edexcel

INTERNATIONAL ADVANCED LEVEL

BIOLOGY

TEACHER MATHEMATICS SUPPORT

Pearson Edexcel International Advanced Subsidiary in Biology (XBI11)
Pearson Edexcel International Advanced Level in Biology (YBI11)
First teaching September 2018
First examination from January 2019
First certification from August 2019 (International Advanced Subsidiary) and August 2020 (International Advanced Level)

Pearson

Contents

Introduction

Mathematical skills are an essential part of IAS and IAL Biology.

In order to be able to develop their skills, knowledge and understanding in science, students need to have been taught, and to have acquired competence in, the appropriate areas of mathematics relevant to biology as indicated in the tables that follow.

The assessment of quantitative skills will include at least 10% level 2 or above mathematical skills for biology in the context of the examinations.

The following tables illustrate where these mathematical skills may be developed and could be assessed. Please note that those skills shown in bold type would only be tested in the full A level course.

	Mathematical skills	**Exemplification of mathematical skill in the context of A Level biology (assessment is not limited to the examples given below)**
(i) A.0 - arithmetic and numerical computation		
A.0.1	Recognise and make use of appropriate units in calculations	Candidates may be tested on their ability to: • convert between units, e.g. mm^3 to cm^3 as part of volumetric calculations • work out the unit for a rate, e.g. breathing rate
A.0.2	Recognise and use expressions in decimal and standard form	Candidates may be tested on their ability to: • use an appropriate number of decimal places in calculations, e.g. for a mean • carry out calculations using numbers in standard and ordinary form, e.g. use of magnification • understand standard form when applied to areas such as size of organelles • convert between numbers in standard and ordinary form • understand that significant figures need retaining when making conversions between standard and ordinary form, e.g. 0.0050 mol dm^{-3} is equivalent to 5.0×10^{-3} mol dm^{-3}
A.0.3	Use ratios, fractions and percentages	Candidates may be tested on their ability to: • calculate percentage yields • calculate surface area to volume ratio • use scales for measuring • represent phenotypic (monohybrid and dihybrid crosses)
A.0.4	Estimate results	Candidates may be tested on their ability to: • estimate results to sense check that the calculated values are appropriate
A.0.5	**Use calculators to find and use power, exponential and logarithmic functions**	**Candidates may be tested on their ability to:** • **estimate the number of bacteria grown over a certain length of time**
(ii) A.1 - handling data		
A.1.1	Use an appropriate number of significant figures	Candidates may be tested on their ability to:

	Mathematical skills	Exemplification of mathematical skill in the context of A Level biology (assessment is not limited to the examples given below)
		● report calculations to an appropriate number of significant figures given raw data quoted to varying numbers of significant figures ● understand that calculated results can be reported only to the limits of the least accurate measurement
A.1.2	Find arithmetic means	Candidates may be tested on their ability to: ● find the mean of a range of data, e.g. the mean number of stomata in the leaves of a plant
A.1.3	Construct and interpret frequency tables and diagrams, bar charts and histograms	Candidates may be tested on their ability to: ● represent a range of data in a table with clear headings, units and consistent decimal places ● interpret data from a variety of tables, e.g. data relating to organ function ● plot a range of data in an appropriate format, e.g. enzyme activity over time represented on a graph ● interpret data for a variety of graphs, e.g. explain electrocardiogram traces
A.1.4	Understand simple probability	Candidates may be tested on their ability to: ● use the terms probability and chance appropriately ● understand the probability associated with genetic inheritance
A.1.5	Understand the principles of sampling as applied to scientific data	Candidates may be tested on their ability to: ● analyse random data collected by an appropriate means, e.g. calculate an index of diversity to compare the biodiversity of a habitat
A.1.6	Understand the terms mean, median and mode	Candidates may be tested on their ability to: ● calculate or compare the mean, median and mode of a set of data, e.g. height/mass/size of a group of organisms
A.1.7	Use a scatter diagram to identify a correlation between two variables	Candidates may be tested on their ability to: ● interpret a scattergram, e.g. the effect of life style factors on health
(ii) A.1 - handling data (continued)		
A.1.8	Make order of magnitude calculations	Candidates may be tested on their ability to: ● use and manipulate the magnification formula

	Mathematical skills	Exemplification of mathematical skill in the context of A Level biology (assessment is not limited to the examples given below)
		$\text{magnification} = \frac{\text{size of image}}{\text{size of real object}}$
A.1.9	Select and use a statistical test	Candidates may be tested on their ability to select and use: • the Chi squared test to test the significance of the difference between observed and expected results • the Student's t-test • the correlation coefficient
A.1.10	Understand measures of dispersion, including standard deviation and range	Candidates may be tested on their ability to: • calculate the standard deviation • understand why standard deviation might be a more useful measure of dispersion for a given set of data, e.g. where there is an outlying result
A.1.11	Identify uncertainties in measurements and use simple techniques to determine uncertainty when data are combined	Candidates may be tested on their ability to: • calculate percentage error where there are uncertainties in measurement
(iii) A.2 – algebra		
A.2.1	Understand and use the symbols: =, <, <<, >>, >, ∝, ~.	No exemplification required
A.2.2	Change the subject of an equation	Candidates may be tested on their ability to: • use and manipulate equations, e.g. magnification
A.2.3	Substitute numerical values into algebraic equations using appropriate units for physical quantities	Candidates may be tested on their ability to: • use a given equation e.g. a formula to calculate an index of diversity $D = \frac{N(N-1)}{\Sigma n(n-1)}$
A.2.4	Solve algebraic equations	Candidates may be tested on their ability to: • solve equations in a biological context, e.g. cardiac output = stroke volume × heart rate
A.2.5	**Use logarithms in relation to quantities that range over several orders of magnitude**	**Candidates may be tested on their ability to:** • **use a logarithmic scale in the context of microbiology, e.g. growth rate of a microorganism such as yeast**
(iv) A.3 – graphs		

	Mathematical skills	Exemplification of mathematical skill in the context of A Level biology (assessment is not limited to the examples given below)
A.3.1	Translate information between graphical, numerical and algebraic forms	Candidates may be tested on their ability to: ● understand that data may be presented in a number of formats and be able to use these data, e.g. dissociation curves
A.3.2	Plot two variables from experimental or other data	Candidates may be tested on their ability to: ● select an appropriate format for presenting data, bar charts, histograms, graphs and scattergrams
A.3.3	Understand that $y = mx + c$ represents a linear relationship	Candidates may be tested on their ability to: ● predict/sketch the shape of a graph with a linear relationship, e.g. the effect of substrate concentration on the rate of an enzyme-controlled reaction with excess enzyme
A.3.4	**Determine the intercept of a graph**	Candidates may be tested on their ability to: ● **read off an intercept point from a graph, e.g. compensation point in plants**
A.3.5	Calculate rate of change from a graph showing a linear relationship	Candidates may be tested on their ability to: ● calculate a rate from a graph, e.g. rate of transpiration
A.3.6	Draw and use the slope of a tangent to a curve as a measure of rate of change	Candidates may be tested on their ability to: ● use this method to measure the gradient of a point on a curve, e.g. amount of product formed plotted against time when the concentration of enzyme is fixed
(v) A.4 - geometry and trigonometry		
A.4.1	Calculate the circumferences, surface areas and volumes of regular shapes	Candidates may be tested on their ability to: ● calculate the circumference and area of a circle ● calculate the surface area and volume of rectangular prisms, of cylindrical prisms and of spheres ● e.g. calculate the surface area or volume of a cell

This booklet is intended to give guidance on the mathematical skills, with worked examples and sample questions, with answers. It is expected that development of mathematical skills would be integrated, where appropriate, within teaching of the theoretical and practical specification content.

The accompanying student booklet includes the same background to the topics, with worked examples and practice questions. The answers to the practice questions are not included in the student booklet.

Core practicals and opportunities for development of mathematical skills

The following tables are intended to suggest how the core practical can be used to develop mathematical skills in the presentation and interpretation of data derived from practical work.

