# Maths skills – M1 Handling data

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

Significant figures demonstrate a level of accuracy and are important in all sorts of biological contexts, when reporting experimental data and for any calculation. Calculated results can only be reported to the limits of the least accurate measurement (the lowest level of accuracy). This means your answer can only have the same number of significant figures as the piece of data with the lowest number of significant figures. Reporting significant figures may involve rounding up (5 or above) or down (4 or below). Be careful of zeros – any at the front of the number are not significant figures. However zeros must be reported if they occur within or at the end of the number.

### M1.2 – Find arithmetic means

The mean is calculated by adding together all the values and dividing by the number of values. As a formula this is written as:



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.

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

In biology you need to be able to understand and interpret frequency tables for a variety of biological contexts. In addition, you must know the difference between histograms and bar charts, when to use them, how to plot them and how to interpret them.

Below are key factors you need to consider when creating frequency tables, bar charts or histograms:

|  |  |
| --- | --- |
| **Frequency tables** | **Bar charts and histograms** |
| Headings | Title |
| Units | Axes labels (with units) |
| Consistent use of decimal places | Independent Variable on x axis |
|  | Dependent Variable on y axis |
|  | Plot data carefully |
|  | Graph is >50% of space available |

You also need to remember the important differences between histograms and bar charts:

|  |  |
| --- | --- |
| **Bar charts** | **Histogram** |
| Discrete independent variable data | Continuous independent variable data |
| Bars the same width | Bar width may differ |
| Bars not touching | Bars touch |

Discrete data plotted with bar charts can only take specific values (there are no ‘in between’ values) and may include variables such as eye colour, species name, . On the other hand continuous data is plotted using histograms for variables measured to a specified accuracy such as height, weight, and age.

For histograms you must make sure that the different category labels do not overlap as your data cannot be in two different categories. Unlike bar charts, the width of bars in a histogram do not all need to be the same. This is because it is the area of the bar that needs to be proportional to the frequency. In situations where the class width of categories differ, you need to calculate the frequency density, and plot these values on the y axis of your histogram.

### M1.4 – Understand simple probability

Estimating probabilities is a fundamental part of working out the likelihood of an event occurring. By understanding probabilities, we can make inferences about the results we are collecting and draw conclusions. When an event is random, it does not mean that it is rare, just that it is unpredictable. Probabilities allow us to predict the likelihood of an event occurring and identify patterns that occur over time, even if we cannot predict the outcome of a single event.

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

Sampling is a way of designing an experiment so that you can measure a representative part of a population. Sampling can be random, where samples are taken in an unbiased way from the whole population or system being studied, or non-random where samples are taken in a pre-defined pattern.

Choosing whether to sample randomly or non-randomly depends on whether distribution is an important part of the question, such as “how does the density of bluebells change as I move away from the base of an oak tree”. This is a good example for non-random sampling. Otherwise random sampling is a good way to avoid bias.

A community dominated by one or two species is considered to be less diverse than one in which several different species have a similar abundance.

Simpson's Diversity Index is a measure of diversity which takes into account the number of species present (species richness), as well as the relative abundance of each species (species evenness). As species richness and evenness increase, so diversity increases.

### M1.6 – Understand the terms 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, mentioned previously in section M1.2, is the sum of the data values divided by the number of data values. The median is the middle value of a data set and requires the data to be put in ascending order first. The mode is the most frequently occurring value and is usually the easiest to spot. Generally the mean is the most useful statistical measure, except when there are outliers in the data when it may be more appropriate to use the median.

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

Drawing a scatterplot allows you to see the relationship between two continuous variables. If a relationship does exist between two continuous variables it can be described as a correlation: a change in one variable tends to come with a change in the other variable. Correlations can be linear or non-linear, positive or negative, and strong or weak. A combination of all three of these terms can be used to describe a correlation e.g. a strong, positive, linear correlation.

Remember correlation does not imply causation!

### M1.8 – Make order of magnitude calculations

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.

