

## Finding rates of reaction in photosynthesis

### Specification references

- 3.5.1
- MS 3.1, MS 3.2, MS 3.5, MS 4.1

### Maths Skills for Biology references

- *Oxygen release from photosynthesis*

### Learning outcomes

After completing this worksheet you should be able to:

- plot graphs to show two variables and draw best fit lines
- calculate the rate of change from graphs showing linear relationships
- recall the formula for calculating volume of a cylinder.

### Introduction

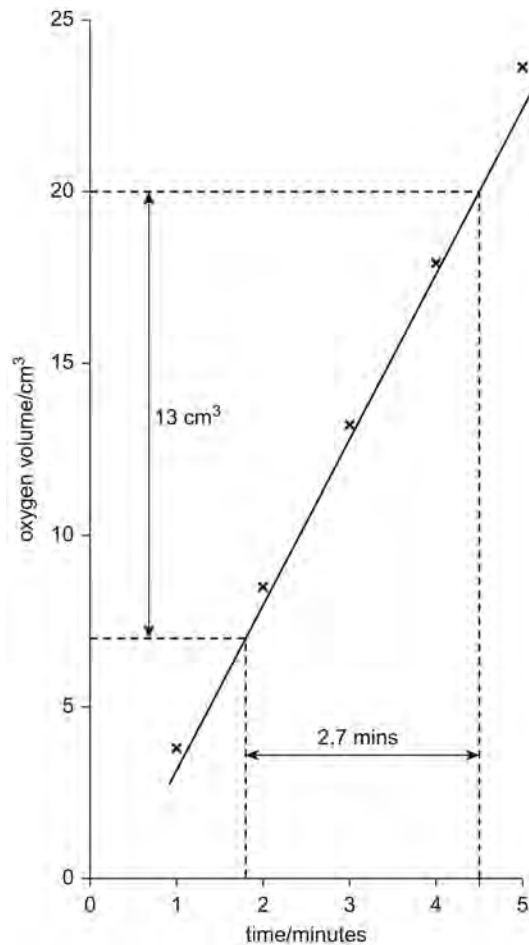
You will have studied photosynthesis in Section 5, covered in Chapter 11 of your Student book. You will know that the rate of photosynthesis can be affected by key factors such as light intensity, temperature or carbon dioxide concentration. Experiments with photosynthesis commonly use these key factors as independent variables, e.g. light intensity or carbon dioxide availability, measuring the production of oxygen gas over time to follow the reaction. Plotting the volume of oxygen produced against time produces a graph from which a rate of reaction can be calculated and related to the factor being varied.

This task is set in the context of experiments in which a capillary tube is used to collect and measure the volume of oxygen produced during photosynthesis. It focuses on developing your graphing skills and also on a method for converting distances in capillary tubing into actual volumes.

**Worked example**

**Question**

In an experiment oxygen was collected for five minutes at a set light intensity. Find the rate of reaction from the graph in Figure 1, which shows the data collected.



**Figure 1**

**Answer**

*Step 1*

Once the straight best fit line has been drawn, select any two points along the line and use a ruler to draw construction lines to each axis, as shown on Figure 1.

*Step 2*

Divide the difference between these points on the *y*-axis by the difference between them along the *x*-axis.

In this example the calculation is  $13 \div 2.7$ , giving a rate of  $4.8 \text{ cm}^3 \text{ minute}^{-1}$ .

## Questions

- 1 Oxygen production was measured by using a capillary tube apparatus to collect the gas. This was repeated at five different temperatures. The results were as follows:

Temperature/°C	15	25	35	45	55
Time/s	Volume of oxygen collected/mm <sup>3</sup>				
30	3	5	7	8	7
60	3	7	10	11	9
90	5	10	15	17	13
120	6	13	19	21	17
150	8	15	24	27	21
180	9	17	27	30	24
210	9	19	28	31	25
240	10	22	30	33	27
270	11	24	34	37	30
300	12	26	36	39	31

- a Plot a neat graph to show the cumulative oxygen collected at each temperature.

(4 marks)

**b** Use the graph to determine the rate of oxygen production at each temperature.

.....  
.....  
.....  
.....

..... (1 mark per temperature)

**c** Plot a graph to summarise the effect of temperature on the rate of oxygen production.

(4 marks)

**d** The oxygen was collected in capillary tubes, in which it formed bubbles. Describe how the length of the bubbles would have been converted to volumes.

.....  
..... (2 marks)

### Maths skills links to other areas

This technique can be used in areas such as oxygen consumption rates in respirometers or water uptake in potometers, where a straight line relationship between volume and time might be expected.

When the graph shows a reaction that reaches an end point, such as an enzyme digestion over time, the technique you should use is tangents. An example of this is in the worksheet 'Finding rates of reaction using tangents', which is also on Kerboodle.

**Answers**

**1 a** A neat line graph should be drawn, with the following criteria:

- graph fills at least half the space available
- axes correct way around and divided equidistantly
- axes labelled with title and units
- points plotted accurately
- straight best fit lines drawn
- a key to identify the line for each temperature.

(Max 4 marks)

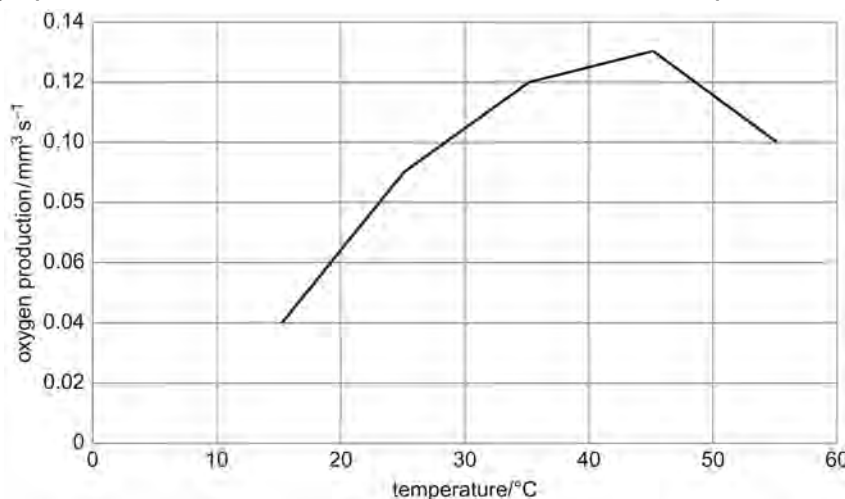
The rates will depend on the positioning of the best fit lines, but they will be in region of:

Temperature/°C	Rate/mm <sup>3</sup> s <sup>-1</sup>
15	0.03
25	0.08
35	0.11
45	0.11
55	0.09

(1 mark per row)

(5 marks)

**c** The graph should also match the relevant criteria listed in part a and will look like this:



(4 marks)

**d** The length of the bubble and radius of capillary tube could be used with the formula  $\pi r^2 h$

(1 mark)

(1 mark)

(2 marks)

## Net primary production and carbon sinks

### Specification references

- 3.5.3
- MS 0.1, MS 0.2, MS 0.3, MS 1.2

### Maths Skills for Biology references

- Numbers and units
- Percentages

### Learning outcomes

After completing this worksheet you should be able to:

- convert between units
- use standard form for very small or large numbers
- calculate percentages
- consider the relative importance of different ecosystems as carbon sinks.

### Introduction

This worksheet links with the material from Section 5 that is described in Chapter 13 of your Student book. The focus is on practising the maths skills involved with converting units and using standard form. These skills are introduced in the context of productivity in different ecosystems. You will be converting units in different ways in the context of carbon fixation or net primary production.

A carbon sink is a reservoir within which carbon is stored in organic compounds. This can be for a brief period, e.g. a few years in temperate grassland, to millions of years in limestone rocks. Carbon dioxide fixed as organic compounds can be measured as net primary production. Photosynthetic organism biomass can, therefore, be regarded as a primary carbon sink.

### Worked example

#### Questions

Convert the following:

- 18 g to kg
- 2.6 tonnes to kg
- $5 \text{ m}^2$  to  $\text{km}^2$
- $2 \text{ km}^2$  to  $\text{m}^2$

**Answers****a***Step 1*

There are 1000 g in one kg, so  $1\text{g} = \frac{1}{1000\text{kg}}$ . To convert g to kg divide by 1000.

