKED EXAMPLE

Transform eqn. 2.6, i.e. $v = \dfrac{V_{\max}[S]}{K_m + [S]}$ so as to give the Hanes–Woolf equation (eqn. 2.10), for which the plot is $[S]/v$ vs. $[S]$.

$$\frac{[S]}{v} = \frac{[S]}{V_{\max}} + \frac{K_m}{V_{\max}} \qquad 2.10$$

**STRATEGY**

The approach is to rearrange the equation so as to be able to separate terms in the $y = mx + c$ form.

**SOLUTION**

By multiplying both sides of eqn. 2.6 by $(K_m + [S])$, we obtain:

$$vK_m + v[S] = V_{\max}\,[S]$$

Rearranging terms:

$$V_{\max}\,[S] = v[S] + vK_m$$

Dividing each term on both sides of the equation by $V_{\max}$:

$$[S] = \frac{v[S]}{V_{\max}} + \frac{vK_m}{V_{\max}}$$

Dividing each term by $v$:

$$\frac{[S]}{v} = \frac{[S]}{V_{\max}} + \frac{K_m}{V_{\max}} \qquad 2.10$$

Because $K_m$ and $V_{\max}$ are constants, this equation is of the form $y = mx + c$, with $y = [S]/v$ and $x = [S]$. A plot of $[S]/v$ vs. $[S]$ will be a straight line with an intercept on the $y$-axis of $K_m/V_{\max}$ and a slope of $1/V_{\max}$. The intercept on the $x$-axis is $-K_m$.

The Hanes–Woolf plot is a better way of analysing enzyme kinetic data than the Lineweaver–Burk plot because the distribution of errors is more uniform (see Chapter 4, section 4.4).

Other examples of rearranging equations to produce straight line graphs include:

### *The Arrhenius equation for variation of reaction rate constant with temperature*

The equation is:

$$k = Ae^{-Ea/RT} \qquad 2.2$$

where $A$, $E_a$, and $R$ are constants. (The Arrhenius equation has been mentioned in section 2.3.2 and is discussed in more detail in Chapter 4, section 4.2.3.) Taking the natural logarithms of both sides of eqn. 2.2, we obtain eqn. 2.11:

$$\ln k = \ln A - \frac{E_a}{RT} \qquad 2.11$$

Remember that the natural logarithm of $e^x = x$, and that the logarithm of the product of two numbers is sum of the logarithms of the numbers (see section 2.3).

Thus, a plot of $\ln k$ vs. $1/T$ is a straight line of slope $-E_a/R$, from which $E_a$ can be calculated (see Chapter 4, section 4.2.3).

### *The equation for a first-order process*

Equation 2.3 describes a first-order process (see section 2.4.3 and Chapter 4, section 4.2.1):

$$[A]_t = [A]_0 e^{-kt} \qquad 2.3$$

where $[A]_0$ and $k$ are constants.

Taking the natural logarithms of both sides of eqn. 2.3, we obtain:

$$\ln [A]_t = \ln [A]_0 - kt \qquad 2.12$$

A plot of $\ln [A]_t$ vs. $t$ is a straight line of slope $-k$, yielding the rate constant directly (see Chapter 4, section 4.2.1).

SELF TEST ?

**Check that you have mastered the key concepts at the start of this section by attempting the following question.**

**ST 2.12** The equation for the osmotic pressure ($P$) exerted by a solution of a protein whose molecular mass equals $M$ and of concentration $c$ g L$^{-1}$ is given by:

$$P/RTc = 1/M + Bc$$

where R, T, and B are constants. How would you determine M from a suitable graph?

Answer

ST 2.12 A plot of $P/RTc$ vs. $c$ will have a $y$-axis intercept of 1/M; $M$ is obtained by taking the reciprocal of this intercept.

In ST 2.12 the equation is already in the form $y = mx + c$, where $c$ is plotted on the $x$-axis and $P/RTc$ on the $y$-axis. For a given protein, $M$ (and hence $1/M$) will be constant.

## 2.6 Some basic statistics

KEY CONCEPTS střední

- Defining the mean, median, and mode of a distribution curve
- Defining the mean and standard deviation of a normal distribution curve
- Testing the difference of two means using the Student's *t* function
- Testing for correlation between variables; linear and non-linear regression

It is important to realize that virtually all the statements we make in an experimental science are statistical ones. We may be very confident, for example that falling out of an airplane at an altitude of 6500 m without a parachute will be fatal, or at a more mundane level, that administration of a statin-type drug (such as simvastatin)

In 1942, Lt I.M. Chisov, a Soviet pilot survived after ejecting at over 6500 m from his Ilyushin 4 plane when his parachute failed to open. Although he sustained significant injuries, he was back in the cockpit a few months later.