Specification A (Salters-Nuffield)

Topic	Core practicals	Mathematical skills
1	Use a semi-quantitative method with Benedict's reagent to estimate the concentrations of reducing sugars and with iodine solution to estimate the concentrations of starch, using colour standards.	Bar charts and histograms.
	Investigate the vitamin C content of food and drink.	Converting volumetric units; using an appropriate number of decimal places.
2	Investigate membrane properties including the effect of alcohol and temperature on membrane permeability.	Tabulating data; plotting an appropriate graph and interpreting results.
	Investigate the effect of temperature, pH, enzyme concentration and substrate concentration on the initial rate of enzyme-catalysed reactions.	Tabulating data and plotting an appropriate graph; drawing a tangent and determining initial rate.
3	Use a light microscope to make observations and labelled drawings of suitable animal cells Use a graticule with a microscope to make measurements and understand the concept of scale	Understanding scale and conversion of units; making order of magnitude calculations.
	Prepare and stain a root tip squash to observe the stages of mitosis.	Tabulating data and calculating percentages, e.g. mitotic index; calculating means.
4	Use a light microscope to: (i) make observations, draw and label plan diagrams of transverse sections of roots, stems and leaves (ii) make observations, draw and label cells of plant tissues (iii) identify sclerenchyma fibres, phloem, sieve tubes and xylem vessels and their location.	Understanding scale and conversion of units; making order of magnitude calculations.
	Determine the tensile strength of plant fibres.	Tabulating data and selecting an appropriate graph; calculating means and conversion of units.

Topic	Core practicals	Mathematical skills
	Investigate the antimicrobial properties of plants, including aseptic techniques for the safe handling of bacteria.	Tabulating data, measuring and using an appropriate number of significant figures, calculating means.
5	Investigate the effects of light intensity, light wavelength, temperature and availability of carbon dioxide on the rate of photosynthesis using a suitable aquatic plant.	Tabulating data, calculating means, handling units and plotting appropriate graphs, calculating rates.
	Carry out a study on the ecology of a habitat, such as using quadrats and transects to determine distribution and abundance of organisms, and measuring abiotic factors appropriate to the habitat.	Constructing and interpreting frequency tables and diagrams, bar charts and histograms; understanding principles of sampling; selecting and using a statistical test; using an index of diversity.
	Investigate the effects of temperature on the development of organisms (such as seedling growth rate or brine shrimp hatch rates), taking into account the ethical use of organisms.	Working out the units for a rate; calculating means and understanding measures of dispersion.
6	Investigate the rate of growth of microorganisms in a liquid culture, taking into account the safe and ethical use of organisms	Use calculators to find and use power, exponential and logarithmic functions; use logarithms in relation to quantities that range over several orders of magnitude; understanding logarithmic values and plot appropriate graphs.
	Investigate the effect of different antibiotics on bacteria.	Tabulating data; measurement and using an appropriate number of significant figures; calculation of means.

7	Use an artificial hydrogen carrier (redox indicator) to investigate respiration in yeast.	Tabulating data, calculating means, handling units and plotting appropriate graphs, calculating rates.
	Use a simple respirometer to determine the rate of respiration and RQ of a suitable material (such as germinating seeds or small invertebrates).	Tabulating data; selecting an appropriate graph; calculating rate; converting units.
	Investigate the effects of exercise on tidal volume, breathing rate, respiratory minute ventilation and oxygen consumption using data from spirometer traces.	Making measurements and converting units; solving equations.
8	Investigate the production of amylase in germinating cereal grains.	Measurements and calculation of means, plotting appropriate graphs.

Maths Skills for Biologists

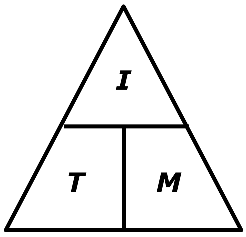
Size and magnification

The use of a microscope allows small structures to be enlarged for observation and study. It is important to know how the size of the image viewed relates to the true size of the structure. Students may be asked to measure the size of an image on a photograph or on a diagram.

The relationship can be expressed as:

$$\text{True size of structure}\,(T) = \frac{\text{Image size}\,(I)}{\text{magnification of microscope}\,(M)}$$

However, questions may ask for the magnification to be worked out when given the other two sets of data, etc. To help do this, you may wish to use the following.



Using 'TIM's pyramid' can help to rearrange the equation as necessary. For example:

$T = \frac{I}{M}$ or $M = \frac{I}{T}$ or $I = T \times M$

Sadly, it is not quite as simple as this because of the units involved. The size of the image is likely to be measured in centimetres (cm) or millimetres (mm) whilst the true size of the structure is likely to be in micrometres (μm).

For example, a mitochondrion shown on an electron micrograph has a maximum image length of 21 mm when measured with a ruler. However, its true size is 7 μm long.

To calculate the magnification of the electron microscope used to produce this image:

Step 1: convert the units so they are all the same

In this case 21 mm is also 21×10^{-3} m and 7 μm is 7×10^{-6} m

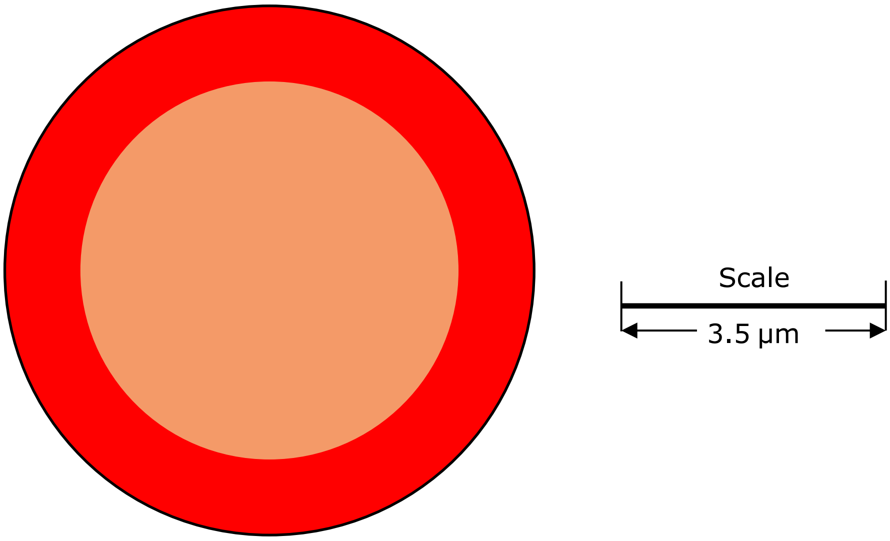
Step 2: Apply 'TIM's pyramid' in the appropriate arrangement

In this case it is $M = \frac{I}{T} = \frac{21 \times 10^{-3}\ \text{m}}{7 \times 10^{-6}\ \text{m}}$

This gives an answer (magnification) of ×3000. There are no units.

Worked example

The diagram shows a red blood cell.



Using the diagram and the scale provided, calculate the magnification of this red blood cell.

Answer

Step 1

Find the true size.

Using the scale provided, the red blood cell is found to be 7 μm across, as its diameter is exactly 2× the length of the scale line.

The true size is 7×10^{-6} m.

The image size, as measured from the diagram is 7.0 cm, which is 7×10^{-2} m.

Step 2

$$M = \frac{I}{T} = \frac{7 \times 10^{-2}\ \text{m}}{7 \times 10^{-6}\ \text{m}} = \times 10\,000 \text{ magnification}$$

Practice question

An investigation was carried out to study the diameter of blood vessels in the wing of a flying mammal.

Using a magnification of ×2000, the mean image diameter for arteries found in the wing was 10 cm.

a Calculate the mean diameter of these arteries in mm and put your answer in the table.

b The mean diameter of the arterioles was found to be 14% that of the arteries and the mean vein diameter is 60% greater than the arteries. Using this additional information, complete the table.

Blood vessel type	Mean diameter of blood vessel / mm
Artery	**0.050**
Arteriole	**0.007**
Vein	**0.080**

Percentage change

The percentage change is, perhaps, the most common calculation asked for when numerical data is presented in a table or graphical form.

