

# Maths tutorial booklet for

## M1: Handling data

Name: \_\_\_\_\_

Target grade: \_\_\_\_\_

### Quiz scores:

M1.1 Use an appropriate number of significant figures =

M1.2 Find arithmetic means =

M1.3 Construct and interpret frequency tables and diagrams, bar charts and histograms =

M1.4 Understand simple probability =

M1.5 Understand the principles of sampling as applied to scientific data =

M1.6 Understand the terms mean, mode and median =

M1.7 Use a scatter diagram to identify a correlation between two variables =

M1.8 Make order of magnitude calculations =

M1.9 Select and use a statistical test =

M1.10 Understanding measures of dispersion including standard deviation and range =

M1.11 Identify uncertainties in measurements and use simple techniques to determine uncertainty when data are combined =

## M1.1 – Use an appropriate number of significant figures

### Tutorials

Learners may be tested on their ability to:

- report calculations to an appropriate number of significant figures given raw data quoted to varying numbers of significant figures
- understand that calculated results can only be reported to the limits of the least accurate measurement.

### Significant figures

When you report your answers it is important to use the correct number of significant figures. The number of significant figures derives from the resolution of the measuring apparatus, which places an upper limit on the level of accuracy achievable with that apparatus (a piece of apparatus with very good resolution could of course be highly inaccurate if it is poorly calibrated). Calculated results can only be reported to the limits of the lowest resolution measurement (the lowest level of accuracy).

#### Lowest level of accuracy

For example, if there are 3 inputs in a particular calculation with 2, 3 and 4 significant figures respectively then the answer can only meaningfully be reported as correct to 2 significant figures.

Reporting answers to a certain number of significant figures may involve rounding. Remember - when the next number is 4 or below you round down, when the number is 5 or above you round up.

#### Rounding

1 - 4 → Round down

5+ → Round up

You need to be careful when you are rounding numbers. Common mistakes include forgetting to include zeros as significant figures. For example, 4.99 rounded to 2 significant figures is 5.0 not just 5.

Round to 2 significant figures:

4.99 → 5.0 ☺

4.99 → 5 X

Another “zero” mistake is to report zeros at the start of the number as significant figures. Remember, any zeros that come at the front of the number are not significant figures. For example, reporting 0.0256 to two significant figures will give you 0.026.

Report to two significant figures:

0.0256 → 0.026

You must also remember to always round numbers in a single step, not sequentially. For example for an answer with two significant figures you don't round 2.4478 first to 2.45 and then to 2.5. You need to round the number in a single step giving 2.4 to two significant figures.

Round to 2 significant figures:

2.4478 → 2.45 → 2.5      X

2.4478 → 2.4      ☺

Significant figures are important in all sorts of biological contexts, when reporting experimental data and for any calculation. Always remember to check whether the question asks you to report your answer to a certain number of significant figures, and if not, remember to use the lowest level of accuracy in the data.

## M1.1 – Use an appropriate number of significant figures

### Quiz

1. In each case convert to the number of significant figures quoted.

a) 2342 to 3 sig fig

b) 2342 to 2 sig fig

c) 456 to 2 sig fig

d) 0.07842 to 3 sig fig

e) 0.07842 to 2 sig fig

f) 0.003004 to 3 sig fig

(Note: for questions 2 to 4 you should be able to identify the appropriate number of significant figures to which to give your answer as well as convert the calculated result to that number of sig figs. If you are finding the calculations themselves difficult please refer to M2.3 and M2.4).

2. A hypothermic patient was rewarmed from 30.6°C to 37.1°C over the course of 3.4 h. What was the rate of warming (use °C h<sup>-1</sup> as your units)?

3. A willow coppice woodland in the UK has an area of 1.15 ha. (ha is the symbol for hectare – an area of land equal to 10,000 m<sup>2</sup>). When the willow harvest is taken each year, and dried, it yields 9 odt (oven-dry tonnes) of biomass. What is the productivity of the land (the amount of biomass produced per unit area) in units of odt ha<sup>-1</sup>?

4. A model cell is made of visking tubing (partially permeable membrane) containing sucrose solution and is immersed in distilled water. In 23.5 min the volume of the model cell increases by 1.0 cm<sup>3</sup> due to inflow of water by osmosis. What is the rate of osmosis in units of cm<sup>3</sup> min<sup>-1</sup>?

## **M1.2 – Find arithmetic means**

### ***Tutorials***

Learners 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.

### ***Calculating means***

The “mean” is another way of saying the “average”. The mean is calculated by adding together all the values and dividing by the number of values.

The symbol for mean is:

$$\bar{x}$$

As a formula this is written as:

mean equals the sum of all the data values ( $x$ ), divided by the number of data values ( $n$ ):

$$\bar{x} = \frac{\sum x}{n}$$

The calculated value for the mean can be quoted to the same number of decimal places as the raw data or to one more decimal place (see section M0.2). Comparing means is very useful for experimental repeats or comparing between different experimental conditions.

### ***Anomalies***

When working with data you need to be able to identify anomalies, or outliers. Anomalies are values which are judged not to be part of the inherent variation. Notice that there is an element of judgement here on the part of the scientist carrying out the investigation and this judgement must be exercised extremely carefully.

You must never discard a value just because it doesn't fit with what you expect or the rest of the data, especially without knowing more about the experiment itself. In general, if you do not have confidence in the data the best thing to do is to collect more data.

However, anomalies may occur because of a failure of the experimental procedure or because of human error, and in these situations the anomalies should be ignored, particularly if you are using the data to calculate the mean.

For example here are the bubble counts recorded in ten repeats of an experiment in which potato discs are added to hydrogen peroxide:

Repeat	1	2	3	4	5	6	7	8	9	10
Bubbles produced in 5 min	55	46	62	49	6	53	51	59	49	54

Repeat 5 has a much lower bubble count than any other repeat.

If there is a reason to think that this repeat might not have been carried out correctly, for example the discs used had been cut and then left out during a pause in the experiment (whereas in all other repeats they were used immediately) and might have become dry, then the experimenter might be justified in excluding this result from the calculation of the mean. If this choice is made it should be clearly recorded:

Repeat	1	2	3	4	5	6	7	8	9	10	Mean using all data	Mean excluding repeat 5
Bubbles produced in 5 min	55	46	62	49	6	53	51	59	49	54	48.4	53.1

## M1.2 – Find arithmetic means

### ***Quiz - calculate the mean:***

- Students measured the number of carrots eaten by rabbits over 24 hours. Calculate the mean number of carrots eaten.

Carrots eaten per rabbit:

6    5    8    5    9    6    7    7    7    8

2. The number of stomata on the upper and lower sides of 5 leaves of a plant were counted.

No. of stomata on lower side of leaf	No. of stomata on upper side of leaf
45	6
48	9
47	11
50	7
46	7

How do the mean numbers of stomata compare on the upper and lower sides of the leaf?

## **M1.3 – Construct and interpret frequency tables and diagrams, bar charts and histograms**

### **Tutorials**

Learners 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.

### **Construct and interpret frequency tables and diagrams, bar charts and histograms**

You should have come across frequency tables, bar charts and histograms in GCSEs. There are a few important rules that you need to follow whenever you represent data in a table or plot a graph. For tables you need to include clear headings and units and also be consistent with decimal places. For graphs including histograms and bar charts you must always:

- include a title
- include axes labels (with units)
- plot the independent variable (IV) on the x axis
- plot the dependent variable (DV) on the y axis
- ensure that the scales for both axes are linear (0, 1, 2, 3, 4...) unless producing a logarithmic plot (see M0.5 and M2.5)
- plot your data carefully
- make sure the graph is large enough to be easily readable (aim to use >50% of the space available).



You also need to remember the important differences between histograms and bar charts. Bar charts are used when the data is qualitative (either as categories such as different plant species or in a rankable form such as ACFOR abundance scores) or is quantitative but *discrete* - it can only take specific values and there are no 'in between' values. Examples of data suitable to be plotted on a bar chart include variables such as eye colour (qualitative categoric) or number of offspring (quantitative discrete). The bars in bar charts are the same width and there are gaps between the bars, they are not touching, when it is plotted.

Histograms on the other hand are used when the data is quantitative and *continuous*. Height, weight, and age are all examples of continuous data where they are measured to a specified accuracy. Unlike bar charts there are *no* gaps between the bars.

Note: the class width of different classes in a histogram can be different as it is the *area* of the bars that is proportional to frequency. However, this can be confusing for people interpreting the histogram so it is good practice, whenever possible, to use classes of the same width.