*Step 2*

So  $18\text{g} = 18 \div 1000\text{kg}$   
 $= 0.018\text{kg}$

**b***Step 1*

There are 1000 kg in one tonne, so  $1\text{kg} = \frac{1}{1000\text{tonne}}$ . To convert tonnes to kg multiply by 1000.

*Step 2*

So  $2.6\text{tonnes} = 2.6 \times 1000\text{kg}$   
 $= 2600\text{kg}$

**c***Step 1*

$1\text{km}^2$  (one square kilometre) is an area  $1000\text{m} \times 1000\text{m}$ , so it is  $1\,000\,000\text{m}^2$ . To convert  $\text{m}^2$  to  $\text{km}^2$  divide by  $1\,000\,000$ .

*Step 2*

So  $5\text{m}^2 = 5 \div 1\,000\,000\text{km}^2$   
 $= 0.000005\text{km}^2$   
 $= 5 \times 10^{-6}\text{km}^2$  expressed as standard form.

To convert to standard form, write down the smallest number between 1 and 10 that can be derived from the number to be converted. In this case it would be 5.0.

Next write the number of times the decimal place will have to shift to expand or reduce this to the original number as powers of ten. On paper this can be done by hopping the decimal over each number like this:



0.000005

until the end of the number is reached. This example requires 6 shifts backwards, so the standard form should be written as  $5 \times 10^{-6}$ .



**d**

*Step 1*

To convert km<sup>2</sup> back to m<sup>2</sup>, multiply by 1 000 000.

*Step 2*

So 2 km<sup>2</sup> = 2 × 1 000 000 m<sup>2</sup>

= 2 000 000 m<sup>2</sup>

= 2 × 10<sup>6</sup> expressed as standard form.

**Questions**

Ecosystem	Area/× 10 <sup>6</sup> km <sup>2</sup>	Mean net primary production/g m <sup>-2</sup> year <sup>-1</sup>	Total annual net primary production/tonnes
Tropical rain forest	17	2200	3.74 × 10 <sup>10</sup>
Temperate deciduous forest	7	1200	8.4 × 10 <sup>9</sup>
Savannah grassland	15	900	1.35 × 10 <sup>10</sup>
Temperate grasslands	9	600	5.4 × 10 <sup>9</sup>
Swamp and wetland	2	3000	
Open ocean	332	125	
Algal bed and reef	0.2	2500	
Cultivated land	14	650	

**Table 1**

*(data quoted from R.H. Whittaker)*

**1** Table 1 contains data for estimated rates of net primary production and surface areas of some ecosystems.

**a** Calculate the area of the tropical rain forest in m<sup>2</sup>, expressed as standard form.

.....

..... (2 marks)

**b** Calculate the total annual net primary production of the temperate grasslands in grams, expressed as standard form.

.....  
..... (2 marks)

**c** Calculate the total annual net primary production of the open ocean in kg, expressed as standard form.

.....  
..... (2 marks)

**2** Complete the last column of the table by calculating the missing figures.

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
..... (8 marks)

**3** Calculate the percentage of the total annual net primary production:

**a** in tropical rain forests

.....  
..... (2 marks)

**b** in cultivated land.

.....  
..... (2 marks)

- 4 Tropical forests and wetlands are both threatened ecosystems. Use the data in Table 1 to assess the possible impact on the carbon cycle if these ecosystems are not protected from human destruction.

.....

.....

.....

.....

.....

.....

..... (6 marks)

**Maths skills links to other areas**

Standard form is useful for dealing with very large and very small numbers, for example in magnification (measuring with the microscope). You could be asked about percentages in many areas, for example, percentage yields of crop or products from fermenter systems.

You will need to be able to convert between units in areas such as magnification and scale calculations measuring from diagrams and photomicrographs.

## Answers

- 1 Allow one mark for correct sum shown but incorrect result in each case below:
- a  $1.7 \times 10^{13} \text{ m}^2$  (2 marks)
- b  $5.4 \times 10^{15} \text{ g}$  (2 marks)
- c  $4.15 \times 10^{13} \text{ kg}$  (2 marks)
- 2 Swamp and wetland  $6 \times 10^9$  (2 marks)  
Open ocean  $4.15 \times 10^{10}$  (2 marks)  
Algal bed and reef  $5 \times 10^8$  (2 marks)  
Cultivated land  $9.1 \times 10^9$  (2 marks)  
(8 marks)
- 3 a 31% (2 marks)
- b 7% (2 marks)
- 4 Together they contribute 36% of the net primary production  
but wetlands only 5%  
If wetland is converted to cultivated land, net primary production reduces by  
 $\frac{3000 - 650}{2350} \text{ g/m}^2/\text{year}$   
which is a  $\left(\frac{2350}{3000}\right) \times \frac{100\%}{78.3\%}$  drop  
This reduces net primary production to  $\frac{0.783 \times (6 \times 10^9)}{4.7 \times 10^9}$  tonnes of carbon per year  
If TRF is converted, net primary production reduces by  $\frac{2200 - 650}{1550} \text{ g/m}^2/\text{year}$   
which is a  $\left(\frac{1500}{2200}\right) \times \frac{100\%}{70.5\%}$  drop  
This reduces net primary production to  $\frac{0.705 \times (3.74 \times 10^{10})}{2.6 \times 10^{10}}$  tonnes of carbon per  
year  
Together, this represents a 25.5% drop in total carbon fixation per year  
Losing rainforests has a much larger impact. (6 marks)

## Using the Hardy–Weinberg equation to estimate allele frequencies

### Specification references

- 3.7.2
- MS 0.3, M. 2.3, MS 2.4

### Maths Skills for Biology references

- *The Hardy-Weinberg equation*

### Learning outcomes

After completing this worksheet you should be able to:

- use the Hardy–Weinberg equation to estimate the frequency of alleles
- substitute values into algebraic equations and solve the equations
- convert decimals to percentages.

### Introduction

In Section 7 you studied the inheritance of alleles. You will have learned that in a monohybrid cross there may be dominant and recessive alleles, so individuals expressing a dominant characteristic could be either homozygous or heterozygous but individuals expressing a recessive allele can only be homozygous. This is described in Chapters 17 and 18 of your Student book. A mathematical requirement is that you are able to apply an equation called the Hardy–Weinberg equation in order to determine allele frequencies. This task is designed to allow you to practise with that equation.

The frequency of dominant and recessive alleles in a population will remain in equilibrium from generation to generation, provided that:

- the population is large
- mating is random
- no mutations occur
- there is no selection, that is, all alleles are equally likely to be passed to the next generation
- the population is isolated, that is, there is no flow of alleles into or out of the population.

Thus it is possible to apply a statistical analysis to the population in order to establish the gene frequencies of the dominant and recessive alleles and the frequency of carriers in a population. This is the Hardy–Weinberg equation.

### Worked example

#### Question

Phenylketonuria is a condition caused by an autosomal recessive allele that occurs at a rate of 0.1 per 1000 live births. What is the frequency of the carrier genotype?

#### Answer

In the equation  $p^2 + 2pq + q^2 = 1$ ,  $p$  represents the frequency of the dominant allele and  $q$  the frequency of the recessive allele.

Consider the cross between two heterozygous parents, Aa:

<b>Parent genotypes</b>	Aa		Aa	
<b>Parent gametes</b>	A	a	A	a
<b>F1 genotypes</b>	AA	Aa	Aa	aa
<b>Substitute <math>p</math> and <math>q</math></b>	$p \times p$	$p \times q$	$p \times q$	$q \times q$
<b>To give</b>	$p^2$	$2pq$		$q^2$
<b>so</b>	$p^2 + 2pq + q^2 = 1$ (i.e. 100%)			

**1** Note the sum of all alleles is always 1, so  $p + q = 1$

*Step 1*

First find the frequency of the sufferer ( $q^2$ ), which is the homozygous recessive.

Here it is  $\frac{0.1}{1000} = 0.0001$

*Step 2*

If  $q^2 = 0.0001$  then  $q = \sqrt{0.0001}$

So  $q = 0.01$

*Step 3*

Since  $p + q = 1$  we can now find  $p$ , which is  $1 - 0.01 = 0.99$

*Step 4*

Therefore  $2pq = 2(0.99 \times 0.01) = 0.0198$  (1.98%)

So 1.98% of people are carriers.