The percentage change could be an increase, where the answer would be positive, or a decrease, where the answer would be negative.

To calculate the percentage change the following equation can be used.

$$\left(\frac{\text{final data set} - \text{initial data set}}{\text{initial data set}}\right) \times 100$$

So if the initial data set is 22 and the final one is 40 then the percentage change is

$$\left(\frac{40-22}{22}\right) \times 100 = \left(\frac{18}{22}\right) \times 100 = 81.8\% \text{ (or } 82\%)$$

Worked example

The table shows the effect of increasing axon diameter on the speed of impulse transmission in myelinated neurones.

Axon diameter / μm	**Impulse speed / m s^{-1}**
1	2.2
2	8.3
3	15.0
4	20.0
5	25.0

Calculate the percentage change in the speed of impulse as the axon diameter increased from 2 μm to 3 μm.

Answer

Percentage change is $\left(\frac{15-8.3}{8.3}\right) \times 100 = \left(\frac{6.7}{8.3}\right) \times 100 = 80.7\%$

Practice question

A student carried out an investigation into the effect of temperature on the growth of seedlings. Seeds of sea plantain were germinated at 18°C. As soon as they germinated, the seedlings were placed in three temperature-controlled rooms at 10°C, 14°C and 18°C. They were allowed to grow for 50 days. Samples of seedlings were taken at 5-day intervals and their mean dry masses were recorded.

The table shows the results of this investigation.

Complete the table by calculating the percentage increase in mean dry mass of the seedlings, for all three temperatures, from day 5 to day 50.

Day	Mean dry mass / mg		
	10°C	**14°C**	**18°C**
5	2	2	2
10	3	4	6
15	4	6	12
20	7	12	20
25	10	19	34
30	13	25	47
35	17	31	85
40	20	40	109
45	24	55	164
50	28	80	210
Percentage change			

Answer

For 10°C = 1300%

For 14°C = 3900%

For 18°C = 10 400%

Teacher comment: some students may not appreciate that a percentage increase can be greater than 100%.

Measures of centre

There are three different measures of centre that may be considered in biology:

- mean
- median
- mode.

Mean

The mean is commonly referred to as 'average'. This is the measure of centre that is most commonly encountered by A level students.

The values are added together and this total is divided by the number of values.

Worked example

Many insects jump to avoid predators. The table below shows the distance jumped by locusts of two different body mass cohorts.

Locust	Distance jumped by two different body mass cohorts / m	
	1.0–1.5 g	2.0–2.5 g
1	0.60	0.30
2	0.35	0.60
3	0.30	0.80
4	0.55	0.45
5	0.35	0.60
6	0.35	0.40
7	0.45	0.80
Mean		

Complete the table, by calculating the mean for each of the two locust body mass cohorts.

Answer

Calculation for the 1.0–1.5 g cohort is:

$$\frac{(0.60 + 0.35 + 0.30 + 0.55 + 0.35 + 0.35 + 0.45)}{7} = \frac{2.95}{7} = 0.42 \text{ m}$$

While the calculation for the 2.0–2.5 g body size cohort is:

$$\frac{(0.30 + 0.60 + 0.80 + 0.45 + 0.60 + 0.40 + 0.80)}{7} = \frac{3.95}{7} = 0.56 \text{ m}$$

Practice question

An investigation was carried out to study whether a marine snail preferred to inhabit red seaweed or brown seaweed. The table below shows the results of this investigation.

a Complete the table by calculating the mean number on the red seaweed.

Quadrat sample	Number on red seaweed	Number on brown seaweed
1	58	12
2	62	14
3	53	13
4	56	13
5	68	54
6	55	11
Mean	**58.7**	12.6

b Describe how the mean value for the brown seaweed was calculated.

Answer

{Sample 5 / 54} was an anomaly so was not included in the calculation.
(remainder / eq} were summated.
(total) divided by 5.

Teacher comment: could discuss with students what constitutes an anomaly.

Median

The median is essentially the centre of a set of values arranged in rank order.

Step 1

Arrange the data in a rank order if not already done.

Step 2

Find the middle value. If there are an odd number of values, such as 7, then the middle one or median one is the 4^{th}. This is because there are 3 data sets with a lower value and three with a higher value.

However, if there are an even number of values, for example 10, then the mean value of the 5^{th} and 6^{th} values needs to be found as they are the middle two. There are equal numbers of values before the 5^{th} value and after the 6^{th} value.

Worked example

The table below shows the abundance of moss, measured as percentage cover, in one habitat.

Sample	Percentage cover of moss (%)
1	65
2	30
3	55
4	60
5	84
6	35
7	59
8	69

Using the table, find the median percentage cover of moss in the habitat.

Show your working.

Answer

Step 1: Rank the values according to size.

30 35 55 59 60 65 69 84

Step 2: Find the middle two values as there is an even number of values in this series.

30 35 55 59 60 65 69 84

Then work out the mean of these two values: $\frac{(59 + 60)}{2} = 59.5$

Practice question

An investigation was carried out to study the percentage recovery of an immobilised enzyme after one usage. This was repeated a total of eight times with a new batch of immobilised enzyme in each case. The table shows the results of this investigation.

Sample number	Percentage recovery (%)
1	100
2	75
3	78
4	100
5	98
6	77
7	80
8	64

The median percentage recovery for this immobilised enzyme after one usage is 79%. Explain how this value was found.

Answer

The percentage recovery values were ranked in (ascending) order.

{an even number of / 8} data sets present so 4^{th} and 5^{th} data sets used / 78% and 80% used.

Mean of 78% and 80% calculated.

Mode

The mode is the most frequently occurring value in a series.

Worked example

The table below shows some data relating to the ABO blood group alleles found in human populations on different islands.

Allele	Frequency of allele occurrence			
	Cuba	**Iceland**	**Madagascar**	**New Zealand**
A	0–5	15–20	15–20	15–20
B	0–5	5–10	15–20	5–10
O	90–100	70–80	60–70	70–80

a Use this data to complete the table, by placing a tick in the box that shows the allele with the mode for each island population.

Island population	Allele A	Allele B	Allele O
Cuba			
Iceland			
Madagascar			
New Zealand			

Answer

If we consider Cuba, then the most common allele is O as between 90 and 100% of the population possess this allele. If we then look at the other islands, again the O allele is most common, though not as frequent as in Cuba.

b Which of the three island populations has the mode for allele B?

Answer

This question asks for a study of the B allele row in the first table and identification of the country with a population showing the highest occurrence of this allele. The answer is Madagascar with a frequency of between 15 and 20% of the population.

If only allele A is considered for populations found on Iceland, Madagascar and New Zealand, then all have the same level of occurrence. In this circumstance, there is no mode.

Practice question

Part **a** is mode and part **b** is an extension

a An investigation was carried out to compare the size of eggs laid by one species of lizard. Egg length was measured for a number of different lizards and a tally chart of some of the data is presented in the table.

Length / mm	**Frequency of occurrence**
26	III
28	~~IIII~~ ~~IIII~~ IIII
30	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~
32	~~IIII~~ ~~IIII~~ ~~IIII~~ IIII
34	~~IIII~~ ~~IIII~~ III
36	~~IIII~~ II
38	III

Using the information provided, state the mode length of egg.

Answer

30

b Using the information provided, calculate the difference between the mode and the mean. Show your working.

Answer

Mean calculated by (38 × 3) + (36 × 7) + (34 ×13) + (32 × 19) + (30 × 20) + (28 × 14) + (26 × 3) = 2486

Divided by 3 + 7 + 13 + 19 + 20 + 14 + 3 = 79

Mean = $\frac{2486}{79}$ = 31.5

Difference = 31.5 – 30 = 1.5

Temperature coefficient (Q_{10})

Q_{10} is the difference in the rate of an enzyme-controlled reaction as the temperature changes by 10°C, especially at temperatures below the optimum temperature.