The statin drugs work by inhibiting a key enzyme involved in the biosynthesis of cholesterol. This can lead to a significant reduction (up to 50%) in blood cholesterol levels, and thus reduce the risk of suffering a heart attack.

will lead to the lowering of blood cholesterol levels, but this is not always the case for every individual. We need some way of estimating the degree of confidence with which we can make statements; this is the realm of statistics. The coverage of this topic for the molecular biosciences is much less than would be needed for the environmental and ecological sciences, principally because in the former we usually perform experiments in which we vary the important parameters (concentration, temperature, pH, etc.) in a systematic way to test some accepted theory or model. In contrast, in the more complex relationships in ecology we may have to consider the effects of many variables at the same time. This would require a much more detailed statistical approach to establish significant correlations between parameters, which could then be investigated in detail to derive the causal mechanisms involved (for example, how A influences B, and subsequently C). Statistics is also useful for establishing the degree of confidence with which we can quote the value of an experimentally measured or derived parameter such as the amount of protein in a solution, the rate constant of a reaction, or the Michaelis constant for the substrate of an enzyme. For many applications of statistics, it is common to use a 95% significance threshold, i.e. that we can be 95% confident about a certain outcome, but in some cases it may be important to be at least 99% confident.

spolehliví mít

### 2.6.1 Distributions of variables

The starting point for our discussion is the way that the values of parameters can be distributed. For example, if we were to measure the speeds at which vehicles were proceeding along an autobahn in Germany (where there is no official speed limit in rural areas), we might find that most of the vehicles were at speeds in the range 90–110 kph (roughly 55–70 miles per hour) but there would be some lorries going slower than this, and some high-performance cars going at speeds of 150 kph or higher. When plotted as a graph with the value of the speed (or rather the range of speeds, such as between 90 and 92 kph) on the $x$-axis and the number of vehicles measured as being within that range on the $y$-axis, we might obtain a distribution curve of the type as shown in Fig. 2.6.

There are three important values associated with a distribution curve. The *mean*, or more strictly the arithmetic mean, $(\bar{x})$ is defined as the average of the values $(x_1,x_2,x_3,x_4$, etc.) of the parameter plotted on the $x$-axis. This is defined mathematically by eqn. 2.13:

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n} = \frac{\Sigma(x)}{n} \qquad \mathbf{2.13}$$

The symbol $\Sigma$ is the Greek capital letter sigma (S).

where $n$ is the number of values of the parameter in question ($x$ in this case), and $\Sigma$ means 'the sum of the values'.

The *median* is defined as the middle value of the parameter, i.e. that value with as many values above it as below it. If we have an even number of values, the

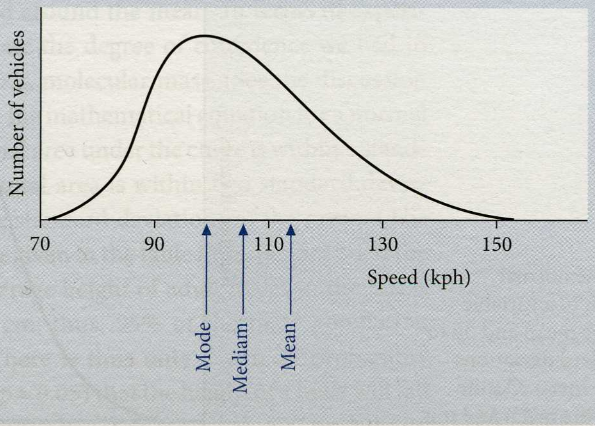


**Fig. 2.6** A skewed distribution showing the hypothetical distribution of vehicle speeds along a German autobahn. The mode, median, and mean of the distribution are indicated.

median is the numerical average of the middle two values. For example, if the median of the values 10.2, 10.7, 11.0, 11.1, 11.5, 11.6, 11.9, 12.2 would be the average of 11.1 and 11.5, i.e. 11.3. The median can be given with its quartiles. The first quartile value has $^1/_4$ of the values below it, the third quartile value has $^1/_4$ of the values above it. The inter-quartile range contains the middle $^1/_2$ of the values.

The *mode* is that value which occurs most commonly. The term mode is valuable in describing a distribution of variables which cannot be ranked (e.g. eye colour) and distribution which might show two peaks (this would be termed bimodal).

The mean, median, and mode of the distribution of vehicle speeds are indicated in Fig. 2.6.

### 2.6.2 The normal distribution

One particularly important type of distribution is known as the normal distribution in which the values of a continuous variable (i.e. one which can take any value, rather than just discrete values) are distributed symmetrically around the mean value. This is shown in Fig. 2.7. This would apply, for example, to physical characteristics such as height or weight, or to examination scores when measured for a suitably large sample size of the population. However, of more importance in the present context is that it also describes the distribution of values of experimental measurements subject to random variations. These would include, for instance, properties of samples taken from individual organisms that have been chosen to be well matched, or replicate determinations of some property of a sample from one particular source. Because of the symmetrical nature of this distribution, the values of the mean, mode, and median all coincide.

The word 'normal' refers to the mathematical form of the distribution curve; it does not mean 'expected' or 'typical'.

A normal distribution is characterized by the values of the mean and the standard deviation.