Bar chart	Histogram
Qualitative data (categoric or rankable) Discrete quantitative data	Continuous quantitative data
Bars the same width	Differing widths of bars possible but not advised
Bars not touching	Bars touch

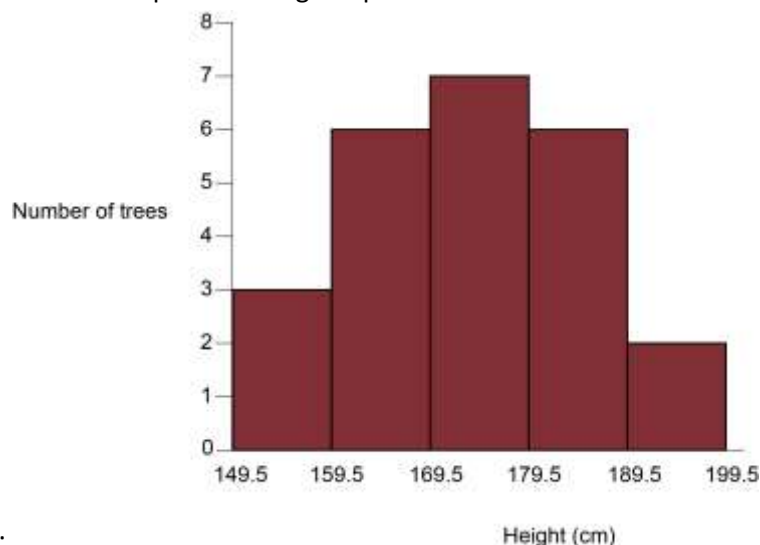
Let's try an example of plotting a histogram. The heights of different plants of the same species were measured to the nearest cm:

<b>Plant</b>	1	2	3	4	5	6	7	8	9	10	11	12
<b>Height (cm)</b>	157	188	163	188	171	184	187	161	175	169	193	175
<b>Plant</b>	13	14	15	16	17	18	19	20	21	22	23	24
<b>Height (cm)</b>	181	199	166	163	177	150	171	156	172	172	183	166

The first step in producing a histogram is to choose the number and width of classes and then create a frequency table using those classes:

<b>Height to the nearest cm</b>	<b>Frequency</b>
$150 \leq x < 160$	3
$160 \leq x < 170$	6
$170 \leq x < 180$	7
$180 \leq x < 190$	6
$190 \leq x < 200$	2

The histogram can now be plotted using the processed data and with each class represented by a bar on



the histogram:

As well as understanding when to use histograms and bar charts, and how to plot them, you must also be able to plot and interpret line graphs and scattergrams. More information about graphs can be found in M3.

## M1.3 – Construct and interpret frequency tables and diagrams, bar charts and histograms

### Quiz

For the below data sets:

- Determine whether a histogram or bar chart is the more appropriate graph to plot with reasons.
- Plot the graph.

- Blood samples were taken from a group of patients and the frequency of blood groups is presented in the table below.

Blood group	Frequency
A	40
B	10
AB	5
O	40

2. The ages of teenage boys and men attending at least one hour of gym class in a week were recorded. Process and present these data to show how the numbers doing this kind of exercise vary with age.

Age (years)	Age (years)	Age (years)	Age (years)	Age (years)	Age (years)	Age (years)	Age (years)	Age (years)	Age (years)
15.7	56.1	50.1	34.1	16.4	44.2	65.5	45.0	57.4	22.2
31.7	35.4	17.8	19.2	32.2	62.9	77.0	28.1	33.4	18.8
23.6	25.6	27.7	48.7	39.9	30.9	34.4	77.8	53.7	52.2
27.0	17.2	43.5	21.1	54.2	31.1	24.4	18.1	34.0	21.5
16.3	25.0	20.6	19.9	22.7	64.0	29.9	24.2	32.4	17.7
36.4	22.0	21.0	50.4	18.6	19.6	49.1	38.6	49.9	46.1
48.8	31.1	39.8	57.3	30.1	33.1	23.5	36.1	41.1	43.7

3. Vitamin C content of fruits

Fruit	Vitamin C content (mg 100g <sup>-1</sup> )
Apple	6
Banana	9
Lemon	46
Kiwi fruit	96
Orange	53
Strawberry	57

## M1.4 – Understand simple probability

### Tutorials

Learners may be tested on their ability to:

- Use the terms *probability* and *chance* appropriately
- Understand the probability associated with genetic inheritance.

### Probability and chance

#### Probability

When I toss a coin there are two possible outcomes, it can land facing heads up, or tails up. There are only two possible outcomes, but with each toss of the coin it is impossible to predict which of these two things will occur. Therefore, we can say that the outcome from the toss of a coin is random. Although the outcome of each individual coin toss is random, we know that (if the coin is unbiased) heads and tails are equally likely therefore we can say that there is a 50% probability of getting heads and a 50% probability of getting tails. This becomes extremely useful when the coin is tossed repeatedly. We can reason 'forwards and backwards' using the concept of probability to make meaningful statements about the results:

In the 'forwards' direction we can make useful predictions about the overall outcome of many repeats, even though each repeat is itself random. So, for example if we know that the coin will be tossed 100 times we can predict that there will be approximately 50 heads and 50 tails.

In the 'backwards' direction we can use the outcome of many repeats to assess whether our understanding of the system is correct. If the actual outcome of 100 tosses of the coin is 48 heads and 52 tails we would not be surprised – this is approximately the ratio of heads to tails we expected based on our assumption that the coin is unbiased. If, however we get 99 heads and only 1 tail we would be surprised and might suspect that the coin is biased. To put

exact numbers on how surprised we are, or more precisely how confident we are that our assumption about the system was wrong, we use statistical tests (see M1.9).

The probability of an event occurring is the likelihood of it occurring. The probability of an event A, written  $P(A)$ , can be between zero and one, with  $P(A) = 1$  indicating that the event will certainly happen and with  $P(A) = 0$  indicating that event A will certainly not happen. When I estimate the probability of a coin toss landing on heads, it would be 0.5. If there is more than one variable involved, then you calculate the probability that an individual has one variable outcome AND the other variable outcome by multiplying the separate probabilities together. For instance if I wanted to calculate the probability of getting heads on TWO coin tosses that would be  $0.5 \times 0.5 = 0.25$ .

### The difference between probability and chance

In statistics we use **probability** as the appropriate term for expressing likelihoods (as ratios, decimals or percentages)

For example – When tossing a coin we have a **probability of 0.5** of the coin landing on heads – with this information we can predict that with ten coin flips, approximately five of them will land heads up.

In everyday speech '**chance**' is often used to mean the same thing as '**probability**'. For example we might say 'I think the **chance** of rain today is about 50/50' meaning that we think there is a 50% **probability** of rain. However, when we use the word 'chance' in commenting on statistics we are using it to talk about the random deviations from probability that can occur (and especially whether we think these random deviations are sufficient to explain why the outcome did not exactly match our expectation based on probability).

Using coin tosses as an example once more, if we actually got six heads from ten tosses of the coin we might say 'getting six rather than the predicted five was due to **chance**'. If we got nine heads we might say 'this is too far from the predicted outcome to be explained by **chance** and so we now believe this coin is biased'. The cut-off point for how far away from the prediction (based on probability) the results have to be before they cannot be explained by chance is where statistical tests come in (see M1.9).

The effect of chance, in percentage terms, is largest when the number of repeats is small. Clearly if we only toss a coin four times, a single result has a big effect on the percentage of heads whereas if we do 1000 repeats the effect of a single result is much smaller. This is the reason that experiments often use many repeats, or take many samples.

### Probability and patterns of genetic inheritance

In a genetic cross of two *Drosophila melanogaster* (fruit flies) that are each heterozygous for the mutant allele of the *vestigial* gene (which is recessive and found on chromosome 2), we predict that the probability of each individual offspring being a recessive homozygote (which means it will have tiny wings) is 0.25, the equivalent of saying that 1 in 4 offspring will have tiny wings.

Parents	Vg / vg x Vg / vg			
Allele inheritance patterns for offspring (four equally likely outcomes):	Vg / Vg	Vg / vg	vg / Vg	vg / vg
Offspring Genotypes (three outcomes):	Vg / Vg	Vg / vg		vg / vg
Offspring Phenotypes (two outcomes):	Normal wings			Tiny wings

This can be displayed as a Punnett square showing the four equally likely offspring:

		<b>Vg / vg female parent gametes</b>	
		<b>Vg</b>	<b>vg</b>
<b>Vg / vg male parent gametes</b>	<b>Vg</b>	<b>Vg / Vg</b> <b>Normal wings</b>	<b>Vg / vg</b> <b>Normal wings</b>
	<b>vg</b>	<b>Vg / vg</b> <b>Normal wings</b>	<b>vg / vg</b> <b>tiny wings</b>

Things get a little more complicated when we include a second genetic trait. Individuals that are homozygous for the recessive mutant version of the *ebony* gene (found on chromosome 3), have very darkly coloured bodies (Heterozygotes sometimes have slightly darker bodies). If the parents are heterozygous at both the *ebony* locus and the *vestigial* locus, then the probability of an individual having tiny wings and dark bodies is  $0.25 \times 0.25 = 0.0625$  or  $1/16$ .

Again this can be shown in a Punnett square giving the 16 equally likely outcomes:

		Vg / vg Eb / eb female parent gametes			
		Vg Eb	vg Eb	Vg eb	vg eb
Vg/vg Eb/eb male parent gametes	Vg Eb	Vg/Vg Eb/Eb Normal wing Normal body	Vg/vg Eb/Eb Normal wing Normal body	Vg/Vg Eb/eb Normal wing Normal body	Vg/vg Eb/eb Normal wing Normal body
	vg Eb	vg/Vg Eb/Eb Normal wing Normal body	vg/vg Eb/Eb tiny wing Normal body	vg/Vg Eb/eb Normal wing Normal body	vg/vg Eb/eb tiny wing Normal body
	Vg eb	Vg/Vg eb/Eb Normal wing Normal body	Vg/vg eb/Eb Normal wing Normal body	Vg/Vg eb/eb Normal wing ebony body	Vg/vg eb/eb Normal wing ebony body
	vg eb	vg/Vg eb/Eb Normal wing Normal body	vg/vg eb/Eb tiny wing Normal body	vg/Vg eb/eb Normal wing ebony body	vg/vg eb/eb tiny wing ebony body



And the numbers of each of those outcomes giving each of the four possible phenotypes is easier to see if we colour code:

		Vg / vg Eb / eb female parent gametes			
		Vg Eb	vg Eb	Vg eb	vg eb
Vg/vg Eb/eb male parent gametes	Vg Eb	Vg/Vg Eb/Eb Normal wing Normal body	Vg/vg Eb/Eb Normal wing Normal body	Vg/Vg Eb/eb Normal wing Normal body	Vg/vg Eb/eb Normal wing Normal body
	vg Eb	vg/Vg Eb/Eb Normal wing Normal body	vg/vg Eb/Eb tiny wing Normal body	vg/Vg Eb/eb Normal wing Normal body	vg/vg Eb/eb tiny wing Normal body
	Vg eb	Vg/Vg eb/Eb Normal wing Normal body	Vg/vg eb/Eb Normal wing Normal body	Vg/Vg eb/eb Normal wing ebony body	Vg/vg eb/eb Normal wing ebony body
	vg eb	vg/Vg eb/Eb Normal wing Normal body	vg/vg eb/Eb tiny wing Normal body	vg/Vg eb/eb Normal wing ebony body	vg/vg eb/eb tiny wing ebony body

So from our probability calculation, illustrated with the Punnett square we can say, for each offspring there is a 1/16 or 0.0625 or 6.25% probability that it will be doubly homozygous for the recessive alleles and therefore display the double mutant phenotype of tiny wings and an ebony-coloured body.