To calculate Q_{10} the following equation can be used.

$$Q_{10} = \frac{\text{rate of enzyme activity at } T + 10^0\text{C}}{\text{rate of enzyme activity at } T^0\text{C}}$$

If temperature *T*°C is 18°C,
then *T* + 10°C is 28°C

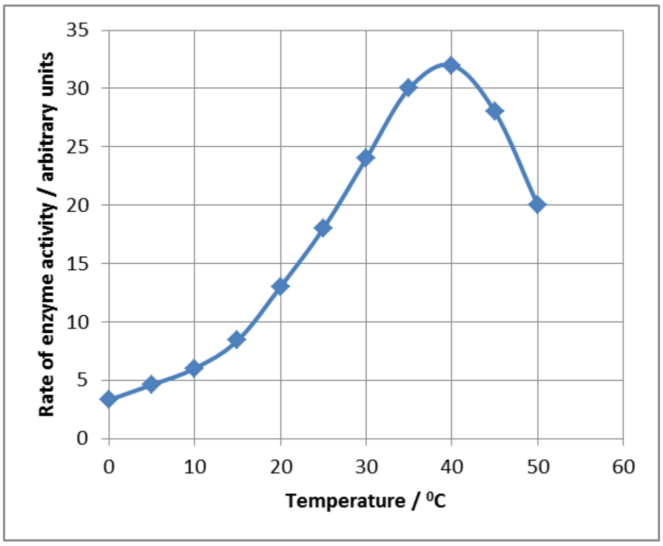
If the rate at *T*°C is 12 units of product produced and at *T* + 10°C 22 units of product produced, then

$$Q_{10} = \frac{22}{12} = 1.83$$

This means that with this 10°C rise in temperature the enzyme produced 1.83 times as much product.

Worked example

The graph below shows an enzyme-controlled reaction.



a Using the graph, calculate Q_{10} as the temperature increases from 20 to 30°C. Show your working.

Answer

Rate of enzyme activity at 20°C is 13 AU and at 30°C it is 24 AU.

So $Q_{10} = \frac{24}{13} = 1.8$

b What is Q_{10} if the temperature increases from 10°C to 20°C?

A 2.0 **B** 2.2 **C** 2.4 **D** 2.6

Answer

The answer is **B**. A common misconception is to simply divide 20°C by 10°C and produce an answer of 2, rather than finding the rate at each of these two temperatures.

Practice question

If Q_{10} for enzyme *X* is 2.00, complete the table.

Temperature / °C	Enzyme *X* activity / arbitrary units
10	10
20	**20**
30	**40**
40	**80**

Teacher comment: as Q_{10} = 2, then the rate of reaction should double. Some students are likely to get the value for 20°C correct, but perhaps for the wrong reason. This will be evident by their subsequent values.

Productivity: GPP, NPP and R

Gross primary productivity (GPP), net primary productivity (NPP) and respiration (R) are inter-related. They can be linked together using the following equation.

$$GPP = NPP + R$$

Needless to say, it is quite possible that students would be given any two from the three components and be required to calculate the third. To do this, the equation may need to be rearranged. The following may help:

1 To calculate R use R = GPP − NPP

2 To calculate NPP use NPP = GPP − R

If, for example, the GPP of an ecosystem is found to be 87 000 kJ m^{-2} $year^{-1}$ and the respiration in the same ecosystem is 50 000 kJ m^{-2} $year^{-1}$ then the NPP is:

$$NPP = 37\,000 \text{ kJ m}^{-2} \text{ year}^{-1}$$

Worked example

A study was carried out to find the respiration per year in two different evergreen forests. To do this, scientists collected data on the NPP and GPP of the forests. The results are given in the table. Productivity is expressed here as grams of carbon fixed per square metre per year (g of C m^{-2} $year^{-1}$).

Type of evergreen forest	NPP / g of C m^{-2} $year^{-1}$	GPP / g of C m^{-2} $year^{-1}$
Broadleaf	1200	2700
Needle	440	820

What is the respiration per year in the two different forests?

Answer

As the same process is being asked for both types of forest, the equation is

$$R = GPP - NPP$$

So respiration per year for the broadleaf evergreen forest is

$$2700 - 1200 = 1500 \text{ g of C m}^{-2} \text{ year}^{-1}$$

For the needle evergreen forest it is

$$820 - 440 = 380 \text{ g of C m}^{-2} \text{ year}^{-1}$$

Practice question 1

The ratio of NPP : GPP can be used by scientists when studying productivity in different types of land. Use the data below to work out the NPP : GPP ratio for these two types of land.

Grassland data: GPP = 400 g of C m^{-2} $year^{-1}$ and R = 140 g of C m^{-2} $year^{-1}$

Scrubland data: NPP = 230 g of C m^{-2} $year^{-1}$ and R = 130 g of C m^{-2} $year^{-1}$

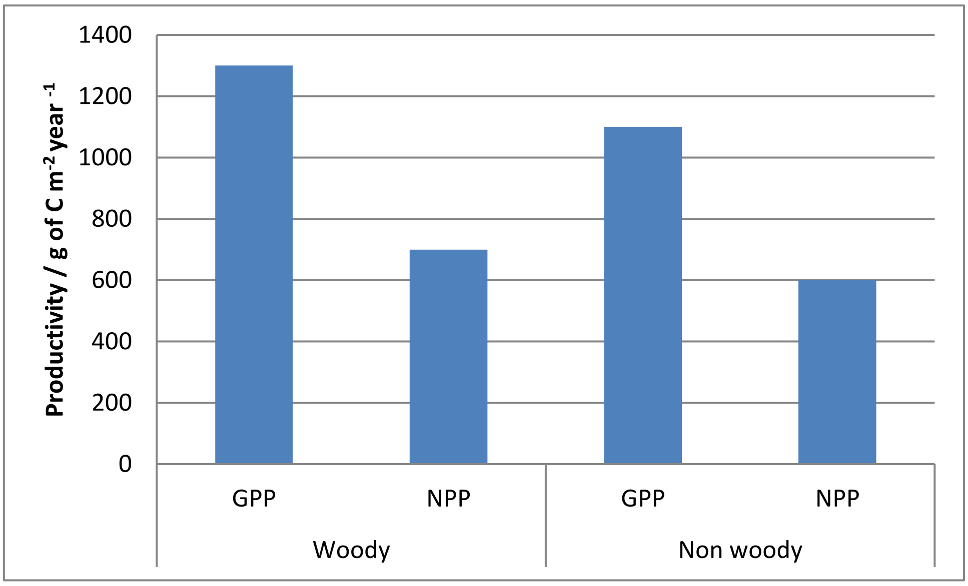
Answer

For the grassland the NPP = 400 – 140 = 260 g of C m^{-2} $year^{-1}$. This leads to a ratio of NPP : GPP of 260 : 400 or 0.65.

For the scrubland the GPP = 220 + 130 or 350 g of C m^{-2} $year^{-1}$. This leads to a ratio of NPP : GPP of 220 : 350 or 0.63.

Practice question 2

The graph shows the GPP and NPP for two types of savannah, woody and non-woody.



Using the graph, calculate the difference in the respiration per year between the two different types of savannah. Show your working.

Answer

Woody savannah respiration = 1300 – 700 = 600 g of C m^{-2} $year^{-1}$

Non woody savannah respiration = 1100 – 600 = 500 g of C m^{-2} $year^{-1}$

NB: Allow some leeway on values read off graph.

Answer = 600 – 500 = 100 g of C m^{-2} $year^{-1}$

Animal population estimation: Lincoln Index

The Lincoln Index is a type of mark-recapture technique used to estimate the population size of mobile animals. Although not mentioned on the specification, it is included here for completeness. All the animals collected on one occasion are marked and counted (M_1). After an appropriate time, another collection in the same area is carried out and the number of animals caught is found (T). Also the number of marked individuals in this second collection are counted (M_2).

To estimate the population of this animal the following equation can be used.

$$\text{Total population} = \frac{(M_1 \times T)}{M_2} = 1.8$$

So if $M_1 = 50$, $M_2 = 15$ and $T = 300$ then the total population is

$$\frac{(50 \times 300)}{15} = \frac{15000}{15} = 1000 \text{ individuals}$$

Worked example

A student wished to estimate the number of one species of aquatic snail in a pond. She collected 110 on day 1 and marked the shell of them all with a small spot of waterproof paint. On day 2 she collected all the snails she could find. She found 35 with paint on them and 230 without.