**Questions**

**1** Around 1 in 5000 people in the UK have Alpha-1 Antitrypsin Deficiency, which is caused by an autosomal recessive allele.

**a** Calculate the frequency of the dominant allele.

.....  
 ..... (2 marks)

**b** Calculate the percentage of carriers in the population.

.....  
 ..... (2 marks)

**c** If the UK population is 65 million, calculate how many people will be carriers of Alpha-1 Antitrypsin Deficiency.

.....  
 ..... (2 marks)

**2** Sickle cell anaemia occurs in 1 in 500 African-American births but has an occurrence of 2% in West Africa. The alleles are H (normal) and Hs (sickle cell). The possible genotypes and phenotypes are HH normal blood, Hs carrier with partial anaemia and resistance to malaria, HsHs fully sickle cell anaemic (normally fatal).

**a** Calculate the frequency of carriers in both the USA and West Africa.

.....  
 .....  
 .....  
 ..... (4 marks)

**b** Suggest an explanation for the difference between the two populations.

.....  
.....  
.....  
.....  
..... (5 marks)

**3** Beta-Thalassaemia is a rare disorder caused by an autosomal recessive allele. It has an incidence of approximately 0.05 out of every thousand births.

**a** Calculate the frequency of the carrier genotype.

.....  
..... (2 marks)

**b** Calculate the percentage of the population that do not have the recessive allele in their genotypes.

.....  
..... (2 marks)

**4** Huntington's Chorea is a condition caused by an autosomal dominant allele. It has an incidence of around 7 per 100 000 of the population.

**a** Calculate the frequencies of the dominant and recessive alleles.

.....  
.....  
.....  
..... (4 marks)

**b** Calculate the percentage of the population that is likely to be heterozygous.

.....  
..... (2 marks)



**Maths skills links to other areas**

You will be given equations and expected to substitute in appropriate values and solve the equations in other areas of biology, for example calculating diversity index, calculating cardiac output, making statistical calculations.

## Answers

- 1 a  $p = 0.986$  (2 marks)  
 b 2.8% (2 marks)  
 c 1 820 000 or 1.82 million (2 marks)

- 2 a USA 8.6%, West Africa 24% (4 marks)

b Carrier is resistant to malaria

Homozygous recessive dies of SCA and homozygous dominant more likely to die from malaria

W. Africa carrier population is selected for resistance to malaria  
 so carrier genotype increases

US descendant carrier population no longer selected as malaria is absent

Homozygous dominant survives more

and increases the frequency of the dominant allele

(Max 5 marks)

- 3 a 0.14 (2 marks)  
 b 98.6% (2 marks)

- 4 a Dominant  $p$  allele =  $3.5 \times 10^{-5}$  (2 marks)  
 Recessive  $q$  allele = 0.999965 (2 marks)  
 (4 marks)

b 0.0069% of total population are heterozygous

Working for Question 4

$$p^2 + 2pq \text{ (all sufferers)} = 7 \text{ in } 100\,000 = 7 \times 10^{-5}$$

$$\text{so } q^2 = 1 - 7 \times 10^{-5} = 0.99993$$

$$\text{so } q = \sqrt{0.99993} \quad q = 0.999965$$

$$p + q = 1 \quad \text{so } p = 1 - 0.999965 = 3.5 \times 10^{-5}$$

$$2pq = 2(0.999965 \times 3.5 \times 10^{-5})$$

$$= 6.9998 \times 10^{-5}$$

$$\text{which is } (6.9998 \times 10^{-5} \div 100\,000) \times 100\%$$

$$= 0.0069\%$$

(2 marks)

## Using hydrogen ion concentration to calculate pH

### Specification reference

- 3.1.4.2
- Maths skill 0.5
- Maths skill 2.2
- Maths skill 2.3
- Maths skill 2.4

### Learning outcomes

After completing the worksheet you should be able to:

- use calculators to find and use power, exponential, and logarithmic functions
- change the subject of a formula
- substitute numerical values into formulae and solve formulae.

### Introduction

The maths skill in this worksheet focuses on the use of power and logarithmic functions. A function is simply a relationship between numbers, for example  $x$  or  $+$ . A power, written using a superscript to the right of the number, tells you the number of times to use the number in a multiplication. For example,  $2^3$  means  $2 \times 2 \times 2$ . To save space, numbers can be written using powers of 10, or 'standard form'. For example,  $5 \times 10^3$  is a way of writing 5 000;  $5 \times (10 \times 10 \times 10) = 5\,000$ . Logarithms are based on powers of 10 and allow large numbers to be accommodated easily on a scale. For example, 1, 10, 100, 1000, 10 000 become logs 0, 1, 2, 3, 4.

You will also be rearranging equations (changing the subject of equations). This means moving the parts or *terms* in the formula.

For example,  $a = \frac{b}{c}$  becomes  $a \times c = b$ .

Equations with logs work like this: to rearrange  $\log_{10}A = x$  so that A becomes the subject, move the base number (in this case 10) to the right and change the x to a power, so that you get  $A = 10^x$ .

### Worked examples

#### Question

A solution has a hydrogen ion concentration of  $1.8 \times 10^{-5}$  M. What is the pH of the solution?

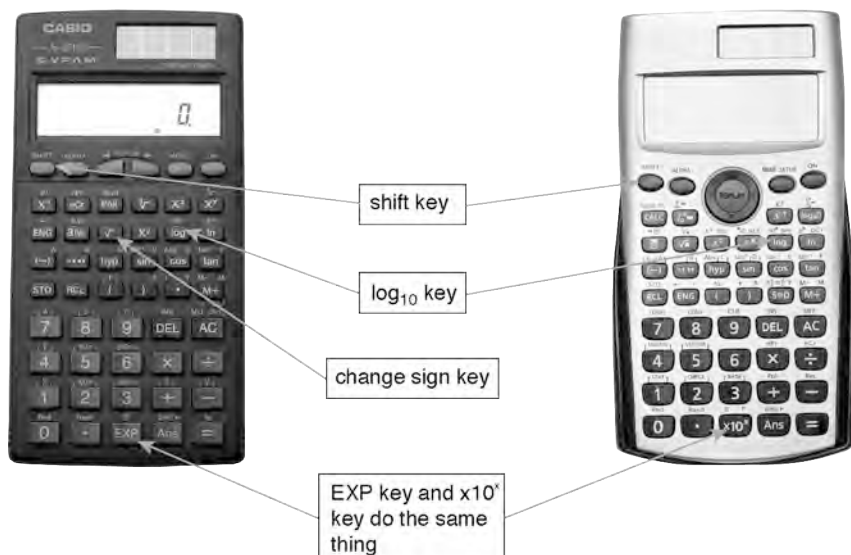
#### Answer

pH and hydrogen ion concentration are related so one may be calculated from the other using the formula  $\text{pH} = -\log_{10}[\text{H}^+]$

$\log_{10}$  is 'log to the base 10' and  $[H^+]$  = concentration of the hydrogen ions in moles per litre.

### Step 1

Type the hydrogen ion concentration figure into the calculator. To do this type 1.8, then press the EXP (or  $\times 10^x$ ) key, type 5 then press the change sign (+/-) key (see Figure 1).



Your display should now show  $1.8^{-05}$

### Step 2

Find  $\log_{10}$  of this hydrogen ion concentration by pressing the log key.

Your display should now show  $-4.7447275$

### Step 3

Press the change sign key (+/-)

Your display should now show 4.744 727 5, which is the pH. Round this to an appropriate number of decimal places, for example to 4.74

### Question

The pH of a solution is 8.6. What is the concentration of hydrogen ions present?

### Answer

To solve this, the formula is rearranged so that  $[H^+]$  becomes the subject of the equation (the thing you are trying to work out). Move the base number (10) to the right and make pH a power so that

$$-\log_{10} [H^+] = \text{pH} \text{ becomes } [H^+] = 10^{-\text{pH}}$$

(As the log was negative, the power has to be negative also.)

### Step 1

Type in the figure for pH so that the display shows 8.6.

**Step 2**

Press the change sign key (+/-) so that the display shows -8.6.

**Step 3**

Remember -8.6 would have been the  $\log_{10}$  of the  $[H^+]$ , so you now have to work out the 'anti-log'. Press the shift key on your calculator, then press the  $\log_{10}$  key. Your display should now show  $2.5118864^{-09}$  which is the value for  $[H^+]$ .