Similarly there is a 3/16 probability of each individual having normal wings but an ebony-coloured body. And 3/16 probability of each individual having tiny wings but a normal-coloured body. There is a 9/16 probability of each offspring appearing normal (although of course many of these flies will be heterozygotes, carrying mutant alleles at one or both loci).

This leads to the classic expected ratio of phenotypes in the offspring in this kind of breeding experiment:

$$9 : 3 : 3 : 1$$

What if the results you get from such an experiment don't exactly match the ratio you'd expect based on your probability calculations?

It could just be due to chance or it could indicate that your assumptions were wrong. Statistical tests (in this case the chi-squared test) (see M1.9) will help you decide if the deviation away from your expectation is big enough to demand a re-think of your assumptions. If so, maybe the loci are linked (on the same chromosome) so they won't be inherited truly independently. Or maybe something else weird is happening (e.g. the double mutant offspring are not surviving to adulthood to be counted).

## M1.4 – Understand simple probability

### Quiz

1. What is the probability of rolling a 5 on a six-sided die?

2. What is the probability of rolling a 3 or a 5 on a six-sided die?

3. What is the probability of rolling at least one 3 when rolling two dice?

4. What is the probability of rolling two 3s one after the other when rolling a single die?

5. We have two cats where one is homozygous for alleles that produce a short tail which is a recessive trait, while the other is homozygous for alleles that produce a long tail which is dominant. With this knowledge we can make predictions about the characteristics of any offspring produced when these two cats are bred together.

- A. What is the probability that offspring would inherit one copy of the short tail allele?

- B. What is the probability that offspring would have short tails?

6. I have two flowers, both heterozygous at two loci. One locus is the petal colour gene where the yellow allele is recessive and the red allele is dominant. The other locus is the petal shape gene where a smooth petal allele is dominant and a wrinkly petal allele is recessive.

A. What is the probability of an individual inheriting just one copy of the wrinkly allele and just one copy of the red allele? (i.e. being doubly heterozygous just like the parents)

B. What is the probability of seeing an individual from this cross which has wrinkly red petals?

## M1.5 – Understand the principles of sampling as applied to scientific data

### Tutorials

Learners may be tested on their ability to:

- analyse random data collected by an appropriate means, e.g. use Simpson's index of diversity to calculate the biodiversity of a habitat.

### Sampling

Sampling is an important part of scientific work: collecting data so that we can make inferences about relationships and processes in the natural world. For example, the only way to know the true mean height of a population of *Sequoia* (redwood) trees would be to measure every individual in the population. In reality this is practically impossible, because of the time it would take to do this, so instead we collect enough measurements in a **sample** to give us an accurate **estimate** of the **true mean** for the height of *Sequoia* trees, and we would refer to this as our sample mean.

There are broadly two ways to take samples: **randomly** or **non-randomly**.

**Random sampling** - When we carry out random sampling, the positions or sampling strategy used should be generated randomly. You as the experimenter do not choose which sites to sample, these are generated by a random number generator (such as the one on your calculator). In this way you do not influence the way in which measurements are taken.

**Non-random sampling** - A systematic, or even, approach to sampling. Measuring bacterial growth every 10 minutes, or taking a sample every 5 m across a woodland. This can be useful for measuring how things change over time or distances, where random sampling could cause strange clusters in sampling.

Stratified sampling – 'Stratified' or 'layered' sampling is a slightly more complex approach to sampling. The sampling approach is split according to predefined layers in the system being studied that could affect the results, for example splitting measurements of height by gender, and making sure that each 'layer' has a representative number of samples, in this case 50:50 to represent equal numbers of men and women.

In another example we might look at the distribution of invertebrates in a field, 10% of the field is water-logged and so I make sure that 10% of the samples are collected from water-logged areas.

When choosing which approach to use:

**Random sampling** is a good default when we have no reason to space things out evenly. An even distribution is not the same as a random distribution!

**Non-random sampling** is best applied when we have a good reason for wanting to know how things might change at regular intervals (time or distance are the most common uses for this.)

When employing **stratified sampling** we should have a good reason for why it is reasonable to split the data collection according to the proposed 'layers'. Within each 'layer' of our stratified sampling we may choose to sample randomly or non-randomly, whichever is most appropriate for the question being asked.

## **Simpson's Index**

Sampling in field ecology is often used to measure abundance of different organisms in a habitat.

Once the abundance of all the organisms present in a habitat has been determined, scientists will often mathematically calculate the biodiversity present in a habitat. This can be done using Simpson's Index of Diversity (D):

D = diversity index

N = total number of organisms in the ecosystem

n = number of individuals of each species

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

Simpson's Index of Diversity always results in a value between 0 and 1, where 0 represents no diversity and a value of 1 represents infinite diversity. The higher the value of Simpson's Index of Diversity, the more diverse the habitat.

## **M1.5 – Understand the principles of sampling as applied to scientific data**

### **Quiz**

1. I want to measure the change in distribution of green alga from the low tide mark to the high tide mark. Should I use a random or non-random sampling method for choosing where to place my quadrats?

2. You want to measure the distribution of flowers in a woodland. The woodland has been divided up into 100 areas of 10 m<sup>2</sup>. You cannot measure them all and so have to choose 10 sampling points. Should you use random or non-random sampling?

3. If in the previous example 19 of the areas were identified as heavily waterlogged how might stratified sampling be employed to improve our sampling technique?

4. A rock pool was sampled for species richness.  
Calculate Simpsons Index of Diversity for this habitat using the formula:

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

Species	Numbers	
Common periwinkle	35	
Dog whelk	41	
Common limpet	8	
Sea urchin	4	
Top shells	24	
Total (N)		

## M1.6 – Understand the terms mean, mode and median

### Tutorials

Learners 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.

### Mean, median and mode

The mean, median and mode are all measures of central tendency and act as representative values for the whole data set. The mean, usually denoted as  $\bar{x}$ , has already been mentioned in section M1.2 and is the sum of the data values divided by the number of data values.

Mean

$$\bar{x} = \frac{\sum x}{n}$$

The median is the middle value of a data set and is usually denoted by  $Q_2$ . Formally the middle value is the  $(n+1)/2$ th piece of data where  $n$  is the number of pieces of data.

Median

$$Q_2 = \frac{n+1}{2}\text{th}$$

Finally, the mode is the easiest to spot as it is the most frequently occurring value.

Mode

Most frequent

For example, take this set of data values:

25,24,27,28,19,31,25,31

To calculate the mean you need to add together all the values and divide by the number of values.  
This gives us 210 divided by 8, which equals 26.25

$$\bar{x} = \frac{210}{8} = 26.25$$

To find the median for the same data set we have to reorder the data from smallest to largest, meaning our data now looks like this:

**19,24,25,25,27,28,31,31**

Here there are 8 pieces of data, so using the  $(n+1)/2$ th formula we know that the median lies on the  $(8+1)/2$ th piece of data which is the 4.5th piece of data

$$(8+1)/2 = 4.5$$

In this example we have an even number of data values, therefore the 4.5th piece of data is halfway between the 4th and 5th values; 25 and 27. To work out the value halfway between these values we add them together and then divide by two.

$$25+27 = 52$$

$$52/2 = 26$$

Therefore, the median for this data set equals 26.

Finally, in this set of data there are two values which occur twice – 25 and 31.

**25,24,27,28,19,31,25,31**

**Mode = 25 and 31**

Therefore this data set is “bi-modal”, it has two different modes, 25 and 31.

Generally the mean is the most useful statistical measure. However, if there are outliers in the data then the median is more representative. For example, if a value of 100 was added to the example data set above, the mean would change to 34.4 but the median would move to 27, a far more representative measure.

**25, 24, 27, 28, 19, 31, 25, 31, 100**

$$\bar{x} = 34.4$$

$$Q_2 = 27$$

You may be asked to calculate and compare these quantities for any biological context where measurements are taken. For example when measuring: the length of leaves, the mass of model cells, or the number of a group of organisms and many more.



## M1.6 – Understand the terms mean, mode and median

### Quiz

Plants were grown in both the sun and the shade and height measurements taken. Calculate the mean, mode and median for each set of data.

Height in sun (cm)	Height in shade (cm)
244	104
265	83
312	131
199	99
278	118
345	150
236	162
197	118
266	146
237	128

	Height in sun (cm)	Height in shade (cm)
Mean		
Mode		
Median		

Numbers of mucus-secreting goblet cells were counted per colonic intestinal crypt in patients with Crohn's disease and healthy patients. Calculate the mean, mode and median for each set of data.