Calculate the population size in the pond for this species.

Answer

$M_1 = 110$, $M_2 = 35$ and $T = 230$

Her population size was $\frac{(110 \times 265)}{35} = 833$ snails

Teacher comment: In this question, the number recaptured on day 2 is not only the 230 unmarked individuals but also the 35 marked ones.

Practice question

A study was carried out to consider whether the population size of a malarial mosquito changed between the dry season and the wet season in Africa. The student performed a mark recapture technique in the same area in the dry season and in the wet season of the same year. The results of her study are shown in the table.

Season	Number marked	Number recaptured	Number recaptured with mark on
Dry	88	89	4
Wet	175	227	4

Using the data provided, state whether the population was increasing or decreasing between the dry season and the wet season. Give reasons for your answer.

Answer

Increasing Dry population = (88 × 89) ÷ 4 = 1958
Wet population = (175 × 227) ÷ 4 = 9931 (or 9931.3 but 0.3 disregarded)

Habitat biodiversity using an index of diversity (D)

This technique is used to help assess the diversity of a habitat by taking into account both species richness (number of different species) and the abundance of each of these species. The index of diversity is useful for comparing the biodiversity of two (or more) habitats. A higher index indicates a higher diversity.

The following formula represents one method for calculating an index of diversity (D).

$$D = \frac{N(N-1)}{\Sigma n(n-1)}$$

Where:

- Σ = sum of
- N = total number of individuals of all species
- n = number of individuals of each species.

Perhaps the easiest way to appreciate this is to work through an example.

Worked example

The table shows the numbers of different animals found in a quadrat on a rocky shore.

Name of animal	**Number of individuals of each species (n)**	**$n(n-1)$**
Sea anemone	12	12(12 – 1) = 132
Fish	1	1(1 – 1) = 0
Shrimps	5	5(5 – 1) = 20
TOTALS	N = 12 + 1 + 5 = 18	$\Sigma n(n-1)$ = 132 + 0 + 20 = 152

The index of diversity of this habitat can now be calculated as:

$$\frac{(18 \times 17)}{152} = \frac{306}{152} = 2.01$$

Practice question

Consider two habitats each with species in them called A, B and C.

Calculating the index of diversity for each habitat and work out which habitat is more diverse.

Habitat 1

Species	**Number of individuals of each species (*n*)**	***n*(*n* – 1)**
A	50	**2450**
B	12	**132**
C	28	**756**
TOTALS	**90**	**3338**

Answer

The index of diversity (*D*) of habitat 1 can now be calculated as:

$$D = \frac{(90 \times 89)}{3338}$$

$$= 2.40$$

Habitat 2

Species	**Number of individuals of each species (*n*)**	***n*(*n* – 1)**
A	37	**1332**
B	24	**552**
C	29	**812**
TOTALS	**90**	**2696**

Answer

The index of diversity (*D*) of habitat 2 is:

$$D = \frac{(90 \times 89)}{2696}$$

$$= 2.97$$

The index of diversity for habitat 2 is higher than the index for habitat 1, so we can conclude that the diversity of habitat 2 is greater than for habitat 1.

The Hardy–Weinberg principle

The Hardy–Weinberg equation is used in population genetics to calculate the frequency of alleles and genotypes. For a population to be in equilibrium, and meet the assumptions of the Hardy-Weinberg principle, the following conditions must be satisfied:

- the population is large
- random breeding
- no mutations
- no selection
- no migration into or out of the population.

The Hardy–Weinberg principle applies to monohybrid inheritance, where a character is controlled by a single gene (e.g. **A**). The genotypes in this population would, therefore, be **AA**, **Aa** and **aa**.

We use the following two equations to calculate the allele and genotype frequencies:

$$p + q = 1 \text{ (or 100\% of the population) [equation 1]}$$

where p represents the frequency of the allele **A** and q represents the frequency of the allele **a**

and

$$p^2 + 2pq + q^2 = 1 \text{ (or 100\% of the population) [equation 2]}$$

where p^2 represents the frequency of the genotype **AA**, $2pq$ represents the frequency of the genotype **Aa** and q^2 represents the frequency of the genotype **aa**.

Worked example

In tobacco plants, the production of chlorophyll is controlled by a single allele, **A**. Seedlings with the genotype **aa** lack chlorophyll and die soon after germination.

In a sample of 600 seedlings, 132 lacked chlorophyll. Calculate the expected number of heterozygous seedlings in this sample.

Answer

From these results, we know that the frequency of seedlings with the genotype **aa** is 132/600 = 0.22 (22%) This represents q^2 in equation 2, so the frequency of the allele **a** is:

$$\sqrt{0.22} = 0.469 \ (= q)$$

We can now find the frequency of the allele **A**, using equation 1:

$$1 - 0.469 = 0.531 \ (= p)$$

We now know the values of p (0.531) and q (0.469), so using equation 2, the frequency of heterozygotes is:

$$2pq = 2 \times 0.531 \times 0.469 = 0.498$$

This represents 49.8% of the total number of seedlings, that is, 49.8% of 600 which is 299 seedlings (rounded up to the nearest whole number).

Teacher comment: it is a good idea to check that the total frequency of both alleles is 1 and that the total frequency of all genotypes is also 1 (or 100%).

Practice question

In fruit flies, wing length is controlled by a single gene. Fruit flies with the homozygous recessive genotype have short wings. In a sample of 350 fruit flies, 73 were found to have short wings.

Use the Hardy-Weinberg principle to calculate the expected number of fruit flies in this sample with the homozygous dominant genotype.

Answer

The frequency of flies with the genotype is $\left(\frac{73}{350}\right) \times 100 = 20.86\%$ (or 0.2086).

This represents q^2, so $q = \sqrt{0.2086} = 0.4567$

The frequency of p will be $1 - 0.4567 = 0.5433$

The frequency of flies with the homozygous dominant genotype is:

$p^2 = (0.5433)^2 = 0.2951$

So the expected number of flies with this genotype is $350 \times 0.2951 = 103$ (to the nearest whole number).

Standard deviation

The standard deviation gives a measure of the spread of a normal distribution around the mean value. If data are tightly clustered around the mean, the standard deviation will be small. If data are more widely spread, then the standard deviation will be larger.

The following formula will be provided in examination questions and is usually used to calculate the standard deviation:

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}}$$

Where:

- S is the standard deviation
- Σ = sum of
- X = each measurement (e.g. width or mass)
- n = the total number of measurements.

This is known as the 'sample standard deviation' and is used when taking a sample from a much larger population.

Worked example

The table shows the lengths of a sample of 20 mollusc shells.

Calculate the standard deviation for this sample.

Mollusc	**shell length / cm (X)**	**X^2**
1	3.2	10.24
2	5.1	26.01
3	2.6	6.76
4	4.6	21.16
5	3.8	14.44
6	6.5	42.25
7	5.0	25.0
8	3.6	12.96
9	6.9	47.61
10	4.8	23.04
11	4.6	21.16
12	4.4	19.36
13	4.0	16.0
14	4.4	19.36
15	4.8	23.04
16	5.9	34.81
17	6.0	36.0
18	4.9	24.01
19	4.4	19.36
20	5.2	27.04
Totals (Σ)	94.7	469.61

Answer

Step 1: calculate X^2 values.

Step 2: find ΣX and ΣX^2.

Step 3: substitute values into the formula to find S.

S = 1.06

Practice question

The table shows the number of snails found in 10 quadrats. Each quadrat measured 1 m × 1 m.

Quadrat number	Number of snails found in each quadrat
1	14
2	15
3	14
4	16
5	12
6	14
7	14
8	12
9	16
10	13

a Calculate the mean number of snails per quadrat.

Answer

mean = $\frac{140}{10}$ = 14 snails per quadrat

b Calculate the standard deviation for this sample.