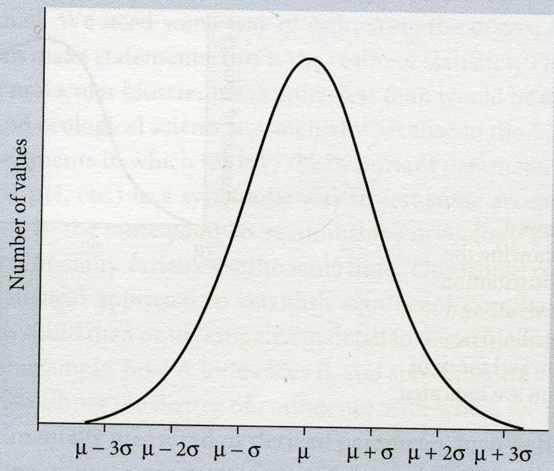


**Fig. 2.7** The normal distribution of a variable showing the mean and up to three standard deviations around the mean. On the graph, $\mu$ is the mean and $\sigma$ is the standard deviation.

The population mean ($\mu$) is defined by eqn. 2.14:

$$\mu = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n} = \frac{\Sigma(x)}{n} \qquad 2.14$$

where $x_1$, $x_2$, etc. are the individual values of the property and $n$ is the number of values in the population.

In practice, we are rarely able to study the entire population so we study the properties of a sample of the population; the sample mean ($\bar{x}$) has already been defined by eqn. 2.13:

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n} = \frac{\Sigma(x)}{n} \qquad 2.13$$

where $n$ is the size of the sample.

The standard deviation ($SD$, also designated as $\sigma$) is defined for a population by eqn. 2.15:

$$SD = \sqrt{\frac{\Sigma(x-\mu)^2}{n}} \qquad 2.15$$

The symbol σ is the Greek small letter sigma (s).

where $x$ is an individual value of the property, $\mu$ is the population mean, and $n$ is the number of values in the population.

For a sample taken from the entire population, the standard deviation ($SD$) is defined in an analogous fashion, except that the term $n - 1$ is introduced into the denominator, as shown in eqn. 2.16:

$$SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \qquad 2.16$$

The introduction of the term $n - 1$ rather than $n$ into the expression for the standard deviation is difficult to explain in simple terms, but it gives a better shape to the distribution curve. For values of $n > 20$, there is only a small (<3%) difference between the two expressions.

The value of the standard deviation relative to that of the mean gives an indication of how tightly the values are grouped around the mean. In terms of experimentally derived values this would indicate the degree of confidence we had in stating the value of the given parameter, e.g. molecular mass. (See the discussion on significant figures in section 2.2.) From the mathematical equation for a normal distribution, it is found that 68.2% of the total area under the curve is within 1 standard deviation of the mean, 95.4% of the total area is within two standard deviations of the mean and 99.7% within three standard deviations of the mean. (The areas of a standard normal distribution are given in the table in Appendix 2.1 at the end of this chapter.) For example, the average height of adult males in the UK is 178 cm with a standard deviation of 5 cm; thus, 95% of the male population is between 168 and 188 cm in height. There is thus only a 1 in 20 probability (which we express using the symbol $p$, i.e. $p = 0.05$) that the height of a male will fall outside that range. For example, there would be a less than 1 in 500 probability ($p < 0.003$) that a male is 195 cm tall.

The coefficient of variation (*CV*) is often used to describe the degree of variability of a population. It is defined as: $CV = 100\ (SD/\bar{x})\%$, where $\bar{x}$ and $SD$ are the mean and standard deviation of the population, respectively.

The way in which the sample mean ($\bar{x}$) might vary from the (true) population mean ($\mu$) is described by the term standard error of the mean (*SEM*) which is defined by eqn. 2.17.

$$SEM = \frac{SD}{\sqrt{n}} \qquad 2.17$$

Clearly, the larger the sample size $n$, the smaller the value of the *SEM*. From the properties of normal distributions:

$\mu \pm 1.96\ SEM$ will include 95% of the sample means

$\mu \pm 2.58\ SEM$ will include 99% of the sample means

### WORKED EXAMPLE ✓

The operation of a pipette was checked by repeatedly dispensing and weighing volumes of water. The volume on the pipette was set at 1 mL, and the following volumes (mL) were dispensed in succession: 0.932, 0.927, 0.948, 0.937, 0.918, 0.929, 0.940, and 0.942. What is the mean and standard deviation of these values? Comment on the reliability of the pipette.

**STRATEGY**

We use eqns. 2.13 and 2.16 to evaluate the mean and standard deviation, respectively.

**SOLUTION**

The mean value is 0.934 mL and the standard deviation is 0.0096 mL. From the properties of the normal distribution 99.7% of the values would be within the range 0.905 to 0.963 mL, which is significantly different from the nominal value of 1.000 mL. Thus, we can conclude that the pipette is *precise* (i.e. it delivers volumes which are reproducibly close to each other, with a low standard deviation), but it is *not accurate* (i.e. it is not sufficiently close to the true, or required, value). If the experiment had given a mean of 1.002 mL with a standard deviation of 0.0096 mL, the pipette would be both *precise* and *accurate*.