Number of goblet cells – Crohn's disease patients	Number of goblet cells – Healthy patients
9	15
11	12
7	14
15	9
10	11
8	13
7	12
12	10
13	16
7	11

	Crohn's disease patients	Healthy patients
Mean		
Mode		
Median		

## M1.7 – Use a scatter diagram to identify a correlation between two variables

### Tutorials

Learners may be tested on their ability to:

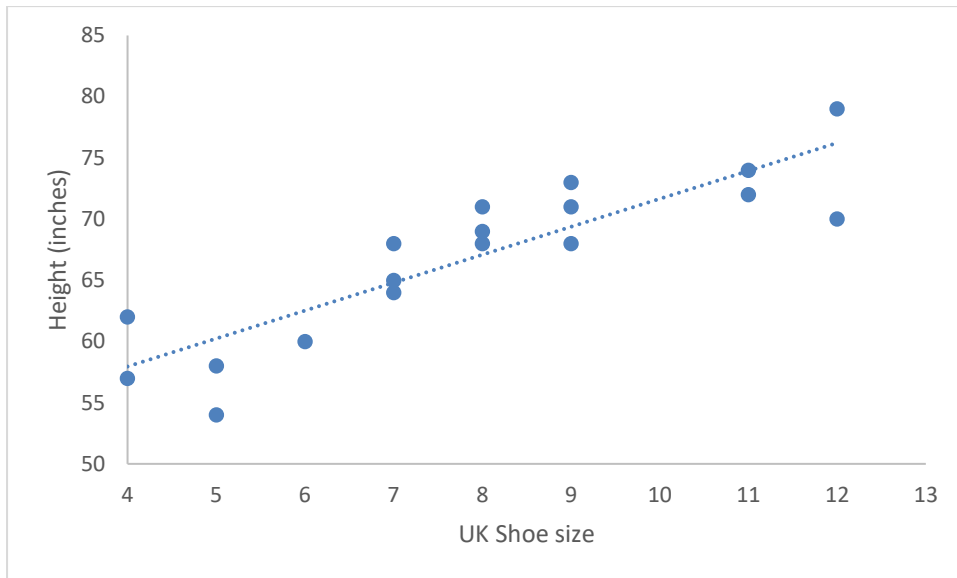
- Interpret a scatter diagram, e.g. the effect of lifestyle factors on health.

### Scatter diagrams

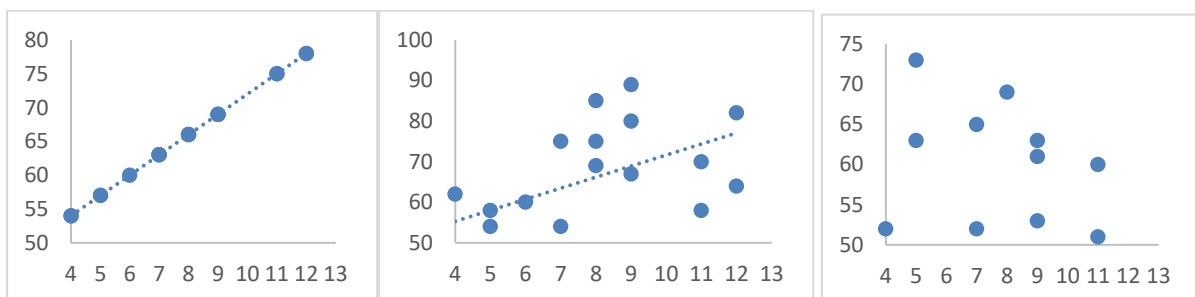
Correlation is a very useful statistical technique that looks at relationships between two variables. For example, taller people tend to have larger shoe sizes than shorter people, therefore we can say that height is correlated with shoe size.

This relationship isn't perfect, two people of the same height can have different shoe sizes, and a shorter person could have a larger shoe size than someone who is taller than them. The **correlation coefficient** tells us how close that relationship is to being perfect (see M1.9 for details on how to calculate and interpret this), but we can plot a range of people's height measurements and shoe sizes on graphs called 'scatter diagrams' or 'scatterplots' to easily visualise this relationship.

	UK Shoe size	Height (inches)
1	4	57
2	4	62
3	6	60
4	7	68
5	8	71
6	9	73
7	11	72
8	11	74
9	12	70
10	12	79
11	5	58
12	5	54
13	8	68
14	8	69
15	9	71
16	9	68
17	7	65
18	7	64



In this example we can see that our graph shows a strong positive linear correlation. A strong correlation is one where we can draw a trend line through the data points, and they are all very close to the line. Although it is a strong correlation, you can see that few of the data points sit perfectly on the trend line, instead there are just as many above and below it. A data point that sits below the trendline describes an individual who is shorter than their shoe size would predict, while a data point above the line describes someone who is taller than their shoe size predicts. The closer the data points cluster to the line, the more accurately the trend line describes individuals in the dataset.



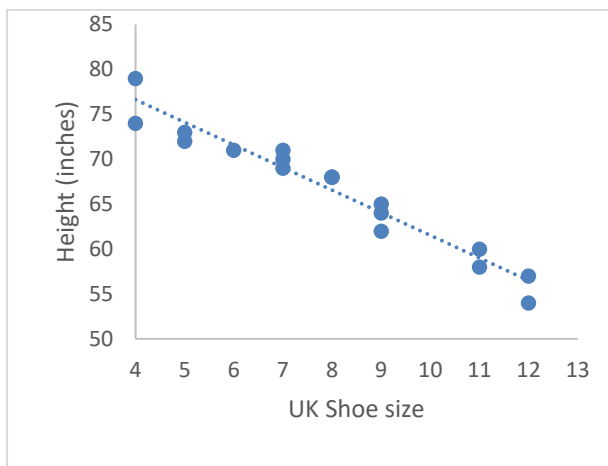
*Perfect positive correlation*

*Weak positive correlation*

*No correlation*

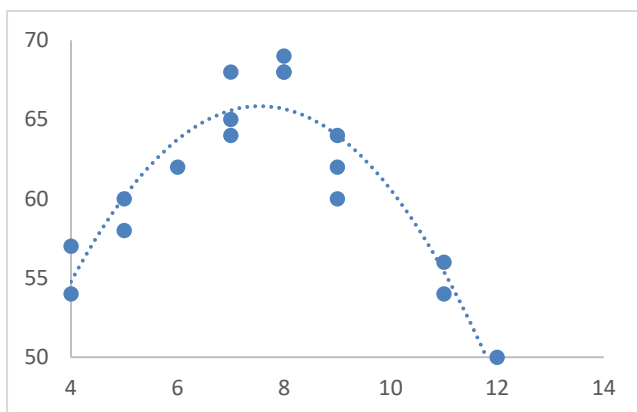
Our example shows positive correlation because as height increases so does shoe size.

Negative correlation would imply the opposite: as height increased, shoe size decreased



*An example of negative correlation*

Our example is linear because the relationship does not change (we do not reach a certain point where very tall people start to have smaller feet).



*Example of quadratic relationship*

It is important to remember that *correlation does NOT imply causation*. While it might seem intuitive to say that being taller **causes** feet to generally be bigger, we cannot conclude this simply from the evidence of correlation. It would be just as consistent with the observed correlation to suggest that the larger feet cause the greater height or that there is a third variable (such as genetics or diet) that affects both height and shoe size in the same way.

## M1.7 – Use a scatter diagram to identify a correlation between two variables

### Quiz

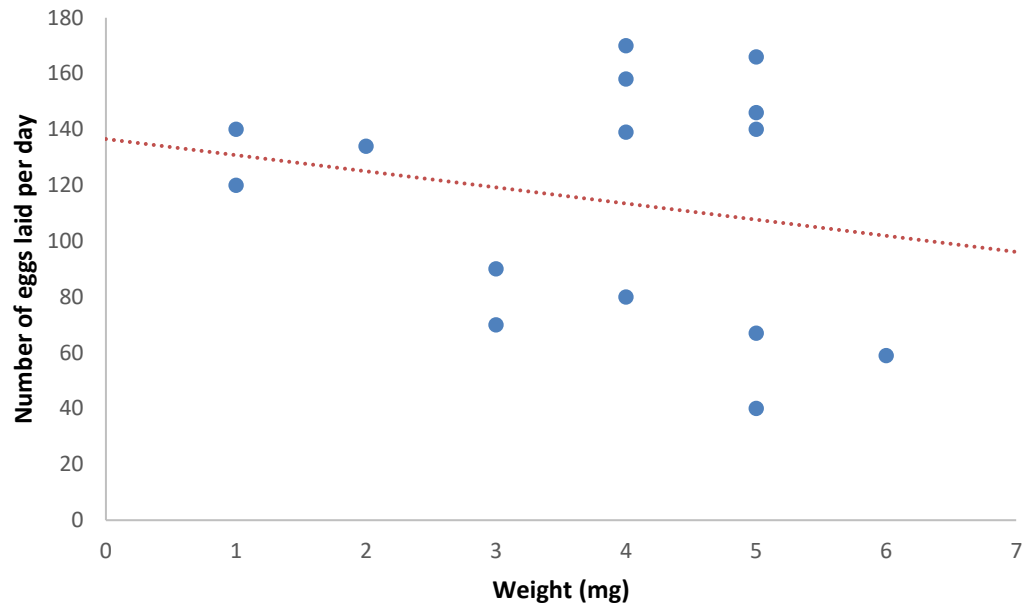
1. Which of the following is/are appropriate to draw as scatterplots?
  - A. The mean horn length of two populations of African rhinos
  - B. The frequency of short-haired and long-haired cats from a cross of two long-haired parents
  - C. The diameter of oak tree trunks and the average number of leaves per branch
  - D. The abundance of insects and the fledging weight of lapwing chicks.