The formula used to calculate magnification is:



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.

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

Statistical tests allow us to draw conclusions about whether the results we have obtained are likely to have come about by chance (the null hypothesis) or whether we have discovered a pattern or process (the alternative hypothesis).

If the data we have gathered allows us to reject the null hypothesis, we do so at a certain confidence level (usually 95%, also known as p=0.05). So when we reject the null hypothesis it doesn’t mean our alternative hypothesis is certainly true, just that it is supported by the evidence collected so far.

On the other hand, if the test we have carried out shows that we cannot reject the null hypothesis, it does not mean the null hypothesis is true. It just means we have failed to disprove it with the data under analysis.

In order for our statistical tests to make meaningful inferences about biological systems and processes, it is important that we use them appropriately.

*t*-tests can be used for comparing mean differences between groups, the Student’s *t*-test if they are two independent groups, or the paired *t*-test if it is the same group measured before and after an event/manipulation.

Spearman’s rank correlation coefficient is used to look at relationships between two variables; is there a consistent pattern where as one value rises the other also tends to rise/fall? Remember a correlation can be positive or negative.

The chi squared test is used to identify whether frequencies of observations deviate from a pattern of equal distribution, or another expected distribution.

The value generated by a statistical test can be used to estimate the probability that the results you have obtained could have occurred by chance. To do this you must use the right table, for example you must look in a *t*-value table when using a *t*-test. By using your test value and the degrees of freedom you can look up the approximate p-value for your data.

The p-value is the probability from 0 to 1 that your results could have been obtained by chance (i.e. that the null hypothesis is true), it is generally agreed that the threshold for rejecting the null hypothesis is set at p<0.05.

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

Dispersion is the variability we find in our data. It can be measured in several different ways. The simplest is the range, which looks at the difference between the highest and lowest scores in a dataset. Although the range is easy to calculate, it is heavily influenced by extreme scores.

Another way of analysing dispersion is to calculate the standard deviation. This is the square root of the average difference between the mean and each of the data points, it is a good measure of the ‘fit’ of our data and allows us to make quantitative inferences about the population from which a sample was taken.

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

It is important to understand the difference between absolute and relative uncertainty.

Absolute uncertainty is a fixed value for any given measuring instrument.

Relative uncertainty is a value that changes dependent on the value of the measurement and the absolute uncertainty.

When adding or subtracting measurements, you must add the absolute uncertainty values together, to get an overall measure of the absolute uncertainty. From this absolute uncertainty you can calculate the relative uncertainty or ‘percentage error’.

### Questions:

**Questions M1.1**

1) 0.30202 to 2 sig fig =

2) 0.675 to 2 sig fig =

3) 7.006 to 3 sig fig =

4) 6.001 to 2 sig fig =

**Questions M1.2**

Students had a competition to grow the tallest sunflower. Their measurements (in cm) are shown below. Calculate the mean sunflower height.

102 95 89 110 79 82 94 87 93 81

Mean = …………………. cm

The lengths of mitochondria in a cell were measured (µm) and recorded. What is the mean mitochondrion length?

2.4 3.4 4.5 7.0 6.8 5.6 5.7 7.0 5.9 6.1

Mean = ……………………….. µm

**Questions M1.3**

For the below data sets:

1. Determine whether a histogram or bar chart is the more appropriate graph to plot with reasons
2. Plot the graph
3. Number of flower heads with different masses of flowers

|  |  |
| --- | --- |
| **Mean mass of flowers per flower head (g)** | **Frequency** |
| 5.0-5.4 | 42 |
| 5.5-5.9 | 22 |
| 6.0-6.4 | 53 |
| 6.5-6.9 | 31 |
| 7.0-7.4 | 20 |
| 7.5-7.9 | 10 |



1. Number of flowers of different colours

|  |  |
| --- | --- |
| **Flower colour** | **Frequency** |
| White | 46 |
| Pink | 92 |
| Red | 42 |