The words 'accurate' and 'precise' are often used interchangeably; it is important to appreciate their correct scientific usage, as in this example.

Having looked at the way in which statistics can be used to describe the distributions of variables, we shall now briefly consider two important applications of statistics in drawing conclusions from such distributions.

### 2.6.3 Testing the difference between two means

A very common use of statistics is to decide whether a change in a parameter is significant. For example, does the administration of a certain drug lead to a significant reduction in blood pressure, or are any changes observed merely due to chance? A trial may be set up with matched pairs of patients half of whom are given the drug and the other half given a dummy 'placebo' which is the control. The blood pressure data are collected and presented in the form of a mean and standard deviation for each group. In order to be able to draw reliable conclusions, the sample sizes should be as large as possible; indeed the trials of new drugs usually involve at least several hundred patients. In each group (drug and placebo) there will be a range of values of blood pressure, each with its own mean and standard deviation (Fig. 2.8).

The large numbers of patients required for the later stages of drug trials is a major factor in the cost of developing new drugs. It is estimated that each new drug would have cost several hundred million dollars to bring to market.

The way we usually proceed is to test the so-called 'null hypothesis', that is that there is no real difference between the mean values for the two groups (i.e. that the drug does not really cause any effect) and that any difference observed reflects random variations between individuals. Testing this hypothesis would certainly be important for the trial of a new drug, since there is an onus on the company to prove that the new drug is more effective than any existing treatments.

We first calculate the standard error of the difference ($SE_d$) between the two means, according to eqn. 2.18:

$$SE_d = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}} \qquad 2.18$$

where $SD_1$ and $SD_2$ are the standard deviations of the two groups (of sizes $n_1$ and $n_2$, respectively).

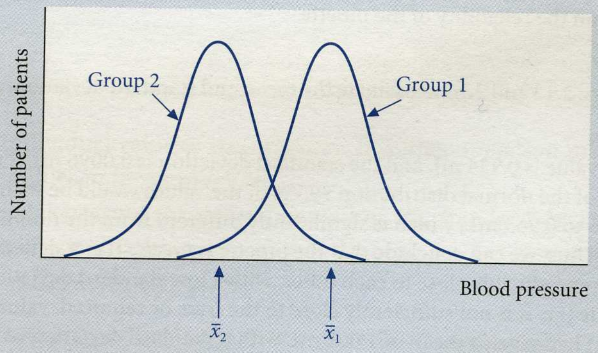


**Fig. 2.8** Distribution curves for the blood pressure data for two groups of patients. The mean values for the two groups are indicated ($\bar{x}_1$ and $\bar{x}_2$).

We then use the $t$ function (more properly known as the Student's $t$ function) to assess the significance of the differences. The $t$ function is defined by eqn. 2.19:

$$t = \frac{(x - \mu)\sqrt{n}}{SD} \qquad 2.19$$

It shows a distribution around a mean value similar to the normal distribution, but is more appropriate for smaller sample sizes.

The $t$ function was defined by William Gosset in 1908. His employers, Guinness Breweries, requested him to publish his work under a pseudonym, so he chose the name 'Student').

In terms of testing the differences between two means, the appropriate definition of $t$ is given by eqn. 2.20:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{SE_d} \qquad 2.20$$

where $\bar{x}_1$ and $\bar{x}_2$ are the two sample means and | | means 'irrespective of sign'.

The probability ($p$) that the two sample means are identical can be deduced from the properties of the $t$ function for the appropriate number of degrees of freedom (equal to $n_1 + n_2 - 2$). The values of the $t$ function are given in the table in Appendix 2.2 at the end of this chapter.

WORKED EXAMPLE ✓

In a small-scale drug trial, the sample mean values of the diastolic blood pressures of the drug and placebo groups were 122.5 and 110.3 mm, respectively. There were 20 patients in each group. The standard deviations for the two groups were 20.5 and 18.1 mm, respectively. Do the data show (at the 95% confidence level) that the drug has an effect on the blood pressure?

STRATEGY

We calculate the standard error of the difference (eqn. 2.18), and from that the value of $t$ (eqn. 2.20). Reference to the table of $t$ values allows us to reach a conclusion about the significance of any change.

SOLUTION

The value of $SE_d = 6.11$ mm. The value of $\bar{x}_1 - \bar{x}_2 = 12.2$ mm. Hence $t = 1.997$. Reference to the table in Appendix 2.2 shows that $t$ is below the entry value (2.02) for 95% confidence. Hence, we cannot reject the null hypothesis and must conclude that the drug has not been shown to have an effect. Since $t$ is quite close to the entry value, it would probably be worthwhile extending the test to include more patients; this may well increase the value of $t$ significantly.