Quadrat number	Number of snails found in each quadrat	x^2
1	14	196
2	15	225
3	14	194
4	16	256
5	12	144
6	14	196
7	14	196
8	12	144
9	16	256
10	13	169
Totals (Σ)	140	1976

Substituting figures into the formula gives a value for the standard deviation of 1.33.

Statistics

Students are expected to know three statistical tests:

- the Chi-squared (χ^2) test
- the Student's t-test
- the correlation coefficient.

The Chi-squared test

This is also sometimes referred to as the 'goodness of fit' test and is used to see how closely experimental results fit expected results. It is particularly useful in genetics, where the expected results can be predicted using Mendelian ratios, such as 9 : 3 : 3 : 1, or 1 : 1. The Chi-squared test can also be used in choice chamber experiments, or prey selection, such as different coloured snail shells selected by thrushes.

The following formula will be provided in examination questions and is used for the Chi-squared test:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where:

- χ^2 is the Chi-squared value
- Σ = sum of
- O = the observed values
- E = the expected values.

Worked example

A breeding experiment was carried out using fruit flies. The table shows the results of this experiment.

Phenotype	long wings and wild-type body colour	long wings and black body colour	vestigial wings and wild-type body colour	vestigial wings and black body colour
Observed numbers (*O*)	76	21	26	5

Use a Chi-squared test to investigate whether these results fit the expected 9 : 3 : 3 : 1 ratio for a cross involving unlinked genes.

Answer

Step 1: Calculate the expected numbers (*E*) of each phenotype, using the expected 9 : 3 : 3 : 1 ratio. The total number of flies in this sample is 128, therefore the expected number of flies with the phenotype long wings and wild-type body colour will be

$$\frac{128}{16} = 8 \times 9 = 72 \text{ (from the ratio parts, } 9 + 3 + 3 + 1 = 16\text{)}$$

Step 2: Calculate the difference between each pair of observed and expected numbers ($O - E$).

Step 3: Square each of these values to obtain $(O - E)^2$.

Step 4: Divide the $(O - E)^2$ values by the expected numbers (*E*).

Step 5: Add these to find the Chi-squared value.

These steps are shown in the following table.

Phenotype	long wings and wild-type body colour	long wings and black body colour	vestigial wings and wild-type body colour	vestigial wings and black body colour
Observed numbers (*O*)	76	21	26	5
Expected numbers (*E*)	72	24	24	8
O* – *E	4	-3	2	-3
$(O - E)^2$	16	9	4	9
$(O - E)^2 \div E$	0.22	0.38	0.17	1.13
Total	0.22 + 0.38 + 0.17 + 1.13 = 1.9			

Interpreting the results of the Chi-squared test.

The larger the value of Chi-squared, the more certain we can be that there is a difference between the observed and expected results.

We need to compare the calculated value of Chi-squared with the critical value at $p = 0.05$ (or 5%) for the appropriate number of degrees of freedom.

In this case, with one row of data, the number of degrees of freedom (df) = the number of columns – 1, i.e. df = 4 – 1 = 3.

Here is a table of critical values for the Chi-squared test.

Degrees of freedom (df)	Levels of significance (*p*)				
	0.05	**0.025**	**0.01**	**0.005**	**0.001**
1	3.84	5.02	6.63	7.88	10.83
2	5.99	7.38	9.21	10.60	13.81
3	7.81	9.35	11.34	12.84	16.27
4	9.49	11.14	13.28	14.86	18.47

The calculated value for Chi-squared (1.9) is **less than** the critical value at $p = 0.05$ for 3 degrees of freedom, so the results of this breeding experiment *are* in the ratio of 9 : 3 : 3 : 1.

Teacher comment: ask students to formulate a null hypothesis for this breeding experiment and to interpret the results in relation to the null hypothesis.

Practice question

A test cross was carried out between two maize plants, one with yellow, smooth cobs and the other with white, wrinkled cobs.

The table shows the results of this cross.

Type of maize cob	**yellow and smooth**	**yellow and wrinkled**	**white and smooth**	**white and wrinkled**
Number of cobs	28	33	30	34

Use a Chi-squared test to investigate whether these results fit the expected 1 : 1 : 1 : 1 ratio.

A

Type of maize cob	**yellow and smooth**	**yellow and wrinkled**	**white and smooth**	**white and wrinkled**
Number of cobs (*O*)	28	33	30	34
Expected numbers (*E*)	31.25	31.25	31.25	31.25
$O - E$	-3.25	1.75	-1.25	2.75
$(O - E)^2$	10.56	3.06	1.56	7.56
$(O - E)^2 \div E$	0.34	0.10	0.05	0.24
Total	0.34 + 0.10 + 0.05 + 0.24 = 0.73			

The calculated value for Chi-squared (0.73) is less than the critical value at $p = 0.05$ (7.81).

We can therefore say that these results **do** fit the expected 1 : 1 : 1 : 1 ratio.

The Student's t-test

This test is used to compare the means of two sets of data, if the data are normally distributed. For example, we could use the t-test to investigate whether the difference between the mean width of shade leaves is significantly different from the mean width of sun leaves.

Incidentally, this test is not named as such because it is widely used by students, but because 'Student' was the pen-name of William Sealy Gosset, who introduced the test in 1908.

The following formula will be provided in examination questions and is used for Student's t-test:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_1^{\,2}}{N_1} + \frac{S_2^{\,2}}{N_2}}}$$

Where

- $\overline{x_1}$ = the mean of the first set of data
- $\overline{x_2}$ = the mean of the second set of data
- S_1 = the standard deviation of the first set of data
- S_2 = the standard deviation of the second set of data
- N_1 = the number of measurements in the first set of data
- N_2 = the number of measurements in the second set of data

Calculations of mean and standard deviation are explained on pages 16 and 31

Worked example

An investigation was carried out to compare the shell lengths of a species of mollusc from two sites, A and B. Ten specimens were collected at random from each site and length of each shell was measured and recorded.

The table below shows the results of this investigation.

Shell lengths / mm	
Site A	**Site B**
35	15
70	18
45	19
50	23
55	26
60	27
64	32
68	34
62	29
78	40
Mean for Site A = 58.7	Mean for Site B =
Standard deviation for Site A = 12.76	Standard deviation for Site B = 7.80

a Calculate the mean shell length for Site B.

Answer

The mean shell length for Site B is 26.3 mm

Teacher comment: it is important to include units, where appropriate, with numerical answers.

b Using the formula for Student's t-test, calculate the value of t for these results.

Answer

Step 1: Calculate the difference between the means.

Step 2: Square each of the standard deviations. Divide S^2 by N for both sets of data and add together.

Step 3: Find the square root of the total from ***Step 2***.

Step 4: Divide the difference between the means by the square root from ***Step 3***, to find the t value.

Note: if the t value is negative, ignore the minus sign to convert it to a positive number. In this example, t = 6.85.

Interpreting the results of the t-test.

We need to compare the t-value with the critical value at p = 0.05 (or 5%) for the appropriate number of degrees of freedom (df).

Calculate the number of degrees of freedom, using the following formula.

$$df = N_1 + N_2 - 2$$

In this example, df = 10 + 10 – 2 = 18

Here is an extract from a table of critical values for a t-test, with 18 degrees of freedom.

Degrees of freedom (df)	**Levels of significance (*p*)**				
	0.05	**0.02**	**0.01**	**0.002**	**0.001**
18	2.101	2.552	2.878	3.610	3.922

The calculated value of t (6.85) is **greater** than the critical value at $p = 0.05$ (2.101), so we can say that the difference between the means **is** significant.

Teacher comment: ask students to formulate a null hypothesis for this investigation and to interpret the results in relation to the null hypothesis.

Practice question

Dog's mercury is a plant that grows in shady areas in woodland and in open clearings. An investigation was carried out to determine whether there was a significant difference between the surface area of the leaves from dog's mercury plants growing in the shaded areas and open clearings of a wood.

Seventeen leaves of dog's mercury were collected from plants growing in a shaded area (Site A) and seventeen leaves were also collected from plants growing in an open clearing (Site B).