2. Plot the following information from the table into a scatterplot – the length of a male peacock's tail against the number of females he courted in a single breeding season

Peacock	Tail length (cm)	Number of females courted
1	140	1
2	135	1
3	156	3
4	147	4
5	152	5
6	164	5
7	154	4
8	162	6
9	139	2
10	149	3
11	153	4
12	159	5
13	154	5
14	157	4
15	161	5

Add a trendline to this scatter plot and describe the relationship you observe.

3. Describe the relationship observed in this scatterplot charting the weight of female house flies against the number of eggs laid per day.



## M1.8 Make order of magnitude calculations

### Tutorials

Learners may be tested on their ability to:

- use and manipulate the magnification formula,

$$\text{Magnification} = \frac{\text{size of image}}{\text{size of real object}}$$

### Order of magnitude

Orders of magnitude are used to make approximate comparisons of size or quantity. If two numbers have the same order of magnitude, they are about the same size. If two numbers differ by one order of magnitude, one is about ten times larger than the other. If they differ by two orders of magnitude, they differ by a factor of about 100, and so on.

Powers of ten	Order of magnitude
0.0001	-4
0.001	-3
0.01	-2
0.1	-1
1	0
10	1
100	2
1,000	3
10,000	4

For example, if you were comparing the surface of an orange with that of the earth, one would say the surface of the earth is many orders of magnitude larger than that of the orange.

The formula used to calculate magnification is:

$$\text{Magnification} = \frac{\text{size of image}}{\text{size of real object}}$$



You can rearrange the formula to calculate any of the three unknowns, as long as you have the other two. However for this to work you must make sure that both quantities/sizes are in the same units.

For example take the following micrograph showing a section of mouse pancreas at high power, one acinus (a cluster of cells) is circled. The magnification of the micrograph is x400 and the length of the circled acinus is 32 mm. We want to calculate the actual size of the acinus.



Magnification → x400

Length of the circled acinus in the image → 32 mm

We want to calculate the actual size of the acinus. This means we need to rearrange the magnification formula (*rearranging formulas is covered in detail in section M2.2*) to make the size of real object the subject of the equation. This gives us the size of the real object equals the size of the image divided by the magnification.

$$\text{Magnification} = \frac{\text{size of image}}{\text{size of real object}}$$

$$\text{Size of real object} = \frac{\text{size of image}}{\text{magnification}}$$

In this example this means the actual size of the acinus is equal to 32 mm (the size of the acinus in the image) divided by 400 (the magnification). Therefore the actual size of the acinus is 0.080mm, or in more appropriate units, 80 μm.

$$32 \text{ mm} / 400 = 0.080 \text{ mm} = 80 \text{ μm}$$

## M1.8 Make order of magnitude calculations

### Quiz

- 1 This is an electron micrograph of a mitochondrion. Its actual length is  $5\text{ }\mu\text{m}$ . Calculate the magnification of the image.



**B0000119** Credit [Prof. R. Bellairs](#), Wellcome Images

TEM of a mitochondrion

A transmission electron micrograph of a mitochondrion in a chick embryo cell.

- 2 This botanical illustration from about 250 years ago shows a banana plant. The image has a scale line where each division represents 30 cm. What is the magnification?

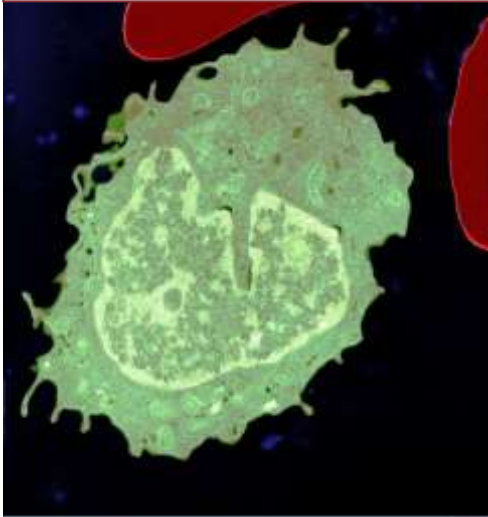


**V0043033 Credit:** Wellcome Library, London

Banana plant (*Musa* species): flowering and fruiting plant with stolons and separate floral segments and sectioned fruit, also a description of the plant's growth, anatomical labels and a scale bar.

3

What is



**B0004162 Credit** [University of Edinburgh](#), Wellcome Images

Monocyte and two red blood cells

Colour-enhanced image of a monocyte and two red blood cells. Monocytes are white blood cells that develop



## M1.9 – Select and use a statistical test

### Tutorials

Learners may be tested on their ability to select and use:

- the chi squared test ( $\chi^2$ ) to test the significance of the difference between observed and expected results
- the Student's *t*-test
- the Paired *t*-test
- the Spearman's rank correlation coefficient.

### Statistical tests

Choosing and using statistical tests can seem daunting at first, but they are very useful tools for analysing data. In simple terms each type of statistical test has one purpose: to determine the probability that your results could have occurred by chance as opposed to representing a real biological effect.

Why do we need statistical tests? As scientists we are interested in finding results that apply as general rules. For example, on average are students in Year 10 taller than students in Year 9? The best and most complete way to do this would be to find every single student across the whole country that is currently in Years 9 or 10 at school and measure every single one. In reality we cannot collect data from every school in the country, it would just take too long. Therefore in this example, and with all experiments, we collect data from a small subset of the population instead (this is our sample). From the sample data (e.g. all of Year 9 and 10 in one school) we infer things about the population as a whole.

Statistical tests allow us to make quantitative statements about the inferences we have made. We can put a number on how confident we are that our conclusion about the whole population is correct based on the sample we have taken.

We will cover four types of statistical test: **the chi squared test, the Spearman's rank correlation, the Student's *t*-test, the paired *t*-test.**

The choice of which statistical tests we use on our data depends on the question being asked. So always look at your data and ask yourself whether you can say yes to these questions – the one that fits best tells you which statistical test to perform.

- 1) Am I looking at frequencies, and whether my observations differ from expected values? For example – count the number of red, purple and white flowers that come from a genetic cross of two purple flowers where I expect a ratio of 1:2:1

Test – **chi squared test**

- 2) Am I looking at the relationship between two variables?

For example – ice-cream consumption and blood sugar levels, to see if people who eat a lot of ice-cream have higher blood sugar.

Test – **Spearman's rank correlation**

- 3) Am I looking at whether there is a difference in the means between two separate/independent groups?

For example – measuring the heights of men and women to see if there is a difference in the average height by gender

Test – **Student's t-test**

- 4) Am I looking at whether there is a difference in the mean between the same group before and after a change?

For example – measuring the cholesterol levels in people before and after switching to a vegetarian diet to see if there is an effect on cholesterol of this dietary change

Test – **Paired t-test**

Which of the four tests is most appropriate for answering the example question we had earlier: 'on average are students in Year 10 taller than students in Year 9'?

## Hypotheses

Statistical tests allow us to test hypotheses about relationships. With every statistical test we generate two competing propositions:

- the null hypothesis ( $H_0$ )
- the alternative ( $H_1$ )

The alternative hypothesis comes from your idea that a particular effect will be present, while the null is simply the opposite, that the effect is absent.

Taking our previous example of height and year group we can generate the following hypotheses:

- $H_1$  : Students in Year 10 are taller on average than students in Year 9
- $H_0$  : On average students in Year 10 do not differ in height from students in Year 9

The reason we have a null hypothesis is because we cannot prove experimental hypotheses but we can reject dis-proven hypotheses. This can be quite confusing but, simply put, it is easier to dis-prove a

theory than prove one. If our data gives us the confidence to reject the null hypothesis then this provides support for our alternative hypothesis, but it does not prove it.

Similarly if our statistical test shows no significant effect, we refer to this as failing to reject the null hypothesis. This is the statistics equivalent of using “not guilty” rather than “innocent” in a court verdict; we have not provided the evidence to reject the null hypothesis at this time but it doesn’t preclude changing our minds if more evidence comes to light at a later date.

## Student’s unpaired *t*-test

This is the best test for looking at average differences between independent groups. So this is the test we would use to compare, for example, the average height of children in Year 9 and the average height of children in Year 10. We would take a sample from each of the year groups (one Year 9 class and one Year 10 class) and measure the variable we’re investigating (height) for all the individuals in each sample. Then, on the basis of these measurements, we use the Student’s *t*-test to say whether we can be reasonably confident that there really is a difference in the mean height of all Year 9 children compared to all Year 10 children.

Year 9		Year 10	
Name	Height (cm)	Name	Height (cm)
Connor	170	Mick	178
Tristan	181	Keith	174
James	178	Charlie	173
Brad	170	Ronnie	175
Alana	174	Danielle	178
Este	180	Taylor	178
Tegan	157	Jenny	174
Sara	157	Gemma	163
<b>Mean</b>	<b>170.9</b>	<b>Mean</b>	<b>174.1</b>
<b>Standard deviation</b>	<b>9.5</b>	<b>Standard deviation</b>	<b>4.9</b>

You can see that the means of our two sample groups **are** different.

<b>Mean</b>	<b>170.9</b>	<b>Mean</b>	<b>174.1</b>
-------------	--------------	-------------	--------------

No-one can say there is **no** difference there.