**Questions M1.4**

1. What is the probability of getting one ‘head’ and two ‘tails’ when three coins are tossed?
2. Two beetles with shiny wings are crossed. The resultant offspring are produced in a ratio of 3:1 shiny to dull wings. If we know a single gene controls this trait, what is the likely reason for the appearance of the dull wing phenotype and why?
3. A *Drosophila melanogaster* cross is established with one parent homozygous for the wild type *vestigial* allele and the other carrying one copy of the wild type allele and one copy of the mutant allele.
4. What is the probability of offspring carrying at least one copy of the mutant allele?
5. What is the probability of offspring displaying the mutant *vestigial* phenotype?

**Questions M1.5**

1. What are the two main approaches to sampling?

1. We wish to estimate the mean wing length in a population of *Drosophila* 10% of which display the *curly wing* mutation. We aim to measure the wing lengths in 250 flies.
2. What method should be employed to take this sample?
3. What numbers of wildtype and *curly wing* flies should be included in this sample?
4. Two equal areas of the New Forest and the Forest of Dean were surveyed for numbers and diversity of tree species.

For both locations calculate the Simpson’s Index of Diversity.

Which of the two forests has the higher biodiversity according to this survey?

|  |  |
| --- | --- |
|  | **Numbers** |
| **Species** | **New Forest** | **Forest of Dean** |
| English oak | 35 | 30 |
| Ash | 52 | 48 |
| Beech | 59 | 74 |
| Birch | 25 | 36 |
| Sweet chestnut | 5 | 0 |
| Yew | 3 | 2 |
|  |  |  |

**Question M1.6**

Ecologists wanted to compare the number of buttercups and dandelions in a field. Using quadrats they counted the numbers of each plant in 10 randomly selected 1 m2 areas. Calculate and compare the mean, median and modes for each data set. Which is the most appropriate statistical measure to report for each data set and why?

|  |  |
| --- | --- |
| **Number of buttercups** | **Number of dandelions** |
| 9 | 18 |
| 15 | 19 |
| 58 | 16 |
| 12 | 22 |
| 10 | 21 |
| 13 | 16 |
| 15 | 20 |
| 11 | 17 |
| 14 | 20 |
| 15 | 72 |

**Questions M1.7**

1. The male gray tree frog produces mating calls at regular intervals, but this interval frequency is thought to be affected by the air temperature. Plot the data collected, add a trendline and describe the relationship observed

|  |  |
| --- | --- |
| **Male call interval (s)** | **Temperature (°C)** |
| 2 | 16 |
| 4 | 26 |
| 2 | 18 |
| 3 | 20 |
| 3 | 24 |
| 4 | 19 |
| 6 | 32 |
| 3 | 29 |
| 6 | 30 |
| 5 | 28 |
| 3 | 21 |
| 2 | 16 |
| 2 | 23 |
| 1 | 11 |
| 1 | 16 |
| 3 | 19 |
| 2 | 11 |
| 5 | 26 |
| 3 | 19 |
| 5 | 27 |
| 2 | 12 |
| 1 | 11 |
| 2 | 17 |
| 2 | 11 |
| 4 | 27 |
| 3 | 22 |
| 4 | 18 |
| 1 | 11 |
| 1 | 11 |

**Questions M1.8**

This image shows the capillaries in a 1 mm2 area of rat retina. What is the diameter of the capillaries?



**B0004116** **Credit** [Jean Wade and Linda Sharp](https://wellcomeimages.org/indexplus/result.html?wi_credit_line%3atext=%22Jean%20Wade%20and%20Linda%20Sharp%22&%24%3dsort=sort%20sortexpr%20image_sort&%2asform=wellcome-images&_IXACTION_=query&_IXFIRST_=1&_IXSPFX_=templates%2fb&_IXFPFX_=templates%2ft&%24%20with%20image_sort=.), Wellcome Images
Branching blood vessels in the retina

Confocal image of the retinal capillary bed of a rat. This image shows an area of 1 square mm.