Round this to a sensible number of decimal places, e.g.  $2.5 \times 10^{-9}$ .

**Questions**

- 1 For each equation, rearrange to make A the subject.
  - a  $C = B - A$  (1 mark)
  - b  $\log_{10} A = B$  (1 mark)
  - c  $-\log_3 A = C$  (1 mark)
- 2 Work out the pH of solutions with the following hydrogen ion concentrations.
  - a  $3.6 \times 10^{-3}$  moles per litre (1 mark)
  - b  $1 \times 10^{-7}$  mol dm<sup>-3</sup> (1 mark)
  - c  $10^{-12}$  M (1 mark)
  - d 0.06 mmol per litre (1 mark)
  - e  $10^{-4}$  mmol/L (1 mark)
- 3 Work out the hydrogen ion concentrations of solutions with the following pH values.
  - a 3.8 (1 mark)
  - b 9.6 (1 mark)
  - c 1.8 (1 mark)
  - d 11.34 (1 mark)
  - e 7.2 (1 mark)

**Maths skills links to other areas**

You may also use logarithms and power functions when doing calculations about population growth of microorganisms

**Answers**

4 a  $A = B - C$    b  $A = 10^B$    c  $A = 3^{-C}$  (1 mark each)

5 a 2.44   b 7.0   c 11.0   d 4.2   e 6 (1 mark each)

6 a  $1.58 \times 10^{-4}$  (1 mark)

b  $2.51 \times 10^{-10}$  (1 mark)

c 0.016 (1 mark)

d  $4.57 \times 10^{-12}$  (1 mark)

e  $6.3 \times 10^{-8}$

## Finding rates of reaction using tangents

### Specification reference

- 3.1.4.2
- Maths skill M3.2
- Maths skill M3.6

### Maths Skills for Biology reference

- Rates of reaction 1
- Rates of reaction 2

### Learning objectives

After completing the worksheet you should be able to:

- plot two variables from experimental data
- draw and use the slope of a tangent to a curve as a measure of rate of change.

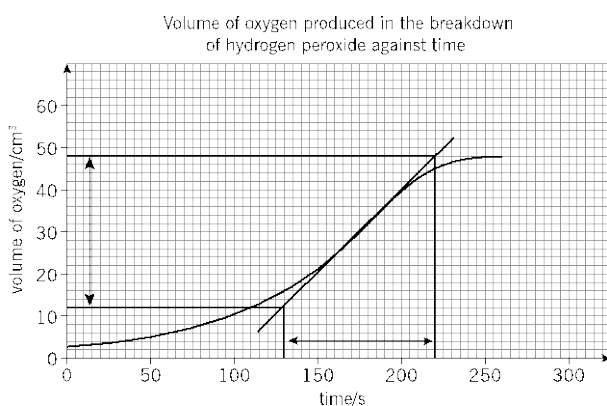
### Introduction

When biological experiments give trends in which the rate of reaction changes, tangents can be used to allow calculation of the rate of reaction at any given point on the curve, for example, the fastest rate.

### Worked example

#### Question

The graph in Figure 1 shows data collected during the reaction between catalase and hydrogen peroxide. The volume of gas collected is plotted against time. What is the maximum reaction rate?



**Figure 1** Graph of volume of oxygen produced in the breakdown of hydrogen peroxide

#### Answer

Step 1

First construct a tangent line by drawing a line using a ruler as shown in Figure 1. The tangent line touches the curve at its steepest point. The exact position of a tangent is achieved by estimation, so the trend line curve is symmetrically diverging from the point at which the tangent line touches it.

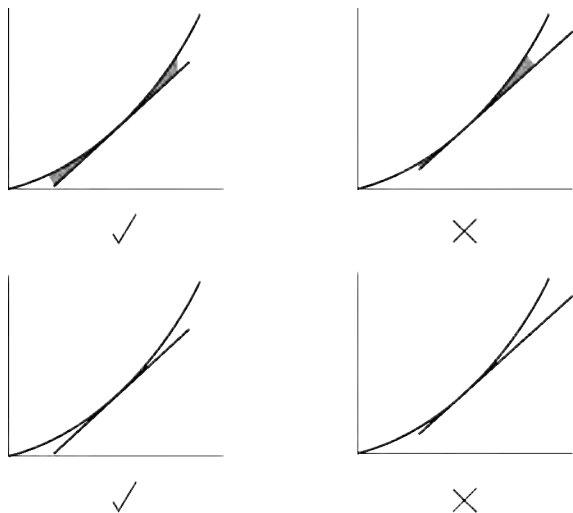


Figure 2 Finding a tangent

Step 2

Next use construction lines to find the values of  $x$  and  $y$  at any two selected points on the tangent, as shown.

Find the change in  $y$  and the change in  $x$  between the two selected points. In the example,  $y$  goes from 12 to 48  $\text{cm}^3$ , a change of 36  $\text{cm}^3$  oxygen while  $x$  changes from 128 to 220 s, a change of 92 s.

Step 3

Calculate the rate by dividing the change in  $y$  by the change in  $x$  ( $dy/dx$ ). In the example:

$$\frac{dy}{dx} = \frac{36}{92} = 0.39 \text{ cm}^3 \text{ s}^{-1}$$

Questions

- The oxygen production by samples of *Elodea* pondweed was investigated at five different concentrations of bicarbonate. All other variables were controlled. The data collected is shown in the table below.

Concentration of bicarbonate (%)	Time (s)					
	30	60	90	120	150	180
	Cumulative volume of oxygen collected ( $\text{mm}^3$ )					
0	0.5	1.5	2.5	4.5	7.0	10.0



1	1.0	2.0	3.5	6.5	10.0	15.0
2	2.5	6.0	11.0	17.0	24.0	40.0
3	6.0	14.0	27.0	41.5	50.5	52.0
4	10.0	25.0	42.5	51.0	52.5	53.0

- a Plot graphs to show the trends in oxygen production at each concentration of bicarbonate. (4 marks)
- b Draw suitable tangents on each curve and use them to calculate maximum rates of reaction at each concentration. (5 marks)
- 2 The table below shows data collected from an experiment in which glucose was being produced by the digestion of starch.

Time (minutes)	Glucose produced (mmol dm <sup>-3</sup> )
5	2
10	6
15	12
20	22
25	25
30	28
35	29

- a Plot the data on a suitable graph. (4 marks)
- b Use a tangent to calculate the maximum rate of reaction. (2 marks)
- c Find the rates of reaction at 8 minutes and at 32 minutes. (2 marks)

## Maths skills links to other areas

You may be asked to use this method of rate calculations when you study movement across cell membranes, water uptake with potometers, photosynthetic rate experiments, and respirometer measurements of oxygen uptake.

**Answers**

- 1 a** Award marks for the following points:  
axes correct orientation, scaled with equidistant divisions;  
both axes labelled with title and units;  
points correctly plotted;  
neat ruler dot-to-dot line or sensible best fit trend line for each trend;  
graphs fit at least half of the space on the paper used; (max 4 marks)
- b** rates are approximately:  
0% =  $0.13 \text{ mm}^3 \text{ s}^{-1}$  (1 mark)  
1% =  $0.19 \text{ mm}^3 \text{ s}^{-1}$  (1 mark)  
2% =  $0.47 \text{ mm}^3 \text{ s}^{-1}$  (1 mark)  
3% =  $0.60 \text{ mm}^3 \text{ s}^{-1}$  (1 mark)  
4% =  $0.63 \text{ mm}^3 \text{ s}^{-1}$  (1 mark)
- 2 a** Award marks for the following points:  
axes correct orientation, scaled with equidistant divisions;  
both axes labelled with title and units;  
points correctly plotted;  
neat ruler dot-to-dot line or sensible best fit trend line for each trend;  
graphs fit at least half of the space on the paper used. (max 4 marks)
- b** (this will depend on the exact positioning of the tangent) Approx  $2.3 \text{ mmol dm}^{-3} \text{ min}^{-1}$ . (2 marks)
- c** (depends on the positioning of the tangent) Approx  $0.9 \text{ mmol dm}^{-3} \text{ min}^{-1}$  and  $0.3 \text{ mmol dm}^{-3} \text{ min}^{-1}$ . (2 marks)

## Calculations of DNA base frequencies

### Specification references

- 3.1.5.1
- Maths skills 0.3

### Maths Skills for Biology references

- Percentages

### Learning outcomes

After completing the worksheet you should be able to:

- use ratios, fractions, and percentages.