But we are not interested in the samples. We are interested in using the data from our samples to say things (with confidence) about the **whole population** (in this case all of year 9 and all of year 10). The important thing for us to find out, therefore, is whether the difference we see between the sample means is **significant** – is it big enough (given the size of the sample and how much variation we see in the data) for us to be confident that it reflects a real difference between the two year groups rather than just chance variations in the samples we happen to have picked?

Our **null hypothesis** is that there is no difference between the heights of Year 9 and Year 10 children. If this is true the difference between the sample means is **not** because there is really any difference between the means for all Year 9s and all Year 10s. It just arose by chance in the particular samples we took. The difference we see in our samples is not big enough to make us confident in saying that the two year groups really are different. We would say that there is '**not a significant difference**'.

The **alternative hypothesis** is that there **is** a difference in height between the two whole year groups. In order to reject the null hypothesis we need to identify a 'significant difference' between the sample means. The difference is big enough that we can be confident it is telling us there is a real difference between the year groups. We can make a statement such as "on average the students in Year 10 are taller than students in Year 9".

The statistical test allows us to find out whether we can confidently reject the null hypothesis.

The Student's *t*-test formula is as follows:

$$t = \frac{|\bar{x}_A - \bar{x}_B|}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

The modulus sign (vertical lines) in the numerator tells us to ignore any minus sign once we have subtracted one mean from the other

Subscript <sub>A</sub> and <sub>B</sub> refer to the two groups – Year 9 and Year 10.

$\bar{x}$  refers to the mean, so that  $\bar{x}_A$  is the group mean of the Year 9 class

*S* is the standard deviation

*n* is the sample size

So that



$$t = \frac{(170.9 - 174.1)}{\sqrt{\frac{(9.5)^2}{8} + \frac{(4.9)^2}{8}}}$$

$$t = 0.4$$

To understand what this means we must look up the value in the Student's  $t$ -test significance tables.

First calculate the degrees of freedom (df) which is  $n-1$  for each group:

$$(8-1) + (8-1) = 14$$

At 14 df the value of  $t = 0.4$  is below the threshold of  $t = 2.15$  which is the threshold at which we consider the difference to be significant

**There is no significant difference between the height of students in Year 9 and Year 10 in our dataset.**

Therefore we state that we have failed to reject the null hypothesis

This does **not** mean that we have proved that the mean height of Year 9 and Year 10 children is the same. It means that we have failed to show a significant difference based on the data we have gathered. Perhaps there really is no difference, or perhaps there really is a difference but our samples failed to show it (which could be for many reasons but most obviously it could simply be that the samples were not large enough).

#### Assumptions

When performing a Student's  $t$ -test the following things are assumed about the data in order to trust the test result.

- We have two independent groups
- For each group we have taken an unbiased sample and measured the same variable
- The variable is continuous
- The continuous variable is normally distributed for each group
- Each group has approximately equal variances (i.e. similar standard deviations) for this variable
- The sample sizes are roughly equal

### **Paired $t$ -test**

When looking for differences between means in two groups we use a  $t$ -test. If the two groups are independent of each other we use the unpaired version of this test. However, if the two groups come as related pairs we can use the paired  $t$ -test, allowing us to identify quite subtle but significant differences

that might be missed with the unpaired test. It is essential to understand that the pairing must be done according to some **genuine relationship** between the members of each pair and must always be done based on that relationship **not based on the data gathered**.

For example if we measure a variable such as systolic blood pressure in a set of patients on Monday and then measure the same variable in the same set of patients on Tuesday we have two groups (patients on Monday and patients on Tuesday) and there is a natural pairing across these two groups (data on patient A on Monday will obviously be paired with data on the same patient the next day). This is a prime example where using the paired *t*-test is appropriate.

But beware! You might think that the following scenario would also allow analysis by the paired *t*-test **but it would not**:

- We measure systolic blood pressure in two groups of ten patients.
- We then rank the data in each group from highest to lowest.

Now can we pair up the highest in each group, then pair up second highest and so on? **No!** This is pairing after data gathering and is using the data itself to guide the pairing. Using a paired *t*-test in this case could easily lead us to mistakenly identify a significant difference where none exists.

As an example we will use the paired *t*-test to compare the mean difference in shell size of the same hermit crabs, before and after they are given the opportunity to swap out their shells for one of a range of others. These are measurements on 15 individual crabs measured twice (before and after shell swapping) .

Hermit crab	Shell size before shell swapping (mm)	Shell size after shell swapping (mm)	Difference $d$	How far is difference from mean difference? $(d - \bar{d})$	$(d - \bar{d})^2$
1	8	8	0	-2.2	4.84
2	10	12	2	-0.2	0.04
3	5	7	2	-0.2	0.04
4	6	8	2	-0.2	0.04
5	4	5	1	-1.2	1.44
6	7	11	4	1.8	3.24

Hermit crab	Shell size before shell swapping (mm)	Shell size after shell swapping (mm)	Difference d	How far is difference from mean difference? $(d - \bar{d})$	$(d - \bar{d})^2$
7	9	10	1	-1.2	1.44
8	11	13	2	-0.2	0.04
9	13	13	0	-2.2	4.84
10	15	14	-1	-3.2	10.24
11	8	11	3	0.8	0.64
12	9	14	5	2.8	7.84
13	7	15	8	5.8	33.64
14	9	10	1	-1.2	1.44
15	11	14	3	0.8	0.64
		<b>Total</b>	<b>33</b>		<b>70.4</b>

$\bar{d}$  = Mean difference =  $33/15 = 2.2$

Calculate for each hermit crab how far the difference in shell sizes is from the mean difference. Square these numbers and add them up to find the total = **70.4**

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

From this we can calculate the standard deviation of the difference

$$= \sqrt{\frac{70.4}{14}} = 2.24$$

The paired t-test is as follows:

$$t = \frac{\bar{d} \sqrt{n}}{s_d} = \frac{2.2 \sqrt{15}}{2.24} = 3.8$$

Now we need to see whether this value of  $t$  is large enough for us to reject our null hypothesis. We can refer to a critical values table, picking the entry for our desired confidence level (95% or  $p=0.05$ ) and the correct degrees of freedom. In a paired  $t$ -test the number of degrees of freedom is  $n-1$ .

15 individuals were used so  $n-1 = 14$

The critical value at  $p = 0.05$  for 14 degrees of freedom is 2.15

$3.8 > 2.15$  so our  $t$  value is greater than the critical value and we can **reject the null hypothesis** that there is no change in average shell size after being given the opportunity to swap shells. Giving hermit crabs the option to change their shells does have an effect on average shell size.

### Assumptions

When performing a paired  $t$ -test the following things are assumed about the data in order to trust the test result.

- We have two groups with some dependency or relationship between specific pairs (one from one group one from the other) (e.g. same subjects measured before and after)
- We have taken an unbiased sample of these pairs and measured the same variable
- The variable is continuous
- The continuous variable is normally distributed for each group with the same variance

## **Spearman's rank correlation coefficient**

If we have data on two variables for a set of items and we want to see if these variables are related we can test them for correlation. Correlation comes in two forms:

Positive correlation – as one variable increases in value, so does the other

Negative correlation – as one variable increases in value, the other decreases in value

As an example we will use the Spearman's rank correlation coefficient to comment on the relationship between the size of a locust and the length of its wings. So in this example the set of items is the locusts in our sample and the two variables we are looking at for each locust are body length and wing length.

When we use the Spearman's rank coefficient to calculate a correlation, we first have to rank the data for each of the variables.

Locust	Body length x (mm)	Wing length y (mm)	Rank x	Rank y	$d$	$d^2$
1	15	7	9	10	-1	1

Locust	Body length x (mm)	Wing length y (mm)	Rank x	Rank y	$d$	$d^2$
2	10	6	10	9	1	1
3	80	32	1	1	0	0
4	45	23	6	5	1	1
5	53	19	5	6.5	-1.5	2.25
6	62	29	2	2	0	0
7	35	18	8	8	0	0
8	41	19	7	6.5	0.5	0.25
9	58	28	4	3	1	1
10	61	27	3	4	-1	1

If two equal values appear e.g. for Rank y at rank 6, then both are given the rank 6.5 (halfway between rank 6 & 7) and no values are given rank 6 or 7.

Next we calculate the difference between the ranks =  $d$  and then square this =  $d^2$

Then we find the sum of all the  $d^2$  values

$$\sum d^2 = 1+1+0+1+2.25+0+0+0.25+1+1 = 7.5$$

Now we can calculate the correlation coefficient using the formula:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 7.5}{7(7^2 - 1)} = 1 - \frac{45}{336} = 0.8661$$

An  $r_s$  value of +1 shows perfect positive correlation

An  $r_s$  value of -1 shows perfect negative correlation

An  $r_s$  value of 0 shows no correlation

To find out whether the  $r_s$  value we have calculated is sufficient evidence to reject our null hypothesis we need to refer to the critical values table. Some versions of this table have entries listed according to  $n$ , the number of items (locusts in this case,  $n = 10$ ). Some versions have entries listed by degrees of freedom.

The number of degrees of freedom is:

$$\text{d.f.} = n - 2 = 8$$

The critical value for the Spearman's rank correlation coefficient for  $n = 10$  or  $\text{df} = 8$  at  $p=0.05$  is 0.6485

Our calculated value for  $r_s$  is therefore greater than the critical value

$$0.8661 > 0.6485$$

Therefore we can reject the null hypothesis that there is no correlation between locust body size and wing size. We can accept the alternative hypothesis that there is a correlation between locust body size and wing size.