[Confocal micrograph](https://wellcomeimages.org/indexplus/result.html?wi_technique%3atext=%22Confocal%20micrograph%22&%24%3dsort=sort%20sortexpr%20image_sort&%2asform=wellcome-images&_IXACTION_=query&_IXFIRST_=1&_IXSPFX_=templates%2fb&_IXFPFX_=templates%2ft&%24%20with%20image_sort=.)
**Collection:** [Wellcome Images](https://wellcomeimages.org/indexplus/result.html?wi_library_dept%3atext=%22Wellcome%20Images%22&%24%3dsort=sort%20sortexpr%20image_sort&%2asform=wellcome-images&_IXACTION_=query&_IXFIRST_=1&_IXSPFX_=templates%2fb&_IXFPFX_=templates%2ft&%24%20with%20image_sort=.)
**Library reference no.:** Contributor Reference IMAGE 05
Copyrighted work available under Creative Commons Attribution only licence CC BY 4.0 <http://creativecommons.org/licenses/by/4.0/>

**Questions M1.9**

1. We are interested in determining whether there is an effect of the addition of nitrate fertiliser on the mean height of Brassica crops. Two fields are treated identically apart from the use of fertiliser on one but not the other. Mature plants from each field were then chosen at random and the heights measured.
2. Generate a null and alternative hypothesis on the effect of fertiliser on crop growth
3. The sample data collected from the two fields is as follows. Calculate the *t*-value and determine if there is a significant difference between the two treatments

|  |  |  |
| --- | --- | --- |
|  | Height of crop when nitrate fertiliser added (cm)n = 23 | Height of crop when no fertiliser added (cm)n = 24 |
| Mean | 142 | 101 |
| Standard deviation | 23 | 18 |



 = 6.8

1. The average time between the production of two RNA transcripts from the same strand of DNA in a cell is known as the ‘synthesis time’, we wish to know whether products that take longer to be synthesised by a cell also last longer or whether they are targeted for degradation at the same rate. To do this the “half-life” (the time needed for half of a batch of mature RNAs to degrade) was measured for each transcript separately.

|  |  |  |
| --- | --- | --- |
| **RNA molecule** | **Synthesis time (s)** | **“Half-life”(s)** |
| 1 | 240 | 98 |
| 2 | 230 | 203 |
| 3 | 1000 | 180 |
| 4 | 78 | 226 |
| 5 | 194 | 162 |
| 6 | 182 | 173 |
| 7 | 675 | 156 |
| 8 | 345 | 146 |
| 9 | 982 | 186 |
| 10 | 112 | 178 |

1. Produce null and alternative hypotheses for this experiment
2. What is the rs value for this data – is this a significant relationship?
3. Describe the results in terms of your hypotheses

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **RNA molecule** | **Synthesis time (s)** | **Rank** | **“Half-life”(s)** | **Rank** | **Difference in rank** | **Difference squared** |
| 1 | 240 | 5 | 98 | 10 | 5 | 25 |
| 2 | 230 | 6 | 203 | 2 | 4 | 16 |
| 3 | 1000 | 1 | 180 | 4 | 3 | 9 |
| 4 | 78 | 10 | 226 | 1 | 9 | 81 |
| 5 | 194 | 7 | 162 | 7 | 0 | 0 |
| 6 | 182 | 8 | 173 | 6 | 2 | 4 |
| 7 | 675 | 3 | 156 | 8 | 5 | 25 |
| 8 | 345 | 4 | 146 | 9 | 5 | 25 |
| 9 | 982 | 2 | 186 | 3 | 1 | 1 |
| 10 | 112 | 9 | 178 | 5 | 4 | 16 |

**

**Questions M1.10**

1. Here is a dataset of the cell sizes of a sample from a population of the single-celled eukaryote *Paramecium bursaria*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **Size (µm)** | 80 | 150 | 95 | 110 | 210 | 140 | 97 | 101 | 85 | 134 |

1. State the interval that covers one standard deviation above and one standard deviation below the mean.

1. Any scores which lie more than three standard deviations above or below the mean could be considered extreme scores, are there any scores which fit this criterion?