### Introduction

A ratio expresses the total contribution of parts, for example a 1 : 1 mixture means equal parts of both quantities. A fraction is an expression of the specific contribution of one part. For example, in a mixture with a ratio of 1 part A : 3 parts B, A is one out of a total of four parts, or  $\frac{1}{4}$ .

A percentage is simply expressing a fraction as a part of 100. For example, the fraction  $\frac{1}{4}$  gives the decimal 0.25, which when multiplied by 100 gives 25%. The molecule DNA has a constant ratio of bases. This is due to the base pairing rule, adenine to thymine (A–T) and guanine to cytosine (C–G). Consequently, for every adenine on one strand there will be a corresponding thymine on the complementary strand, so knowing the proportion of one base will allow the calculation of all the others quite easily.

### Worked example

#### Question

Analysis of a sample of DNA revealed that it was made up of 32% adenine. Calculate the proportion of guanine in the sample.

#### Answer

*Step 1:* There will be the same percentage of thymine and adenine because they form pairs A–T. Find the total for A + T in the sample:  $2 \times 32\% = 64\%$

*Step 2:* Find the total for G + C, which is the remainder of the molecule by subtracting the value for A + T from the total:  $100\% - 64\% = 36\%$

Step 3: Guanine pairs with cytosine on a 1 : 1 ratio, so half the G + C proportion will be guanine. Find the proportion of guanine:  $36\% \div 2 = 18\%$

## Questions

- Convert these fractions to decimals. Give your answer to 2 significant figures.
  - $\frac{3}{12}$  (1 mark)
  - $\frac{16}{56}$  (1 mark)
- What is the fraction of *A* in the following ratios?
  - $5A : 10B$  (1 mark)
  - $11B : 2A$  (1 mark)
- What is the percentage of *A* in parts **a** and **b** of question 2? (2 marks)
- In a sample of DNA, there was 17% thymine.
  - Calculate the percentage of cytosine in the molecule. (1 mark)
  - What is the percentage of purine bases in the sample? (1 mark)
- A sample of DNA from the lambda bacteriophage virus was sequenced and was found to be made up from 48 502 base pairs. 12 336 base pairs were adenine and thymine.
  - How many of the bases in the molecule will be cytosine? (1 mark)
  - Calculate the percentage of the DNA molecule which is made up from guanine. (1 mark)
- Cytosine makes up 29.50% of a DNA sample.
  - Find the percentage of adenine. (1 mark)
  - Predict the percentage of guanine that will be found in a molecule of mRNA that is made by transcribing this DNA sample. Assume that all the bases are equally distributed on each of the two strands in the DNA molecule. (1 mark)
- A sample of messenger RNA was analysed and gave the following base sequence:  
AAGCGGUUACGGAAUAG
  - What is the percentage of pyrimidines in the original coding DNA strand? (1 mark)
  - What proportion of bases in the original coding DNA molecule was contributed by guanine? (1 mark)

## Maths skills links to other areas

You will find a very similar calculation in your second year of A Level when using the Hardy–Weinberg ratio to find allele frequencies.

## Answers

- 
- 8 a 0.25 (1 mark)  
b 0.29 (1 mark)
- 9 a 5/15 (1 mark)  
b 2/13 (1 mark)
- 10 A is 33.3% and 15.4% (2 marks)
- 11 a 33% (1 mark)  
b 50% (1 mark)
- 12 a 36,166 (1 mark)  
b 37.3% (1 mark)
- 13 a 20.5% (keep same number of decimal places as in the question) (1 mark)  
b 7.38% (1 mark)
- 14 a 70.6% (12 of the bases in the coding strand are C + T) (1 mark)  
b 23.5% (8 of the 34 bases in the two strands of the DNA molecule were G) (1 mark)

## Cardiac output

### Specification reference

- 3.3.41
- Maths skill 2.2
- Maths skill 3.1

### Maths Skills for Biology reference

- *Electrocardiograms*

### Learning objectives

After completing the worksheet you should be able to:

- change the subject of an equation
- translate information between graphical, numerical, and algebraic forms.

### Introduction

The individual parts or *terms* in equations are all related. Sometimes you might know all the values of the terms except one. The equation can be re-written so that the unknown term can be calculated. This is called rearranging or *changing the subject* of an equation. A very useful example of this arises during the study of cardiac output.

The different terms are cardiac output, stroke volume, and heart rate. The equation that relates them together is:

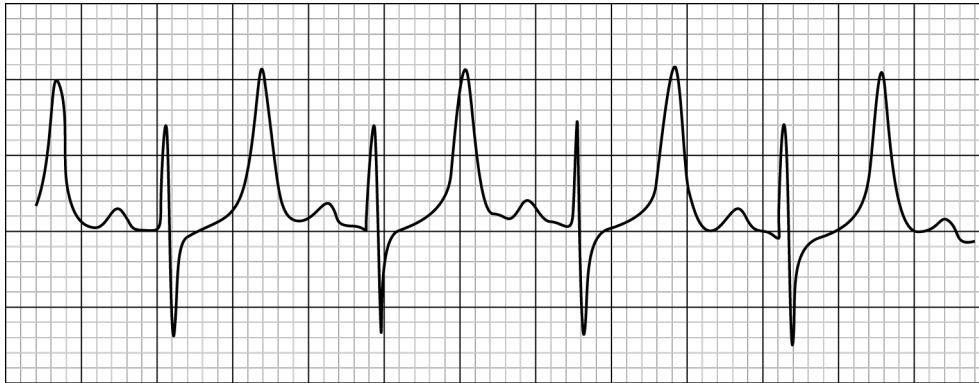
$$\text{cardiac output} = \text{stroke volume} \times \text{heart rate}$$

where cardiac output is the subject of the equation. Changing the subject of the equation means rearranging it so that heart rate can be calculated as:

$$\text{heart rate} = \frac{\text{cardiac output}}{\text{stroke volume}}$$

Cardiac output is the volume of blood being pumped by the heart into the circulatory system in one minute. Usually this measurement relates specifically to the output from the left ventricle. It depends on the stroke volume (the volume of blood pumped out of the ventricle each contraction) and the heart rate.

Heart rate may be measured quickly using pulse counts in beats per minute or by measuring intervals on ECG peaks. Stroke volume is harder to measure precisely and several methods exist. Echocardiograms may be used to measure ventricle volume just before and after one beat and stroke volume measured as the difference between the two. A simpler method is to measure systolic and diastolic blood pressure using a blood pressure monitor and take the difference between the two (pulse pressure) to be equivalent to stroke volume in  $\text{cm}^3$ .

**Worked example****Question****Figure 1**

The stroke volume of the ECG shown in Figure 1 is  $69 \text{ cm}^3$ . Each small square represents  $0.2 \text{ s}$ . What is the cardiac output?

**Answer***Step 1*

Find the heart rate. Begin by measuring the distance between the high peaks. Here it is  $13.5$  small squares.

*Step 2*

Convert this to time in seconds using the scale.

$$13.5 \times 0.2 = 2.7 \text{ s}$$

*Step 3*

Divide to find the number of beats in one minute ( $60 \text{ s}$ ).

$$60 \div 2.7 = 22.2 \text{ beats per minute}$$

*Step 4*

Transfer the numbers into the formula to find the cardiac output (CO).

$$\text{CO} = 69 \times 22.2 \text{ cm}^3 \text{ per minute}$$

$$\text{CO} = 1531.8 \text{ cm}^3 \text{ per minute}$$

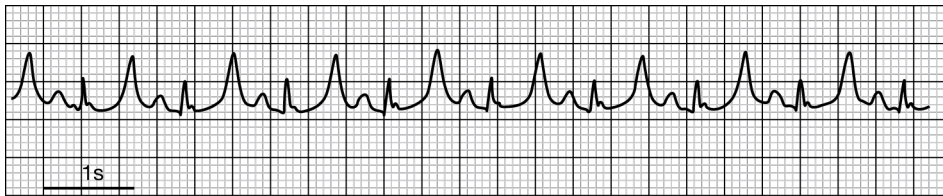
*Step 5*

Convert to  $\text{dm}^3$  per minute by dividing by  $1000$ .

$$1531.8 \div 1000 = 1.53 \text{ dm}^3 \text{ per minute (} 1.53 \text{ dm}^3 \text{ min}^{-1}\text{)}$$