### 2.6.4 The correlation coefficient and linear regression

The term 'correlation' refers to how strongly two variables are related. For example, if we were to plot a scatter diagram showing the shoe sizes of individuals against their height, we would expect to see a positive relationship between the two (tall

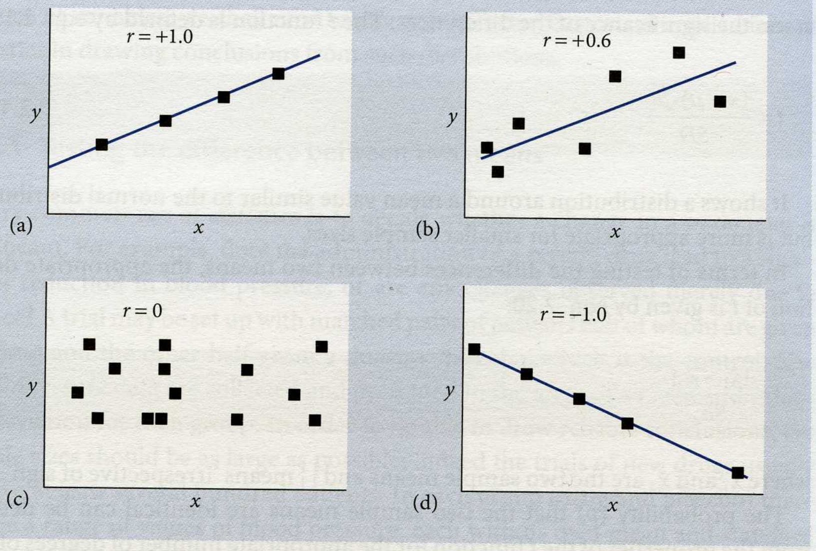


**Fig. 2.9** Scatter diagrams showing the correlation between variables $x$ and $y$. (a) Correlation coefficient ($r$) = +1.0, perfect positive correlation; (b) $r$ = +0.6, partial correlation, the significance depends on the number of degrees of freedom, see Appendix 2.2; (c) $r$ = 0, no correlation; (d) $r$ = −1.0, perfect negative correlation.

people generally have large feet). We can define the correlation coefficient, $r$, for the variables $x$ and $y$ by eqn. 2.21:

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\Sigma(y-\bar{y})^2}} \qquad 2.21$$

where $\bar{x}$ and $\bar{y}$ are the means of the values of $x$ and $y$, respectively, in the data set..

Values of $r$ can range from −1 (perfect negative correlation) to +1 (perfect positive correlation). Some examples of scatter diagrams and the associated values of $r$ are shown in Fig. 2.9.

The value of $r$ which indicates a significant correlation between two variables depends on the number of $(x,y)$ data points we have (strictly speaking on the degrees of freedom ($n$), which equals the number of $(x,y)$ data points −2). Values of $r$ which are used to establish a correlation are listed in the table in Appendix 2.3 at the end of this chapter. For example, we could say (at the 95% confidence level) that two variables are positively correlated if $r \geq 0.754$ ($n = 5$) or $r \geq 0.576$ ($n = 10$). Drawing this sort of conclusion is important if one is trying to establish a correlation between two variables before trying to propose a mechanism for a causal relationship.

In the molecular biosciences it is more likely that we are investigating the validity of a model and are testing experimental data against that model. If we are testing an equation where we would expect a straight line relationship (see section 2.5.2), then we can plot the appropriate parameters on a graph to check that the equation and hence the model are obeyed. In an ideal world (perfect data), all the points would fall on the straight line and determination of the slope and

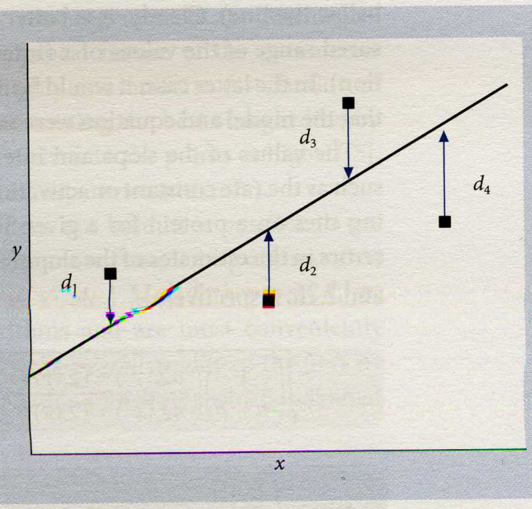


**Fig. 2.10** Least-squares line for the data points shown by filled squares. The deviations from the straight line are indicated by $d_1$, $d_2$, etc.

intercept would be trivial. In practice, due to errors in measurements, it is likely that the points would be scattered around a straight line. The determination of the best straight line is known as linear regression and the line produced is known as the regression line of $y$ on $x$. The most widely used method to do this is the least-squares method, in which the sum of the squares of the differences between the calculated and observed values of $y$ at each value of $x$ is minimized (Fig. 2.10).

Note in Fig. 2.10 that the squares of the differences are used so that deviations below the line and above the line both contribute to the total deviation.