Our  $r_s$  is positive (+0.8661), therefore we have positive correlation: as the size of a locust increases there is a tendency for its wing length to increase.

Remember - correlation does not equal causation – we have shown that these two variables tend to change together, but we have not shown that there is a cause and effect.

### Assumptions

When performing a Spearman's rank correlation the following things are assumed about the data in order to trust the test result:

- We have a set of items and we have data on the same two variables from every one of those items
- Both variables are ordinal (i.e. they can be placed in order (ranked)) or a measurement.

We also assume that there are few (or no) tied ranks. In cases where there are many ties we can correct for this by using a slightly different formula but that is beyond the scope of the A Level Biology maths requirements.

### **The chi squared test**

When we want to look at distributions of frequencies and whether they differ from expected values we can use the chi squared ( $\chi^2$ ) test. Our expected frequencies can be based on previous observations from experiments, or simply an expectation that there should be equal proportions in each category.

For example, we cross two flowers with pink petals – we know that both of these plants are heterozygotes and they carry two co-dominant alleles, one for red petals and one for white petals. We then count the frequency of offspring that develop with either red, white or pink petals.

Our null hypothesis is that any differences in the observed numbers of offspring with white, red and pink petals from the expected numbers are due to chance. The alternative hypothesis would be that the differences are being caused by something other than chance – an unknown interaction with another gene perhaps.

We start by working out what our expected frequencies should be. A Punnett square is a good way to do this. In the table below the alleles present in the parental gametes are shown and then, within the outlined section, the four equally likely outcomes of each fertilisation event, giving the alleles present in the offspring and the **resulting appearance**

		Parental gamete alleles	
		Red	White
Parental gamete alleles	Red	Red/Red <b>Red</b>	Red/White <b>Pink</b>
	White	Red/White <b>Pink</b>	White/White <b>White</b>

From this simple table we can see that we expect to see frequencies in the offspring of White:Red:Pink petals at a ratio of 1:1:2.

Let's say we count a total of 160 offspring from the cross, we can therefore calculate the expected numbers of white, red and pink petals – let's compare that to the observed numbers in the table below

	Expected	Observed
<b>White</b>	40	28
<b>Red</b>	40	46
<b>Pink</b>	80	86

Is there a significant difference between the expected and observed frequencies?

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

The  $\chi^2$  value is calculated as follows:

$f_o$  = observed frequencies

$f_e$  = expected frequencies

$$\chi^2 = \frac{(28-40)^2}{40} + \frac{(46-40)^2}{40} + \frac{(86-80)^2}{80} = 4.95$$

Once again we must look up this value in the appropriate statistics data table and compare it to the critical value at the appropriate degrees of freedom.

There are 3 offspring types – red, white and pink, so  $n = 3$

Therefore degrees of freedom =  $n-1 = 2$

On the  $\chi^2$  table the critical value where  $p = 0.05$  and  $df = 2$  is 7.81

$4.95 < 7.81$  therefore our  $\chi^2$  value does not reach the critical value for significance at 2 degrees of freedom.

Therefore we cannot reject the null hypothesis: differences between our observed frequencies of petal colour and expected frequencies could be due to chance alone.

### Assumptions

When performing a Chi-squared test the following things are assumed about the data in order to trust the test result. There is a minimum sample size for performing the chi squared test – this is indicated by each expected value in a cell being  $>5$



## M1.9 – Select and use a statistical test

### Quiz

1. We measured the mass of nine sample adult males in each of two separate populations of elephants (**A** and **B**), and want to know if the means of the two populations are different.

Sample number	1	2	3	4	5	6	7	8	9
Population <b>A</b> mass of adult male (kg)	6000	5590	6124	5800	5987	6020	5900	6143	5699
Population <b>B</b> mass of adult male (kg)	4100	5900	4867	5010	5534	5321	5987	5350	5478

- a) Calculate the means and standard deviations for the two populations

**A LEVEL**  
**BIOLOGY A**  
**BIOLOGY B (ADVANCING BIOLOGY)**  
**Learner Activity**

- b) Which statistical test is appropriate for testing the hypothesis that there is a difference in the mean mass of adult male elephants between these two populations?

- c) Calculate whether there is a significant difference between these means

2. For which one or more of the following is a Spearman's rank correlation coefficient the appropriate statistical test to use?

- A** Comparing the relationship between grey seal pup size and fat reserves
- B** Comparing the frequency of different species of bluebell in a woodland
- C** Describing the relationship between the numbers of ladybirds and the numbers of aphids in 10 different meadows
- D** Comparing the average growth of bacteria on two types of agar plate, where one has been treated with penicillin

3. Equal amounts of two types of the bacteria *E.coli* are mixed together in a volumetric flask, one of these populations of *E.coli* is carrying an antibiotic resistance gene. The mixture is then poured out onto agar plates that have been inoculated with penicillin and incubated for 24 hours. Based on previous experiments, when we count the bacteria, we expect there to be twice as many colonies on the plate with the resistance gene as without. If we found 846 colonies on our plates the next day, and 432 of them carried the resistance marker, does this differ significantly from our expected frequency?

## M1.10 – Understanding measures of dispersion including standard deviation and range

### Tutorials

Learners 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

### Standard Deviation

You will already be familiar with calculating the “average” or “mean” of a dataset. This is a very useful summary value, but it can also be interesting to look at the distribution of scores in the data. One way to do this is to calculate the range of scores in the data. Calculating the range of scores is easy – simply subtract the smallest score from the largest score.

Here is an example - There are 11 people on a bus and their ages are as follows: 15, 17, 19, 22, 36, 39, 40, 44, 53, 54 & 90

The oldest person on the bus is 90, and the youngest person is 15

The **interval** within which all the data in the set falls is 15 years to 90 years. This sometimes gets confused with the range, as it is natural in everyday speech to say ‘the ages of people on the bus ranged from 15 to 90 years’. However, to find the range, by its strict definition, we need to do that very simple calculation:

Subtract smallest score from largest score

the **range** is  $90 - 15 = 75$  years.

One problem with using the range is that because it only uses the highest score and lowest score it is dramatically affected by extreme scores.

For example, the age range on the bus was 75 years, but if we compute the range while excluding the score of 90 and use the next highest age score, which was 54, we see that the range drops dramatically from 75 years to 39 years. Almost half the size!

Another problem with the range is that it is little help in making inferences about a whole population when the data we have is just a sample.

In this case if we were using the people on the bus as a sample of the whole population of a town, knowing the range in our sample is 75 only really tells us that the range in the whole population must be at least 75.

A different, and much more useful, way of measuring dispersion in sample data is to calculate the sample standard deviation. This is less influenced by outlying results and it allows us to make inferences about the whole population from which our sample is drawn.

The greater the standard deviation, the greater the spread of the data.

The standard deviation is calculated in a few steps.

One simple way to describe what we are aiming to get out of these steps is that we are going to work out **on average** how different from the mean each of our data points is. If the data are widely spread we'll get a big number, whereas if they are generally close to the mean we'll get a small number. But we're going to do this in two slightly strange ways, which give us a far more useful statistic than a 'straightforward' mean of the differences:

- We will square the differences, to deliberately accentuate the effect of the larger ones
- We will divide by one less than the number of data points (i.e.  $n-1$  instead of  $n$ )

So first we need to know the mean of our data. Let's use the ages of the people on the bus again.

$$\bar{x} = \frac{15+17+19+22+36+39+40+44+53+54+90}{11} = 39$$

Now we subtract the mean value from each of our observed values (the individual ages of people on the bus)

$$(x_i - \bar{x}) = (15-39) (17-39) (19-39) (22-39) (36-39) (39-39) (40-39) (44-39) (53-39) (54-39) (90-39)$$

$$(x_i - \bar{x}) = -24, -22, -20, -17, -3, 0, 1, 5, 14, 15, 51$$

Next we square all these values. This increases the effect on the final result of the numbers further away from the mean (which we want to do to give us the most useful statistic for drawing inferences about the population – it is **not** done simply to make all the deviations positive!). We sum the values next and you can see that the '51' makes a big contribution (2601) to the resulting sum now that it is squared.

$$\sum (x_i - \bar{x})^2 = (-24)^2 + (-22)^2 + (-20)^2 + (-17)^2 + (-3)^2 + (0)^2 + (1)^2 + (5)^2 + (14)^2 + (15)^2 + (51)^2 = 4806$$

This sum of the squares is a good measure of the accuracy of our model, but it needs to reflect how much data we collected (i.e. we are getting an **average** squared difference) and so we divide the sum of squares by the number of observations minus 1 ( $n-1$ ), (More on this in a moment)

$$\frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{4806}{10} = 480.6$$

There is only one problem with this number – it gives us the average difference between the mean and the observations made, but it gives us this in units squared. So if we square root this value it will turn back into the original unit of measurement (age in years).

$$\sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{4806}{10}} = \sqrt{480.6} = 21.9$$

The standard deviation (s) is a very useful tool; the smaller this value is, on average the closer our individual data points sit towards the mean.

With a normal distribution of data we would expect the following to hold true:

- 68% of the data lie within one standard deviation either side of the mean
- 95% of the data lie within 2 standard deviations either side of the mean.

### **n-1**

Why do we divide by n-1 rather than n?

If we really want to know the 'root mean square difference from the mean' (which is a pretty good description of what the standard deviation is) why don't we divide by n just like calculating other means?

In fact if we were calculating the standard deviation of a whole population (e.g. if we were only interested in the people on the bus as a (tiny) whole population and did not want to use them as a sample to make estimates about a larger population) we **would** divide by n.