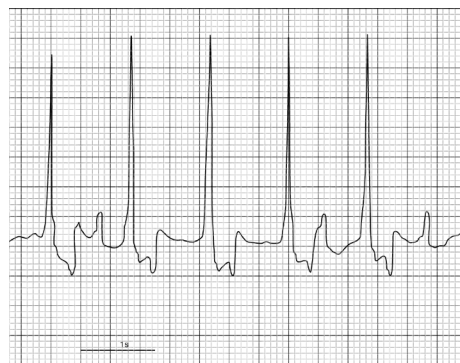
**Questions**

- 1 For each equation, rearrange to make A the subject.
  - a  $C = B - A$  (1 mark)
  - b  $P + Q = A \times B$  (1 mark)
- 2 An individual at rest had a stroke volume of  $75 \text{ cm}^3$  and a heart rate of 60 beats per minute.
  - a Calculate the cardiac output in  $\text{dm}^{-3} \text{ min}^{-1}$ . (1 mark)
  - b Calculate the cardiac output when the same individual exercises with a stroke volume of  $98 \text{ cm}^3$  and a heart rate of 103 beats per minute. (1 mark)
  - c What is the percentage increase in cardiac output? (2 marks)
- 3 Figure 2 shows the ECG of a healthy person at rest.



**Figure 2**

- a What is the heart rate of the person? (1 mark)
- b If the stroke volume of the person is  $80 \text{ cm}^3$ , what is the cardiac output in litres per minute? (1 mark)
- c What would the heart rate of the person be if their CO rose to  $9.6 \text{ dm}^{-3} \text{ min}^{-1}$  and their stroke volume increased by 50%? (2 marks)
- 4 Figure 3 shows the ECG of a person at rest. The same person measured their blood pressure with a portable meter and found it to be 123 / 82 mmHg.
  - a Use the graph to find the person's heart rate. (1 mark)
  - b Use the following formula to estimate the person's stroke volume  
 systolic pressure – diastolic pressure = pulse pressure  
 Assume that pulse pressure = stroke volume in  $\text{cm}^3$ . (1 mark)
  - c Calculate the cardiac output for this person. (1 mark)
  - d When exercising, the person's heart rate rose to 105 beats per minute. What would be the percentage increase in stroke volume required to give them a cardiac output of  $5.6 \text{ dm}^{-3} \text{ min}^{-1}$ ? (2 marks)



**Figure 3**



- 5 During a period of exercise the oxygen demand increased. It was necessary for the blood to deliver  $10 \text{ cm}^3$  of oxygen per minute per 100 g of respiring skeletal muscle.
- Oxygenated blood carries  $20 \text{ cm}^3$  of oxygen per  $100 \text{ cm}^3$ .
  - 80% of the CO is directed at skeletal muscles when exercising heavily.
  - Skeletal muscle accounts for 50% of lean body mass.
- a** Calculate the volume of blood required to deliver sufficient oxygen to the muscles of a person weighing 108 kg during the exercise. (2 marks)
- b** What is the required cardiac output to supply oxygen at this rate? (1 mark)
- c** If the person had a heart rate of 193 bpm during the exercise, what would their stroke volume have been? (1 mark)

### Maths skills links to other areas

You will find the ability to apply and rearrange the subject of equations like this in several places in the specification. For example, calculation of magnifications, working with rates of reaction, diversity indices.

**Answers**

- 6 a**  $A = B - C$  (1 mark)
- b**  $A = \frac{P+Q}{B}$  (1 mark)
- 7 a**  $4.5 \text{ dm}^{-3} \text{ min}^{-1}$  (1 mark)
- b**  $10.1 \text{ dm}^{-3} \text{ min}^{-1}$  (1 mark)
- c** 124.4% (allow one for the correctly calculated increase but incorrect %, two for the correct %) (2 marks)
- 8 a** 52.9 beats per minute (1 mark)
- b**  $4.2 \text{ dm}^{-3} \text{ min}^{-1}$  (1 mark)
- c** 80 beats per minute (one mark for the new stroke volume =  $120 \text{ cm}^3$ , one mark for the correct rate) (2 marks)
- 9 a** 56.6 bpm (1 mark)
- b**  $41 \text{ cm}^3$  (1 mark)
- c**  $2.3 \text{ dm}^{-3} \text{ min}^{-1}$  (1 mark)
- d** 30% (one mark for the change in stroke volume correctly found 12.3, one for the correct %) (2 marks)
- 10 a**  $27 \text{ dm}^3 / 27,000 \text{ cm}^3$  (2 marks if correct, or one mark each correct step, e.g. muscle mass is 54 kg) (2 marks)
- b**  $33\,750 \text{ cm}^3 / 33.8 \text{ dm}^3$  (1 mark)
- c**  $174.9 \text{ cm}^3$

## Species diversity index

### Specification reference

- 3.4.6
- Maths skill 1.1
- Maths skill 1.2
- Maths skill 1.5
- Maths skill 2.3
- Maths skill 2.4

### Maths Skills for Biology reference

- Species diversity

### Learning outcomes

After completing the worksheet you should be able to:

- find arithmetic means
- substitute numerical values into algebraic equations
- solve algebraic equations
- understand the principles of sampling as applied to scientific data.

### Introduction

Species diversity is a calculated measure of the balance between the different species present in a community. High diversity suggests that the community is evenly balanced and likely to be stable, e.g. ancient woodland. Low diversity can be indicative of communities that have over-representation by some species and may be liable to change as the years pass. For example, early stages of succession, such as pioneer communities. Low values may also indicate dominant populations in specialised habitats, for example barnacles on rocks in the upper shore. Suitable data for the calculation depends on random sampling, such as frame quadrats placed using randomly generated grid coordinates.

Diversity is measured by calculating the value of Simpson's index of diversity,  $D$ . This value does not have units but is a fraction of one. If all species had exactly the same populations, then the balance between them is equal and  $D$  will be 1. The more uneven the numbers of species, the smaller  $D$  becomes.

The formula required is  $D = 1 - \left( \sum \left( \frac{n}{N} \right)^2 \right)$

### Worked example

#### Question

Sampling a community gave the following data: species A 22 plants; species B 30 plants; species C 25 plants; species D 23 plants. Calculate the species diversity.

#### Answer

It helps to lay out the data in a table like the one below to make the calculation easy to follow through.

Plant species	Number of each species of plant ( $n$ )	$\frac{n}{N}$	$\left(\frac{n}{N}\right)^2$
<b>A</b>	22	$\frac{22}{100} = 0.22$	0.05
<b>B</b>	30	$\frac{30}{100} = 0.30$	0.09
<b>C</b>	25	$\frac{25}{100} = 0.25$	0.06
<b>D</b>	23	$\frac{23}{100} = 0.23$	0.05
<b>Totals</b>	Total plants found (all species) $N = 100$		$\sum\left(\frac{n}{N}\right)^2 = 0.25$
Species richness	4		

#### Step 1

Find the value of  $N$ , the total number of all plants present. Here, it is  $22 + 30 + 25 + 23 = 100$ , shown in column 2.

#### Step 2

Calculate and fill in the values of  $n/N$ , where  $n$  is the number of plants of each individual species. The results are shown in the third column.

#### Step 3

Square the values of  $n/N$  to find  $\left(\frac{n}{N}\right)^2$  for each plant species. These are shown in the fourth column.

#### Step 4

Add up the values of  $\left(\frac{n}{N}\right)^2$  to find the total, which is  $\sum\left(\frac{n}{N}\right)^2$ . In this example, the result is 0.25, shown at the bottom of column 4.

#### Step 5

Finally, subtract the value of  $\sum\left(\frac{n}{N}\right)^2$  from 1 to give a value for  $D$ . Note: this will always be a fraction of 1. Here  $1 - 0.25 = 0.75$ , which is a high diversity. In this

example, the reason is that all species are evenly represented and the community is consequently well balanced.

Note: this method will work just as well using % cover data from quadrats.