The equation for the least-squares fit straight line is $y = mx + c$, where the slope $m$ is given by eqn. 2.22:

$$m = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} \qquad 2.22$$

Knowing the values of $m$, $\bar{x}$, and $\bar{y}$, the value of the $y$-axis intercept $c$ can be calculated from eqn. 2.23:

$$\bar{y} = m\bar{x} + c \qquad 2.23$$

Once the best straight line has been determined, it can be used to predict values of $y$ for given values of $x$. The standard error of the estimate of $y$ ($s_e$) is given by eqn. 2.24:

$$s_e = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n-2}} \qquad 2.24$$

where $n$ is the number of $(x,y)$ points.

We can then establish a 95% confidence band around the regression line, which will be marked by two lines (one 1.96 times $s_e$ above the line, the other 1.96 times $s_e$

below the line). Clearly, it is better if one is using this approach within the measured range of the values of $x$ (interpolation) than outside this range (extrapolation). In the latter case, it would be important to establish (or necessary to assume) that the model and equation were valid outside the range.

The values of the slope and intercept may then be used to derive parameters such as the rate constant or activation energy of a reaction, or the number of binding sites on a protein for a given ligand. It is possible to calculate the standard errors in the estimates of the slope ($s_m$) and the $y$-axis intercept ($s_c$) using eqns. 2.25 and 2.26, respectively.

$$s_m = \sqrt{\left(\frac{1}{n-2}\right)\left(\frac{n\Sigma(y^2)-(\Sigma(y))^2}{n\Sigma(x^2)-(\Sigma(x))^2} - m^2\right)} \qquad 2.25$$

$$s_c = s_m\sqrt{\frac{\Sigma(x^2)}{n}} \qquad 2.26$$

✓ WORKED EXAMPLE

The rate ($v$, in units of μM min$^{-1}$) of an enzyme-catalysed reaction was studied as function of substrate concentration ([$S$], in units of μM). The data were analysed by the Hanes–Woolf plot (eqn. 2.10) in which [$S$]/$v$ is plotted against [$S$]. The following values were obtained:

| [$S$]/$v$ | 13.2 | 18.0 | 18.2 | 22.9 | 25.3 | 27.0 | 30.7 |
|---|---|---|---|---|---|---|---|
| [$S$] | 10 | 20 | 30 | 40 | 50 | 60 | 70 |

Calculate the correlation coefficient for the plot of [$S$]/$v$ vs. [$S$], and use linear regression to calculate the best straight line.

**STRATEGY**

This is an application of eqns. 2.21, 2.22, and 2.23. It is a good idea to draw up a table to calculate the various terms required for these equations.

**SOLUTION**

[$S$]/$v$ is designated as $y$ and [$S$] as $x$. The values of $\bar{y}$ and $\bar{x}$ are 22.19 and 40, respectively. The values of $\Sigma(y-\bar{y})^2$ and $\Sigma(x-\bar{x})^2$ are 220.03 and 2800, respectively. The value of $\Sigma(x-\bar{x})(y-\bar{y})$ is 776. From this, using eqn. 2.21, $r = 0.9886$; this is highly significant correlation ($p < 0.001$ for 5 degrees of freedom, i.e. the number of $(x,y)$ data points (7) −2). The slope ($m$) and $y$-axis intercept ($c$) of the least-squares line are 0.277 and 11.1, respectively. Further analysis using eqns. 2.25 and 2.26 shows that $s_m = 0.029$ and $s_c = 1.3$.

### 2.6.5 Non-linear regression

Although linear regression is a very useful method, there are many occasions when the relationships between variables cannot be expressed in terms of a simple straight line equation, or where such a relationship could cause problems. (One example of the latter is the Lineweaver–Burk rearrangement of the Michaelis–Menten equation (see section 2.5.1). The reciprocal nature of the parameters plotted ($1/v$ and $1/[S]$) means that there is a highly non-uniform distribution of errors over the range of values, so that in determining the best straight line by the least-squares method, the greatest weight is given to the points at high $1/[S]$, i.e. low $[S]$, which are associated with the greatest experimental errors).

In cases where a straight line relationship does not hold, it is possible to use non-linear regression, where the data are fitted to more complex equations, often involving higher power dependence on $x$ (e.g. $x^2$, $x^3$, etc.). Most fitting procedures involve the use of complex numerical algorithms and are most conveniently performed by computers. One way of assessing the overall quality of the fit is by evaluating the normalized root mean square deviation (NRMSD), which is defined by eqn. 2.27:

$$NRMSD = \sqrt{\frac{\Sigma(y_{\text{obs}} - y_{\text{cal}})^2}{\Sigma(y_{\text{obs}})^2}} \qquad 2.27$$

where $y_{\text{obs}}$ and $y_{\text{cal}}$ are the observed and calculated (according to the fitting equation) values of $y$, at each specified value of $x$. The NRMSD can take values ranging from 0 (perfect fit) to 1 (no fit whatsoever); generally, values less than 0.1 are considered satisfactory.

There are many programs commercially available for the direct fitting of enzyme kinetic data to theoretical models such as that described by the Michaelis–Menten equation. The majority of these use the Levenberg–Marquardt algorithm, which employs an iterative approach to find the values of the parameters in the chosen model which give the best fit to the experimental data as judged by the sum of the squares of the differences being minimized. Initial trial values of these parameters are either supplied or guessed and these are then varied in an incremental fashion and the effect on the goodness of fit assessed. In the Levenberg–Marquardt approach, the sizes of the incremental changes can be automatically adjusted according to how well the values of the parameters are converging towards their final values.