But in most biological data handling situations we actually have a sample (because we don't have time to measure every single individual in the whole population) and we use our calculation of the sample standard deviation to make inferences about the population. Using n-1 in our calculation is essentially a correction factor that makes the number we come up with from our sample a better estimate of the standard deviation of the population.

So for AS and A Level Biology we stick to using the n-1 formula.

If you are using a calculator or a spreadsheet to calculate standard deviation make sure you press the right button, or use the right embedded formula.

## M1.10 – Understanding measures of dispersion including standard deviation and range

### Quiz

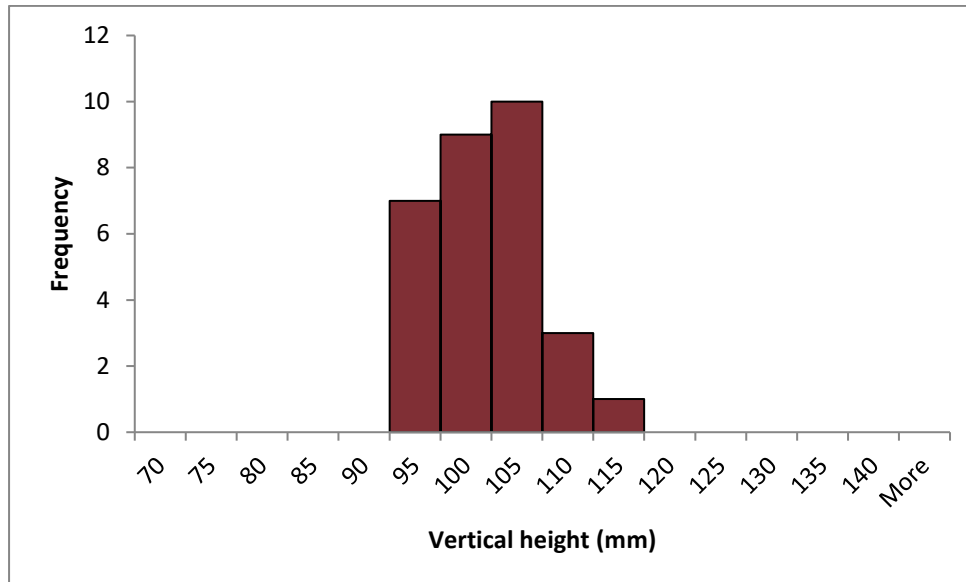
- Below are the ages (in months) of Queen ants of the genus *Cardiocondyla* from two geographically isolated populations. For each population a random sample of 11 queens was taken and the ages recorded. Calculate the mean age and standard deviation for queens from each population. Which of these two populations has the smallest standard deviation?

	Queen ant	1	2	3	4	5	6	7	8	9	10	11
Queen age in months	Population A	6	8	10	8	9	6	7	12	14	11	9
	Population B	8	8	9	14	16	9	15	13	12	11	8

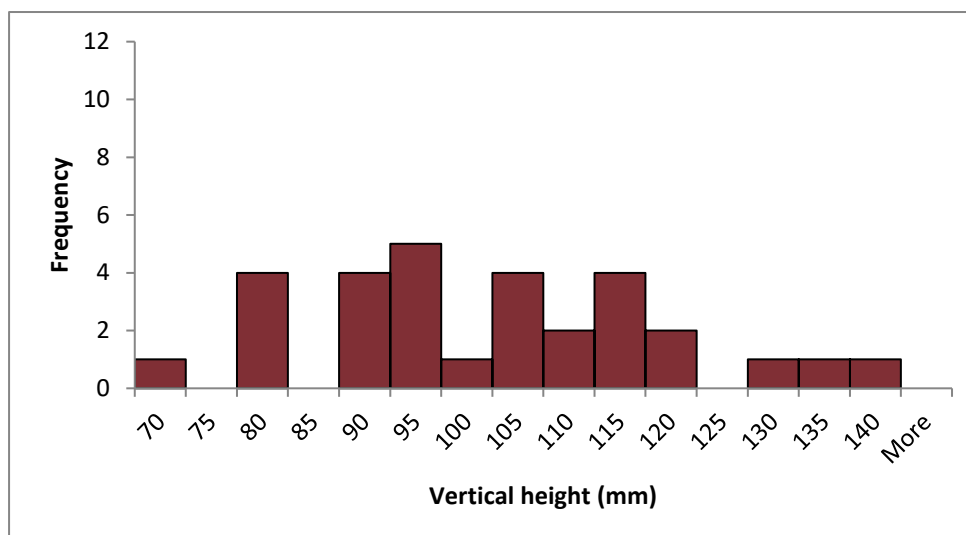
**A LEVEL**  
**BIOLOGY A**  
**BIOLOGY B (ADVANCING BIOLOGY)**  
**Tutorial**

2. The vertical jump height (mm) was measured of two separate populations (A & B) of fleas. Below are two histograms of the distributions of jump heights in the two populations. Both populations had a normal distribution around a common mean jump height of 100 mm. Which population has the greatest standard deviation?

A)



B)





## M1.11 Identify uncertainties in measurements and use simple techniques to determine uncertainty when data are combined

### Tutorials

Learners may be tested on their ability to:

- calculate percentage error where there are uncertainties in measurement.

### Uncertainties

Any measurement we make tells us about a property of something. It might tell us how long something is, how heavy it is, how much coverage something has. The measurement gives a number to that property.

To make measurements we almost always need an instrument of some kind: a ruler, a thermometer, a quadrat, a top pan balance etc.

The measurement is recorded as both a number and a unit:

e.g. how hot is this flask of water?

37, degrees Celsius

When we make a measurement there is always some level of uncertainty. A well-made instrument should be trustworthy and give precise, repeatable measurements. But for every measurement there is always a margin of doubt. We might describe this as accurate to within a given value of uncertainty. An example might be a measuring cylinder, which accurately measures volumes of liquid to within  $0.5 \text{ cm}^3$ . For example we might measure out  $300 \text{ cm}^3$  of liquid  $\pm 0.5 \text{ cm}^3$ , which means the true volume might be anywhere between  $299.5 \text{ cm}^3$  and  $300.5 \text{ cm}^3$ .

For this instrument  $\pm 0.5 \text{ cm}^3$  is the absolute uncertainty. It doesn't matter how much liquid we measure; the measurement could always be out by this amount.

The relative uncertainty or **percentage error** is the ratio of absolute uncertainty to the original measurement, expressed as a percentage.

For example, when measuring  $300 \text{ cm}^3$  the relative uncertainty is  $\frac{0.5}{300} * 100 = 0.17\%$

However, if I measured out only 50 cm<sup>3</sup> the relative uncertainty is  $\frac{0.5}{50} * 100 = 1\%$

So we can see that the amount of relative uncertainty is not fixed, it is dependent on the amount of absolute uncertainty afforded by the instrument, **and** the size of the measurement we are making.

When making more than one measurement, we must account for the absolute uncertainty for each new measurement.

If I measure out 65 cm<sup>3</sup> of hydrochloric acid then there is an absolute uncertainty of 0.5 cm<sup>3</sup>, If I measure out and add 30 cm<sup>3</sup> of sodium hydroxide solution, then there is also an absolute uncertainty of 0.5 cm<sup>3</sup>. Giving a total absolute uncertainty of 1 cm<sup>3</sup>. Therefore my combined volume is 95 cm<sup>3</sup> ± 1 cm<sup>3</sup>.

My relative uncertainty for this measurement is  $\frac{1}{95} * 100 = 1.05\%$

The same idea applies when we are looking at the change in a value, by subtracting the initial value from the final value.

For example we could use a top pan balance for simple potometry (measuring the water loss from a plant by change in mass).

For this example assume we are using a top pan balance with an absolute uncertainty of +/- 0.1 g

We take our initial mass measurement: 117.3 g

24 h later we take our final measurement: 110.0 g

To find the change in mass we subtract the initial value from the final value:

$$110.0 \text{ g} - 117.3 \text{ g} = - 7.3 \text{ g}$$

The plant has changed mass by – 7.3 g (i.e. its mass has fallen by 7.3 g)

The **absolute** uncertainty associated with this figure is the sum of the absolute uncertainties of each measurement:

$$0.1 \text{ g} + 0.1 \text{ g} = 0.2 \text{ g}$$

The **percentage error** or **relative uncertainty** is found, as in the previous examples, by dividing the absolute uncertainty by the measured value and expressing the result as a percentage:

$$\frac{0.2}{7.3} * 100 = 2.7\%$$

## M1.11 Identify uncertainties in measurements and use simple techniques to determine uncertainty when data are combined

### Quiz

1. A microscope graticule allows fine-scale measurements to be made under a microscope. If the graticule is accurate to  $\pm 0.5 \mu\text{m}$ , and a protozoan parasite *Trypanosoma* is measured as  $50 \mu\text{m}$ , calculate the percentage error for this measurement.
2. Cell cultures of the bacteria *E. coli* can be measured by a spectrophotometer to give an accurate (to within 2%) reading of bacteria  $\text{cm}^{-3}$   
A sample has been calculated as containing  $3 \times 10^9$  bacteria  $\text{cm}^{-3}$   
Calculate the absolute uncertainty of this measurement.
3. A plant shoot is measured for growth over a 5-day time period. Every morning it was measured with a ruler accurate to  $\pm 0.5 \text{ mm}$  and the height recorded as show below.  
Calculate the difference in height between days 1 and 5 and state the percentage error in this measurement.

Day	1	2	3	4	5
Height (mm)	8	11	16	21	24

**Produced in collaboration with the University of East Anglia**