The surface area of each leaf was measured. The table shows the results of this investigation.

Surface area of leaves / cm^2	
Shaded area (Site A)	**Open clearing (Site B)**
21	15
14	17
16	18
18	17
19	17
21	19
19	13
22	14
18	21
16	13
13	16
22	13
21	16
23	12
19	14
18	12
15	20
Mean for Site A = 18.53	Mean for Site B =
Standard deviation for Site A = 2.87	Standard deviation of Site B = 2.70

a Calculate the mean surface area of the leaves from Site B.

Answer

The mean surface area for the leaves from Site B is 15.71 cm^2.

b State the number of degrees of freedom for this investigation.

Answer

The number of degrees of freedom is 32 (17 + 17 – 2)

c A statistical table showed that the critical value at $p = 0.05$ is 2.04.

Use Student's t-test to determine whether the difference in mean surface area is significant.

Answer

The calculated value for t is 2.96. This value is greater than the critical value at p = 0.05 (2.96 > 2.04). We can therefore say that there is a significant difference between the mean surface areas.

Correlation coefficient

The correlation coefficient is used to test the strength of the relationship between two variables, such as x and y.

The correlation coefficient can vary from − 1, through 0, to +1. A correlation coefficient of -1 indicates a perfect negative correlation and a coefficient of +1 indicates a perfect positive correlation. A correlation coefficient of 0 indicates that there is no relationship between the two variables.

One correlation test is known as **Spearman's rank correlation test**. This is used to test the relationship between two variables, where the data are not normally distributed.

The following formula will be provided in examination questions and is used for Spearman's rank correlation test:

$$r_s = 1 - \frac{6(\Sigma d^2)}{n(n^2 - 1)}$$

Where:

- r_s = the correlation coefficient
- Σ = sum of
- d = the difference between each pair of ranks (explained in the worked example)
- n = the size of the sample (number of pairs of values)

Note: $n(n^2 - 1)$ is sometimes written as $n^3 - n$.

Worked example

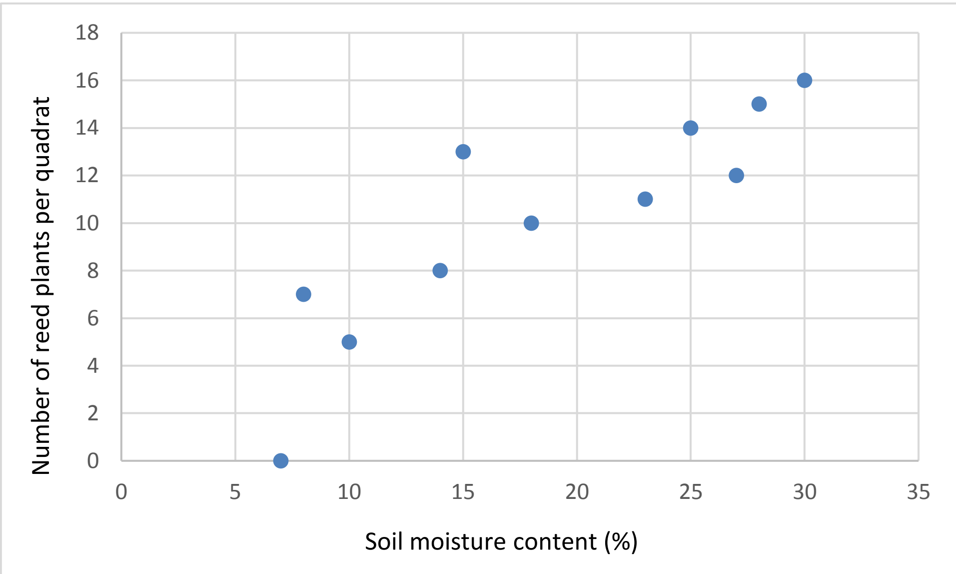
An investigation was carried out into the relationship between the soil moisture content and the number of reed plants growing in the soil.

Eleven samples were taken and the number of reed plants were counted, using a 0.25 m^2 quadrat. The soil moisture content of each sample, expressed as a percentage, was also determined.

The results are shown in the table below.

Sample	1	2	3	4	5	6	7	8	9	10	11
Number of reed plants per quadrat	11	15	10	12	5	0	8	16	7	13	14
Soil moisture content (%)	23	28	18	27	10	7	14	30	8	15	25

It is a good idea to start by plotting a scatter diagram to see if there is any relationship between these two variables.



The scatter diagram shows a relationship between the variables; overall, as the soil moisture content increases, the number of reed plants per quadrat also increases.

contd.

Step 1

Rank the data in each set, so that the highest value is given the ranking '1', the second highest value is given the ranking '2' and so on. If two values are the same (tied numbers) they are each given the mean rank.

Sample	1	2	3	4	5	6	7	8	9	10	11
Number of reed plants per quadrat	11	15	10	12	5	0	8	16	7	13	14
Rank	6	2	7	5	10	11	8	1	9	4	3
Soil moisture content (%)	23	28	18	27	10	7	14	30	8	15	25
Rank	5	2	6	3	9	10	8	1	11	7	4

Step 2

Find the differences (*d*) between each pair of ranks. For example, for the first pair of measurements, *d* = 6 – 5 = 1. Add another row to the table for these *d* values.

Sample	1	2	3	4	5	6	7	8	9	10	11
Number of reed plants per quadrat	11	15	10	12	5	0	8	16	7	13	14
Rank	6	2	7	5	10	11	8	1	9	4	3
Soil moisture content (%)	23	28	18	27	10	7	14	30	8	15	25
Rank	5	2	6	3	9	10	8	1	11	7	4
d	1	0	1	2	1	1	0	0	-2	-3	-1

Step 3

Square the *d* values. Add another row to the table for these d^2 values.

Sample	1	2	3	4	5	6	7	8	9	10	11
Number of reed plants per quadrat	11	15	10	12	5	0	8	16	7	13	14
Rank	6	2	7	5	10	11	8	1	9	4	3
Soil moisture content (%)	23	28	18	27	10	7	14	30	8	15	25
Rank	5	2	6	3	9	10	8	1	11	7	4
d	1	0	1	2	1	1	0	0	-2	-3	-1
d^2	1	0	1	4	1	1	0	0	4	9	1

contd.

Step 4

Calculate $\sum d^2$ by adding all the d^2 values.

$\sum d^2 = 1 + 0 + 1 + 4 + 1 + 1 + 0 + 0 + 4 + 9 + 1 = 22$

Step 5

Substitute the figures into the formula to find r_s.

$r_s = 0.9$

Interpreting the results of the correlation test

We need to compare the calculated value for r_s with the critical value at $p = 0.05$, for the appropriate number of pairs of values (n).

Here is an extract from a table of critical values for the Spearman rank correlation coefficient, at $p = 0.05$.

Number of pairs of values (*n*)	4	5	6	7	8	9	10	11	12
Critical values	1.00	0.90	0.83	0.79	0.74	0.68	0.65	0.61	0.59

The calculated value for the correlation coefficient, 0.9, is greater than the critical value in the table, for 11 pairs of values (0.61). We can therefore say that there is a significant positive correlation between the soil moisture content and the number of reed plants.

Note: If there is an inverse relationship between two variables, the correlation coefficient will have a negative value, for example, -0.85. When interpreting the table of critical values, the sign is ignored and if the calculated correlation coefficient is larger than the critical value, we say that there is a significant inverse (or negative) relationship between the two variables.

Teacher comment: ask students to formulate a null hypothesis for this investigation and to interpret the results in relation to the null hypothesis.

Practice question

An investigation was carried out into the relationship between length and mass in a species of tuna, caught in the Pacific Ocean.