**Questions**

1 Three examples of data are shown in the table below.

Moorland		Rocky shore		Potato field	
Species	Mean % cover per quadrat	Species	Mean number per quadrat	Species	Mean % cover per quadrat
Milkwort	5	Limpet	82	Potato	89
Carnation grass	16	Grey top shell	12	Dandelion	8
Purple moor grass	29	Toothed top shell	36	Clover	13
Bell heather	6	Edible periwinkle	35	Nettle	7
Starry sedge	5	Chiton	4	Dock	14
Common heather	48	Beadlet anemone	23		
Tormentil	13	Snakelocks anemone	7		
Sheep's fescue	82	Dog whelk	34		
Sheep's sorrel	11	Barnacle	197		
Mat grass	23				

- a Calculate the species diversity for each site. (6 marks)
- b Comment on the diversity of the three sites. (3 marks)

2 Samples were taken from three randomly located positions in a freshwater pond. The results are shown in the table below.

Number of specimens of each species found in freshwater pond samples			
Species	Sample number		
	1	2	3
Water boatman	12	9	5
Great diving beetle	4	0	5
Water flea	56	78	36
Freshwater shrimp	35	12	24
Ramshorn pond snail	0	11	2
Dragonfly nymphs	9	10	7

- a Calculate the mean number of each species present. (3 marks)
- b Calculate the value of  $D$  for this pond by using the formula  $D = 1 - \left( \sum \left( \frac{n}{N} \right)^2 \right)$  (2 marks)
- c Suggest how the random sampling positions might have been determined. (2 marks)
- d Discuss how the data might have been different had the sampling **not** been random. (3 marks)
- e Dragonfly nymphs were counted together and not as separate species. Suggest how this might have influenced the value of  $D$  that you have calculated. (3 marks)

### Maths skills links to other areas

You will need to be confident with the calculation of arithmetic means in many areas of the specification where numerical data are used. Understanding the importance of adequate random sampling and how it affects the data you collect is very important, for example when trying to prove a trend, such as in enzyme reaction rate experiments, or when using data in statistical tests to try and prove significance of patterns.

**Answers**

- 1** 2 marks per correct value. Allow one for a correct intermediate stage if not, e.g.  $N = 238$  in the moorland.
- a** Moorland  $D = 0.81$ ; Rocky shore  $D = 0.73$ ; Potato field  $D = 0.51$
- b** Any **three** of the following:  
Moorland in the most diverse. Potato field is the least diverse;  
balance of species is more even in the moorland and shore samples;  
moorland and shore likely to be more stable communities/later in the succession;  
potato field is least stable/most likely to change. (max 3 marks)
- 2 a** water boatman 9, diving beetle 3, flea 57, shrimp 24, snail 4, dragonfly nymphs 9. Lose one mark per wrong answer.
- b** 0.64 (1 mark for correct workings; 1 mark for correct answer)
- c** Any **two** of the following:  
A grid or system of tapes laid out; reference to some kind of laser range finder system  
random coordinates generated;  
samples taken at those coordinates. (max 2 marks)
- d** May have sampled only species from one part of the pond;  
e.g. only the edge/only the deep water;  
species from some parts of the pond would not be collected/ORAs;  
species richness may have been lower. (max 3 marks)
- e**  $D$  would be inaccurate/unrepresentative;  
might have reduced the value of  $D$ ;  
because lumping would give a bigger  $(n/N)^2$  value;  
would not allow comparison between this and other sites. (max 3 marks)

## Standard deviation

### Specification reference

- 3.4.7
- Maths skill M1.10

### Specification reference

- Standard deviation

### Learning objectives

After completing this worksheet you should be able to:

- calculate standard deviation using a formula
- relate the calculated values of standard deviation to error bars on graphs
- compare means of data.

### Introduction

Standard deviation shows us how much variation there is between the data and the mean value. It is therefore a measure of spread in the data and gives an indication of the likely reliability of the mean value.

It is possible to calculate standard deviation automatically in a spreadsheet, but you may also be asked to calculate it using a given formula when analysing data.

The formula to use is  $s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$

In the formula:

- $s$  is standard deviation
- $x$  refers to each measured value in the dataset
- $\bar{x}$  refers to the arithmetic mean
- $n$  is the number of values measured
- $\sum$  = 'the sum of', meaning add together the values indicated to the right of this symbol.

Standard deviation becomes more meaningful as the number of measurements in the dataset increases. In general, having fewer than five repeats will not give a meaningful value for standard deviation.



**Worked example**

**Question**

As part of an investigation into enzyme action, 10 measurements were made of the gas volume generated by an enzyme reaction at 20 °C. The volumes collected (cm<sup>3</sup>) were as follows:

45, 48, 12, 39, 52, 47, 45, 45, 36, 41

- a Calculate the mean volume of gas.
- b Calculate the standard deviation of this mean.
- c Find the graph error bar values for the mean value.

**Answer**

- a Find the mean by adding together all the values of  $x$  and dividing by the number of values.

$$45 + 48 + 12 + 39 + 52 + 47 + 45 + 45 + 36 + 41 = 410$$

$$410 \div 10 = 41$$

so  $\bar{x} = 41$

It is easier to do the standard deviation calculation, if you use a table like the one below.

Values of $x$ (volumes in cm <sup>3</sup> )	Values of $x - \bar{x}$	Values of $(x - \bar{x})^2$
45	4	16
48	7	49
12	-29	841
39	-2	4
52	11	121
47	6	36
45	4	16
45	4	16
36	-5	25
41	0	0
Number of values $n = 10$		Total of this column is $\sum(x - \bar{x})^2$
The mean, $\bar{x}$ , is 41		$\sum(x - \bar{x})^2 = 1124$

- b Find the values of  $x - \bar{x}$ . These are shown in the second column of the table. Find the square of each of the  $x - \bar{x}$  values, e.g. for the first row:  $4 \times 4 = 16$ . The completed values are shown in the third column of the table.

**Step 4**

Find  $\sum(x - \bar{x})^2$  by totalling the values of  $(x - \bar{x})^2$ , i.e. the sum of the values in the third column, which is 1124.

**Step 5**

Divide the total by  $n - 1$  (number of values  $- 1$ , here it is  $10 - 1 = 9$ ).

$$1124 \div 9 = 124.9$$

**Step 6**

Take the square root of this number (calculator button marked  $\sqrt{\quad}$ ) to complete the calculation.

$$\sqrt{124.9} = 11.2$$

so SD = 11.2

Note: this is quite a high value because an anomalous outlier point has been included (i.e. the value 12, which is clearly out of line with the general pattern).

**c** When the mean value is plotted as a point on a graph the standard deviation can be used to mark on an error bar. This is simple to do.

**Step 7**

Find the value for the top of the error bar by adding the standard deviation to the mean:

$$41 + 11.2 = 52.2$$

**Step 8**

Find the value for the bottom of the error bar by subtracting standard deviation from the mean:

$$41 - 11.2 = 29.8$$

**Step 9**

Plot the point on the graph (41). Mark the positions of the top and bottom of the error bar using small pencil dash marks directly above and below the plotted point, in this example at 52.2 and 29.8 as measured from the y-axis. Use a ruler and sharp pencil to draw a vertical line through your plotted point between these limits.

Note that an error bar is different from a range bar, which is drawn by simply using the highest and lowest values as top and bottom rather than the standard deviation. This is not as useful, because an outlying value will have a bigger effect and make the data appear more scattered. Try testing this visually by plotting this example on a sketch graph.

**Questions**

1 The table shows the results of an experiment with pondweed producing oxygen bubbles at different temperatures.

Temperature / °C	Rate of reaction / bubbles per minute				
0	3.0	1.0	3.0	6.0	5.0
5	15.0	13.0	15.0	12.0	13.0
10	20.0	23.0	24.0	16.0	21.0
15	22.0	14.0	21.0	12.0	31.0
20	26.0	32.0	36.0	31.0	39.0

- a Find the mean value for each row in the table. (5 marks)
  - b Calculate the standard deviation of each mean value. (5 marks)
  - c Comment on the reliability of the mean values. (2 marks)
  - d What are the upper and lower limits of the error bar you would plot for the mean value at 5 °C? (2 marks)
- 2 An identical experiment was carried out by five different students. Not all were as thorough as they should have been, so they were inconsistent in how many repeat measurements they took. Their results are shown in the table.