In the case of enzyme kinetic data, these direct fitting procedures can be used to fit the data ($v$ as a function of $[S]$) directly to the Michaelis–Menten equation (see Chapter 4, section 4.4). The program will produce estimates of the parameters $K_m$ and $V_{\text{max}}$, together with the standard errors of the estimates in these quantities. Ideally, the errors should be ≤5% of the values of the parameters. Low values of these errors give confidence that the equation (and the model on which it is based) is obeyed, and that the measurements are not subject to excessive random errors.

A low error value does not, however, exclude the possibility of a systematic error. For example, if a stock solution of substrate had been made up at the wrong concentration, then the value of $K_m$ would be incorrect, even if there were no errors in pipetting or measurement of rates.

As well as the NRMSD, a useful further check on the appropriateness of the analysis of the data is to look at the so-called pattern of residuals (the differences between the calculated and observed values of the $y$ parameter at each value of $x$). When these differences are plotted against the values of the $x$ parameter, there

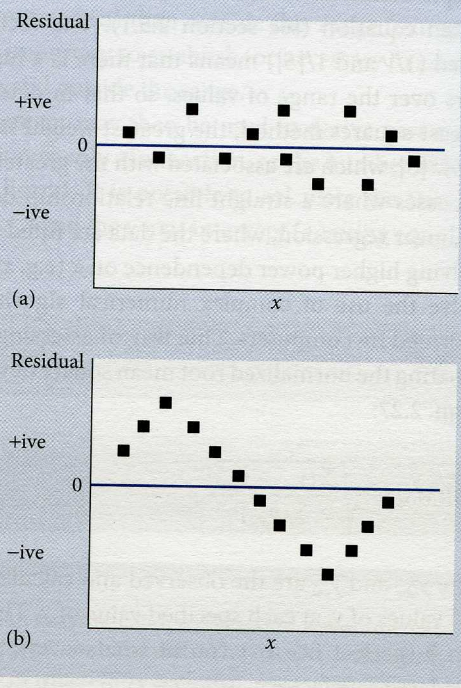


**Fig. 2.11** Pattern of residuals to test methods of data analysis. (a) The data points show a random pattern of deviations around the zero residual line; (b) the data points show a systematic pattern of deviations around the zero residual line.

should be a random pattern around the line $y = 0$ (Fig. 2.11(a)). If there is a systematic pattern of residuals (as, for example in Fig. 2.11(b)), this indicates that the method of analysis is inappropriate, and should be altered to yield a random pattern of residuals.

## ? SELF TEST

With respect to ST 2.13, the current guidelines for blood cholesterol levels are that they should be kept below 5.2 mM, although the distribution between various lipoprotein complexes is also important.

**Check that you have mastered the key concepts at the start of this section by attempting the following questions.**

**ST 2.13** The following values of blood cholesterol (mM) were found in a sample of eight healthy females: 4.3, 4.1, 5.8, 5.0, 3.9, 5.5, 5.2, and 4.9. What is the median, mean and standard deviation of these values? What is the 95% confidence limit for the population mean ($\mu$)?

**ST 2.14** A group of 21 healthy students undertook a glucose tolerance test in which they fasted overnight and then ingested 75 g glucose. Before taking the glucose, the blood glucose levels of the group had a mean value of 4.73 mM (SD = 0.48 mM). 30 min after taking the glucose, the blood glucose levels had a mean value of 7.95 mM (SD = 1.63 mM). After further 90 min, the levels had a mean value of 5.25 mM (SD = 1.53 mM). Are the levels at 30 and 120 min significantly different from that at the start?

**ST 2.15** A Hanes–Woolf plot (eqn. 2.10) used to analyse a set of enzyme kinetic data obtained at eight values of substrate concentration showed a correlation coefficient ($r$) of 0.669. What would you recommend to the investigator who produced the data?

**Answers**

**ST 2.13** The values are: median, 4.95 mM; mean, 4.84 mM; standard deviation, 0.68 mM; 95% confidence limit for $\mu$, 4.37–5.31 mM.

**ST 2.14** Comparing 30 min and start values, $t = 8.68$; this gives $p < 0.01$, i.e. the null hypothesis can be rejected with >99% confidence. Comparing 120 min and start values, $t = 1.49$; this gives a $p$ value between 0.2 and 0.1 ($0.2 > p > 0.1$); i.e. the null hypothesis cannot be rejected with at least 95% confidence. Thus, the 30 min value is significantly higher than the start value, but the 120 min value is not.

**ST 2.15** The value of $r$ is below the value required for 95% confidence of a positive correlation. It would not therefore be appropriate to use the plot to try to obtain reliable values of the kinetic parameters ($K_m$ and $V_{max}$) for the enzyme. It would be sensible to try to improve the experimental technique and to obtain more data points.