**Questions M1.11**

1. A 10 µl pipette is guaranteed by its manufacturer as accurate to ±0.03 µl, this was tested in the lab by drawing up 10 µl of water repeatedly and verifying the volumes independently. The volumes were verified as follows:

9.9978; 10.0062; 10.0020; 10.0047; 9.9998; 9.9982; 10.0023; 10.0034; 10.0012; 9.9994

Is this pipette accurate to within the manufacturer’s specifications?

1. 2.5 ml of a liquid bacterial culture is transferred into a larger volume of 9 ml of fresh broth. The pipette used to transfer the culture is accurate to ±0.03 ml, and the measuring cylinder for the broth is accurate to ±0.05 ml.
2. What is the new volume and the absolute uncertainty?
3. What is the relative uncertainty of the pipette measurement?

### Answers:

**Questions M1.1**

0.30

1) 0.30202 to 2 sig fig =

0.68

2) 0.675 to 2 sig fig =

7.01

3) 7.006 to 3 sig fig =

6.0

4) 6.001 to 2 sig fig =

**Questions M1.2**

Students had a competition to grow the tallest sunflower. There measurements (in cm) are shown below. Calculate the mean sunflower height.

102 95 89 110 79 82 94 87 93 81

Mean = 912/10 = 91.2 cm

The lengths of mitochondria in a cell were measured (µm) and recorded. What is the mean mitochondria length?

2.4 3.4 4.5 7.0 6.8 5.6 5.7 7.0 5.9 6.1

52/9 = 5.77778

Mean = 5.78 (or 5.8) µm

**Questions M1.3**

For the below data sets:

1. Determine whether a histogram or bar chart is the more appropriate graph to plot with reasons
2. Plot the graph
3. Number of flower heads with different masses of flowers

|  |  |
| --- | --- |
| **Mean mass of flowers per flower head (g)** | **Frequency** |
| 5.0-5.4 | 42 |
| 5.5-5.9 | 22 |
| 6.0-6.4 | 53 |
| 6.5-6.9 | 31 |
| 7.0-7.4 | 20 |
| 7.5-7.9 | 10 |

Histogram – continuous data



1. Number of flowers of different colours

|  |  |
| --- | --- |
| **Flower colour** | **Frequency** |
| White | 46 |
| Pink | 92 |
| Red | 42 |

Bar chart – discrete data

**Questions M1.4**

1. What is the probability of getting one ‘head’ and two ‘tails’ when three coins are tossed?

3/8

1. Two beetles with shiny wings are crossed. The resultant offspring are produced in a ratio of 3:1 shiny to dull wings. If we know a single gene controls this trait, what is the likely reason for the appearance of the dull wing phenotype and why?

The emergence of a new phenotype suggests that the parents were both carrying a recessive allele for dull wings. 3:1 is the expected ratio that would result from such a cross – other theorised parental combinations of alleles for a single gene trait would not produce the ratio of offspring observed.

1. A *Drosophila melanogaster* cross is established with one parent homozygous for the wild type *vestigial* allele and the other carrying one copy of the wild type allele and one copy of the mutant allele.
	1. What is the probability of offspring carrying at least one copy of the mutant allele?

0.5

* 1. What is the probability of offspring displaying the mutant *vestigial* phenotype?

0

**Questions M1.5**

1. What are the two main approaches to sampling?

Random and non-random

1. We wish to estimate the mean wing length in a population of *Drosophila* 10% of which display the *curly wing* mutation. We aim to measure the wing lengths in 250 flies.
2. What method should be employed to take these samples?