Student	Results of each repeat done									
1	14.0	23.0	9.0							
2	17.0	19.0	27.0	20.0	31.0					
3	24.0	18.0	21.0	29.0	19.0	26.0				
4	19.0	25.0	24.0	18.0	25.0	26.0	23.0	22.0		
5	22.0	18.0	27.0	21.0	22.0	23.0	25.0	23.0	23.0	24.0

- a Find the mean value for each student's data. (5 marks)
- b Calculate the standard deviation of each student's data. (5 marks)
- c Compare the mean values by describing and comment on the patterns in these calculations. (5 marks)

**Maths skills links to other areas**

Whenever you collect data you should consider using standard deviation to test the dispersion of the results and add error bars to your graphs as a normal part of your processing. To save time you can use spreadsheets to find standard deviation automatically.

**Answers**

- 1 a** The mean values for five temperatures rows are 3.6, 13.6, 20.8, 20.0, and 32.8 (one mark deducted per error. Values must have same decimal places as the original data). (3 marks)
- b** **One** mark for each correct value:  
The SD values are 1.9, 1.3, 3.1, 7.5, 5.0. (5 marks)
- c** Any **two** from:  
The mean, at 5 °C is the most reliable/at 15 °C is the least reliable. (1 mark)  
Because SD is smallest at 5 °C/ largest at 15 °C. (1 mark)  
Data at 5 °C are less spread/have less variation/are less dispersed from the mean/OR A. (1 mark)
- 2** The limits are 14.9 and 12.3 (one mark for each correct limit). (2 marks)
- 3 a** **One** mark for each correct answer:  
The means for the students from 1 to 5 are 15.3, 22.8, 22.8, 22.8, 22.8. (5 marks)
- b** **One** mark for each correct answer:  
The SD for the students from 1 to 5 are 7.1, 5.9, 4.3, 2.9, 2.4. (5 marks)
- c** Any **five** from:  
The true mean is (close to) 22.8;  
All students with five or more repetitions succeed in getting an accurate mean  
As the number of repeats increases the standard deviation becomes smaller  
There is less, spread/dispersion, in the data/outliers have less effect;  
Student 1 has insufficient data to calculate an accurate mean  
Error bars become smaller as the number of repeats increase  
Although the means for students 2–5 are the same, students 4 and 5 have the most certainty.  
Students 4 and 5 will be able to plot the points with the greatest certainty.  
There is no significant benefit in repeating this experiment more than eight times. (max 5 marks)

## Using the chi-squared test to analyse genetic crosses

### Specification references

- 3.7.1
- MS 1.4, MS 1.9, MS 2.3, MS 2.4

### Maths Skills for Biology references

- *Chi-squared goodness of fit test in genetics*

### Learning outcomes

After completing this worksheet you should be able to:

- apply the chi-squared test to compare observed and expected results
- understand the role of chance in genetic inheritance.

### Introduction

The results of genetic crosses can be predicted using Punnett squares. In these diagrams there is an equal expectation that each combination will occur just by chance, because every gamete has an equal chance of being the one that is randomly combined at fertilisation.

## Worked example

### Question

In the *Drosophila* fly, the allele for long wing is dominant over that for vestigial wing. A cross between two heterozygous long winged flies gave 68 offspring in the ratio 55 long winged and 13 vestigial winged. Use the chi-squared test to determine the significance of the difference between observed and expected results.

### Answer

The null hypothesis is that any difference between the observed and expected results is due to chance.

*Step 1:* Work out the possible results of the cross using a Punnett square.

Long wing = L; vestigial wing = l

Parental genotypes Ll × Ll

Possible gametes L or l × L or l

F1 combinations

	♂ gametes	
♀ gametes	L	I
L	LL	LI
I	LI	II

Ratio expected 3 Long winged : 1 vestigial winged

Step 2: The formula for the chi-squared ( $\chi^2$ ) test is  $\chi^2 = \sum \frac{(O - E)^2}{E}$

It is easiest to lay out the calculation in a table like this:

	Observed (O)	Expected (E)	(O-E)	(O-E) <sup>2</sup>	$\frac{(O - E)^2}{E}$
Long winged flies	55	51	4	16	0.314
Vestigial winged flies	13	17	-4	16	0.941

Step 3: Total the values in the last column to find chi-squared,  $0.314 + 0.941 = 1.255$

Step 4: Find the degrees of freedom (df) in the data using

$$df = \text{number of categories} - 1$$

Here two categories (long or vestigial wing) so  $2 - 1 = 1$  df

**Step 5:**

df	0.99	0.95	0.90	0.50	0.10	0.05	0.01	0.001	df
1	0.00016	0.0039	0.016	0.46	2.71	3.84	6.63	10.83	1
2	0.02	0.10	0.21	1.39	4.60	5.99	9.21	13.82	2
3	0.12	0.35	0.58	2.37	6.25	7.81	11.34	16.27	3
4	0.30	0.71	1.06	3.36	7.78	9.49	13.28	18.46	4
5	0.55	1.14	1.61	4.35	9.24	11.07	15.09	20.52	5
6	0.87	1.64	2.20	5.35	10.64	12.59	16.81	22.46	6
7	1.24	2.17	2.83	6.35	12.02	14.07	18.48	24.32	7
8	1.65	2.73	3.49	7.34	13.36	15.51	20.09	26.12	8
9	2.09	3.32	4.17	8.34	14.68	16.92	21.67	27.88	9
10	2.56	3.94	4.86	9.34	15.99	18.31	23.21	29.59	10

P values

P < 0.5 and > 0.1

1 degree of freedom uses this row

The calculated value lies between these values

Look up the value in the chi-squared distribution table, reading along the row corresponding with your df until you find the position of your calculated value. The *p* values at the tops of the columns indicate the probability that the results were just due to chance, i.e. there is no underlying causal factor skewing the outcome in some way. In this example the *p* values are between 50% and 10%, so chance alone did cause the outcome.

You can only decide that chance is not causing the outcome and that there is another causing factor where the *p* value is less than or equal to 5% (0.05). The values of chi-squared corresponding to the 5% probability level are often referred to as **critical values**.

**Questions**

1 A character for stem height in pea plants has a dominant allele giving tall stems and a recessive allele giving short stems. In a cross between two heterozygous individuals, 315 tall and 101 short plants were produced. Use the chi-squared test to determine the significance of the difference between observed and expected results.

.....

.....

..... (3 marks)

2 When two pink antirrhinum plants produced seed, the 96 F1 plants grew in the ratio 26 red : 51 pink : 19 white.

a Calculate the value of chi-squared.

.....

..... (2 marks)

- b** The critical values of chi-squared at two degrees of freedom are 4.60 at the 10% level and 5.99 at the 5% level. Comment on your value of chi-squared in the light of these values.

.....  
 .....  
 ..... (3 marks)

- 3** In a species of plant, leaf pigments are controlled by two alleles, G gives green leaves, g gives white leaves and the heterozygous Gg has variegated leaves (partially green and partially white). Two variegated plants were crossed and the seeds collected and grown. The resulting F1 plants were examined after two weeks and the leaf colours were 16 green leaved plants and 21 variegated plants.

- a** Determine the significance of the difference between the observed and expected results using the chi-squared test.

.....  
 .....  
 ..... (3 marks)

- b** Explain the absence of white leaved plants in the F1.

.....  
 ..... (2 marks)

**Maths skills links to other areas**

The chi-squared test is used to test how well observed data fit with expectation. It is used in situations where the expected outcome can be clearly predicted, genetics being the example most often used. You will also need to apply statistical tests to compare the difference between sets of data (*t* test) and to test the correlation between variables (correlation coefficient). Worksheets on these two tests are also available on the Kerboodle site.



**Answers**

- 1** Chi-squared = 0.115 with one degree of freedom (1 mark)
- $p < 0.9$  and  $> 0.5$  /  $< 90\%$  and  $> 50\%$  mean that the difference between observed and expected values is due to chance (1 mark)
- the difference is due to chance alone (1 mark)
- (3 marks)
- 2 a** Chi-squared = 1.396 (2 marks)
- b** The value is less than the critical values at both the 5% and the 10% levels of probability (1 mark)
- Therefore the  $p$  value is greater than 10% (1 mark)
- The results are due only to chance (1 mark)
- (3 marks)
- 3 a** Chi-squared = 14.514 with two degrees of freedom (1 mark)
- $p < 0.001$  /  $< 0.1\%$  shows that the data are due to chance alone (1 mark)
- The distribution is very unlikely to be due to chance (only 0.1%), so some factor is causing the outcome to be skewed (1 mark)
- (3 marks)
- b** White leaves contain no chlorophyll so they are unable to photosynthesise; they all die (2 marks)