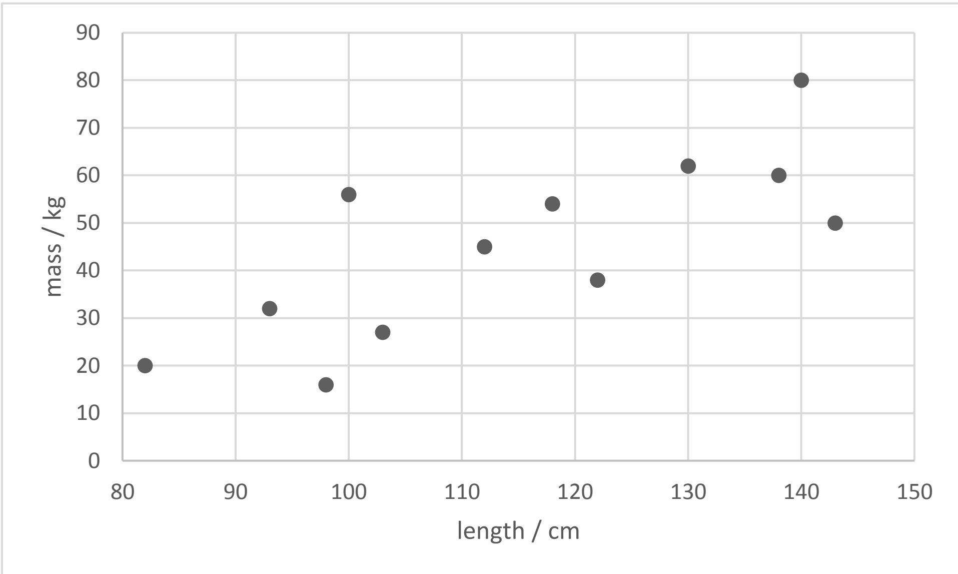
Twelve specimens were caught and the length and mass of each fish was recorded.

The table below shows the results of this investigation.

Fish number	1	2	3	4	5	6	7	8	9	10	11	12
Length / cm	100	112	130	143	103	98	93	118	140	138	122	82
Mass / kg	56	45	62	50	27	16	32	54	80	60	38	20

a Plot a scatter diagram of the data and describe the relationship.

Answer



The scatter diagram indicates that there is a positive correlation between mass and length of the fish.

Teacher comment: the fish are caught at random, so the two variables here could be plotted the other way round i.e. with mass on the x-axis

contd.

b Calculate the Spearman's rank correlation coefficient for the data.

Fish number	1	2	3	4	5	6	7	8	9	10	11	12
Length / cm	100	112	130	143	103	98	93	118	140	138	122	82
Rank	9	7	4	1	8	10	11	6	2	3	5	12
Mass / kg	56	45	62	50	27	16	32	54	80	60	38	20
Rank	4	7	2	6	10	12	9	5	1	3	8	11
d	5	0	2	-5	-2	-2	2	1	1	0	-3	1
***d*²**	25	0	4	25	4	4	4	1	1	0	9	1

Answer

$\sum d^2 = 78$

$n = 12$

Substituting these figures into the formula for Spearman's rank correlation coefficient gives: $r_s = 0.73$.

c What does your calculated value indicate about the relationship between these two variables?

Answer

The critical value for r_s, at $p = 0.05$, for 12 pairs of values, is 0.59.

The calculated value for r_s (0.73) is greater than 0.59. We can therefore say that there is a significant positive correlation between length and mass in this sample of fish.

Further questions

Question 1

Ultraviolet light (UV) can be used to kill microorganisms. A student carried out an investigation into the effect of UV light on the survival of bacteria.

The student spread bacteria evenly on three agar plates. Each plate was then exposed to UV light for one minute. The student then placed the agar plates in an incubator.

The procedure was repeated for exposure times of 2, 3, 4 and 5 minutes.

After 48 hours, the number of bacterial colonies on each plate were recorded.

The results are shown below.

Exposure time / min	**Number of bacterial colonies**		
1	302	282	322
2	187	215	231
3	174	108	129
4	70	94	82
5	37	47	21

a Write a suitable null hypothesis for this investigation.

Answer

There is no significant correlation between time of exposure to UV light and the mean number of bacterial colonies.

Or: the correlation between time of exposure to UV light and the mean number of bacterial colonies is not significant.

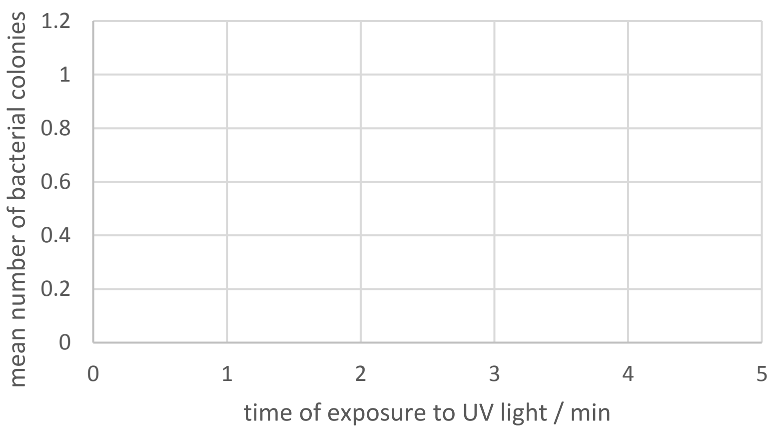
b Complete the following table, by calculating the mean number of bacterial colonies for each UV exposure time.

Exposure time / min	**Mean number of bacterial colonies**
1	**302**
2	**211**
3	**137**
4	**82**
5	**35**

contd.

c Plot a scatter diagram to show the relationship between the mean number of bacterial colonies and the exposure time.

Answer



d The student used a statistical test to investigate the significance of the correlation between the number of bacterial colonies and exposure time to UV light.

A correlation coefficient of - 0.99 was found.

The following table shows the critical values, for p = 0.05, for this statistical test.

Number of means	Critical values at p = 0.05
4	1.00
5	0.90
6	0.83
7	0.71
8	0.64
9	0.60

What conclusion can be drawn from the results of this investigation?

Answer

The graph shows an inverse relationship (or negative correlation) between the time of exposure to UV light and the mean number of bacterial colonies. The calculated value of the correlation coefficient (0.99) is greater than the critical value at p = 0.05. The negative sign indicates an inverse relationship. We can therefore say that there is a significant inverse relationship between the time of exposure and the mean number of bacterial colonies and reject the null hypothesis.

Question 2

A group of nine athletes (A to I) wanted to see if training for two weeks at a mountain camp, 2000 m above sea level, had an effect on the number of red blood cells in their blood.

Samples of blood were taken from each of the athletes at their normal training camp at sea level. Blood samples were taken again after two weeks of training at the mountain camp.

The following table shows the results.

Athlete	Number of red blood cells × 10^{12} per dm^3 of blood	
	Before mountain training	**After mountain training**
A	5.0	4.9
B	5.1	5.3
C	4.9	5.7
D	5.3	5.5
E	5.4	5.6
F	5.0	5.4
G	4.8	5.3
H	5.1	5.6
I	5.5	5.1
mean =	**5.1**	**5.4**

a Complete the table by calculating the mean numbers of red blood cells × 10^{12} per dm^3 of blood for the athletes before and after mountain training.

b Calculate the percentage increase in the number of red blood cells after mountain training.

Answer

Percentage increase = $\left(\frac{0.3}{5.1}\right) \times 100 = 5.9\%$

contd.

c Write a null hypothesis for this investigation.

Answer

The difference between the mean number of red blood cells before and after mountain training is not significant.

d A Student's t-test was applied to the data to test the null hypothesis.

i State the number of degrees of freedom for these results.

Answer

The number of degrees of freedom = (9 + 9) – 2 = 16.

ii The calculated value of t was found to be 2.24.

The following table gives critical values of t, for the appropriate number of degrees of freedom.

Significance level (*p*)	0.20	0.10	0.05	0.01	0.001
Critical value of t	1.34	1.75	2.12	2.92	4.02

What conclusion can be drawn from the results this investigation?

Answer

The calculated value of t (2.24) is greater than the critical value of t at p = 0.05 (2.12). We can therefore say that there is a significant difference between the mean numbers of red blood cells before and after mountain training and reject the null hypothesis.