Stratified random sampling

1. What numbers of wildtype and *curly wing* flies should be included in this sample?

25 curly wings and 225 wildtype

1. Two equal areas of the New Forest and the Forest of Dean were surveyed for numbers and diversity of tree species. For both locations calculate the Simpson’s Index of Diversity. Which of the two forests has the higher biodiversity according to this survey?

The New forest has a greater D value – therefore the higher biodiversity

|  |  |
| --- | --- |
|  | **Numbers** |
| **Species** | New Forest | Forest of Dean |
| English oak | 35 | 30 |
| Ash | 52 | 48 |
| Beech | 59 | 74 |
| Birch | 25 | 36 |
| Sweet chestnut | 5 | 0 |
| Yew | 3 | 2 |
| Simpson’s Index | 0.75 | 0.72 |

**Question M1.6:**

Ecologists wanted to compare the number of buttercups and dandelions in a field. Using quadrats they counted the numbers of each plant in 10 randomly selected 1 m2 areas. Calculate and compare the mean, median and modes for each data set. Which is the most appropriate statistical measure to report for each data set and why?

|  |  |
| --- | --- |
| **Number of buttercups** | **Number of dandelions** |
| 9 | 18 |
| 15 | 19 |
| 58 | 16 |
| 12 | 22 |
| 10 | 21 |
| 13 | 16 |
| 15 | 20 |
| 11 | 17 |
| 14 | 20 |
| 15 | 72 |

|  |  |  |
| --- | --- | --- |
|  | Buttercups | Dandelions |
| Mean | 17.2 | 24.1 |
| Median | 13.5 | 19.5 |
| Mode | 15 | 16 |

There are outliers in both data sets therefore the median is more representative to report.

**Questions M1.7**

1. The male gray tree frog produces mating calls at regular intervals, but this interval frequency is thought to be affected by the air temperature. Plot the data collected add a trendline and describe the relationship observed

|  |  |
| --- | --- |
| **Male call interval (s)** | **Temperature (°C)** |
| 2 | 16 |
| 4 | 26 |
| 2 | 18 |
| 3 | 20 |
| 3 | 24 |
| 4 | 19 |
| 6 | 32 |
| 3 | 29 |
| 6 | 30 |
| 5 | 28 |
| 3 | 21 |
| 2 | 16 |
| 2 | 23 |
| 1 | 11 |
| 1 | 16 |
| 3 | 19 |
| 2 | 11 |
| 5 | 26 |
| 3 | 19 |
| 5 | 27 |
| 2 | 12 |
| 1 | 11 |
| 2 | 17 |
| 2 | 11 |
| 4 | 27 |
| 3 | 22 |
| 4 | 18 |
| 1 | 11 |
| 1 | 11 |

A strong positive correlation.

**Questions M1.8**

This image shows the capillaries in a 1 mm2 area of rat retina. What is the diameter of the capillaries?



**B0004116** **Credit** [Jean Wade and Linda Sharp](https://wellcomeimages.org/indexplus/result.html?wi_credit_line%3atext=%22Jean%20Wade%20and%20Linda%20Sharp%22&%24%3dsort=sort%20sortexpr%20image_sort&%2asform=wellcome-images&_IXACTION_=query&_IXFIRST_=1&_IXSPFX_=templates%2fb&_IXFPFX_=templates%2ft&%24%20with%20image_sort=.), Wellcome Images
Branching blood vessels in the retina

Confocal image of the retinal capillary bed of a rat. This image shows an area of 1 square mm.

[Confocal micrograph](https://wellcomeimages.org/indexplus/result.html?wi_technique%3atext=%22Confocal%20micrograph%22&%24%3dsort=sort%20sortexpr%20image_sort&%2asform=wellcome-images&_IXACTION_=query&_IXFIRST_=1&_IXSPFX_=templates%2fb&_IXFPFX_=templates%2ft&%24%20with%20image_sort=.)
**Collection:** [Wellcome Images](https://wellcomeimages.org/indexplus/result.html?wi_library_dept%3atext=%22Wellcome%20Images%22&%24%3dsort=sort%20sortexpr%20image_sort&%2asform=wellcome-images&_IXACTION_=query&_IXFIRST_=1&_IXSPFX_=templates%2fb&_IXFPFX_=templates%2ft&%24%20with%20image_sort=.)
**Library reference no.:** Contributor Reference IMAGE 05
Copyrighted work available under Creative Commons Attribution only licence CC BY 4.0 <http://creativecommons.org/licenses/by/4.0/>