With respect to **ST 2.14**, the current guidelines are that the fasting blood glucose levels should be in the range 3.3–6.1 mM; 30 min after the ingestion of glucose the level should be below 11.1 mM; after further 90 min, the level should have dropped to below 7.8 mM. Fasting blood glucose levels greater than 7.8 mM, and greater than 11.1 mM at the 120 min point indicate diabetes.

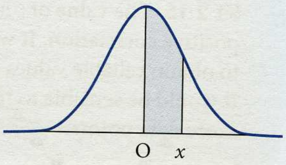
## References for Chapter 2

Cornish-Bowden, A. (1999) *Basic Mathematics for Biochemists*, 2nd edn. Oxford University Press, Oxford, 221 pp.

# Appendix

## Appendix 2.1 Table of areas of a standard normal distribution

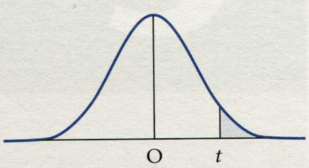
The entries in the table show the proportion of the total area under the curve which lies between $x = 0$ and the actual value of $x$. The areas for negative values of $x$ are obtained by symmetry.



| $x$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0753 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2257 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2517 | .2549 |
| 0.7 | .2580 | .2611 | .2642 | .2673 | .2703 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2995 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |

## Appendix 2.2 Table of values of Student's *t* function

The first column lists the number of degrees of freedom, *n*. The other columns show the probabilities (*p*) for t to be greater than the values listed. The values of *p* for negative values of *t* are obtained by symmetry.



| *p*<br>*n* | .10 | .05 | .025 | .01 | .005 |
|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 |

## Appendix 2.3 Table of critical values of correlation coefficient

The value of ***p*** shown is the probability that the absolute value of ***r*** exceeds the value shown in the table. Thus, for 8 degrees of freedom, the probability that *r* exceeds 0.632 when its true value is 0 (no correlation) is 0.05.

The number of degrees of freedom is the number of (*x*,*y*) data points −2.

THE CORRELATION COEFFICIENT

| *Degrees of freedom* | *Value of p* | | | | |
|---|---|---|---|---|---|
| | **0.10** | **0.05** | **0.02** | **0.01** | **0.001** |
| 1 | 0.9877 | 0.99692 | 0.99951 | 0.99988 | 0.9999988 |
| 2 | 0.900 | 0.950 | 0.980 | 0.990 | 0.999 |
| 3 | 0.805 | 0.878 | 0.934 | 0.959 | 0.991 |
| 4 | 0.729 | 0.811 | 0.882 | 0.917 | 0.974 |
| 5 | 0.669 | 0.754 | 0.833 | 0.875 | 0.951 |
| 6 | 0.621 | 0.707 | 0.789 | 0.834 | 0.925 |
| 7 | 0.582 | 0.666 | 0.750 | 0.798 | 0.898 |
| 8 | 0.549 | 0.632 | 0.715 | 0.765 | 0.872 |
| 9 | 0.521 | 0.602 | 0.685 | 0.735 | 0.847 |
| 10 | 0.497 | 0.576 | 0.658 | 0.708 | 0.823 |
| 11 | 0.476 | 0.553 | 0.634 | 0.684 | 0.801 |
| 12 | 0.457 | 0.532 | 0.612 | 0.661 | 0.780 |
| 13 | 0.441 | 0.514 | 0.592 | 0.641 | 0.760 |
| 14 | 0.426 | 0.497 | 0.574 | 0.623 | 0.742 |
| 15 | 0.412 | 0.482 | 0.558 | 0.606 | 0.725 |
| 16 | 0.400 | 0.468 | 0.543 | 0.590 | 0.708 |
| 17 | 0.389 | 0.456 | 0.529 | 0.575 | 0.693 |
| 18 | 0.378 | 0.444 | 0.516 | 0.561 | 0.679 |
| 19 | 0.369 | 0.433 | 0.503 | 0.549 | 0.665 |
| 20 | 0.360 | 0.423 | 0.492 | 0.537 | 0.652 |
| 25 | 0.323 | 0.381 | 0.445 | 0.487 | 0.597 |
| 30 | 0.296 | 0.349 | 0.409 | 0.449 | 0.554 |
| 35 | 0.275 | 0.325 | 0.381 | 0.418 | 0.519 |
| 40 | 0.257 | 0.304 | 0.358 | 0.393 | 0.490 |
| 45 | 0.243 | 0.288 | 0.338 | 0.372 | 0.465 |
| 50 | 0.231 | 0.273 | 0.322 | 0.354 | 0.443 |
| 60 | 0.211 | 0.250 | 0.295 | 0.325 | 0.408 |
| 70 | 0.195 | 0.232 | 0.274 | 0.302 | 0.380 |
| 80 | 0.183 | 0.217 | 0.257 | 0.283 | 0.357 |
| 90 | 0.173 | 0.205 | 0.242 | 0.267 | 0.338 |
| 100 | 0.164 | 0.195 | 0.230 | 0.254 | 0.321 |