object is 1 mm x 1 mm

Image has sides of 143 mm therefore magnification is x 143

Image of capillary has diameter of 1 mm

Therefore capillary object has diameter of 1 / 143 = 0.007 mm = 7 µm

**Questions M1.9**

1. We are interested in determining whether there is an effect of the addition of nitrate fertiliser on the mean height of Brassica crops. Two fields are treated identically apart from the use of fertiliser on one but not the other. Mature plants from each field were then chosen at random and the heights measured.
2. Generate a null and alternative hypothesis on the effect of fertiliser on crop growth

Null – There is no effect of fertiliser application on the mean height of Brassica crops

Alternative – There is an effect of fertiliser application on the mean height of Brassica crops *or* The Brassica crop exposed to fertiliser will be taller than the brassica not exposed to fertiliser.

1. The sample data collected from the two fields is as follows. Calculate the *t*-value and determine if there is a significant difference between the two treatments

|  |  |  |
| --- | --- | --- |
|  | Height of crop when nitrate fertiliser added (cm)n = 23 | Height of crop when no fertiliser added (cm)n = 24 |
| Mean | 142 | 101 |
| Standard deviation | 23 | 18 |

Should apply the Student’s *t*-test – two independent groups



 = 6.8

*t* = 6.8

degrees of freedom = n1 - 1 + n2 - 1 = 45

6.8 is well in excess of the threshold for p=0.05 and indeed for p=0.01. Therefore there is a significant difference in the sample means and we can reject the null hypothesis.

The crop treated with nitrate fertiliser grew taller.

1. The average time between the production of two RNA transcripts from the same strand of DNA in a cell is known as the ‘synthesis time’, we wish to know whether products that take longer to be synthesised by a cell also last longer or whether they are targeted for degradation at the same rate. To do this the “half-life” (the time needed for half of a batch of mature RNAs to degrade) was measured for each transcript separately.

|  |  |  |
| --- | --- | --- |
| **RNA molecule** | **Synthesis time (s)** | **“Half-life”(s)** |
| 1 | 240 | 98 |
| 2 | 230 | 203 |
| 3 | 1000 | 180 |
| 4 | 78 | 226 |
| 5 | 194 | 162 |
| 6 | 182 | 173 |
| 7 | 675 | 156 |
| 8 | 345 | 146 |
| 9 | 982 | 186 |
| 10 | 112 | 178 |

1. Produce null and alternative hypotheses for this experiment

Null – There is no correlation between synthesis time and half-life for RNA production. Alternative – The synthesis time correlates with the half-life of RNA molecules

1. What is the rs value for this data – is this a significant relationship?

Sum of difference squared = 202

1. Describe the results in terms of your hypotheses

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **RNA molecule** | **Synthesis time (s)** | **Rank** | **“Half-life”(s)** | **Rank** | **Difference in rank** | **Difference squared** |
| 1 | 240 | 5 | 98 | 10 | 5 | 25 |
| 2 | 230 | 6 | 203 | 2 | 4 | 16 |
| 3 | 1000 | 1 | 180 | 4 | 3 | 9 |
| 4 | 78 | 10 | 226 | 1 | 9 | 81 |
| 5 | 194 | 7 | 162 | 7 | 0 | 0 |
| 6 | 182 | 8 | 173 | 6 | 2 | 4 |
| 7 | 675 | 3 | 156 | 8 | 5 | 25 |
| 8 | 345 | 4 | 146 | 9 | 5 | 25 |
| 9 | 982 | 2 | 186 | 3 | 1 | 1 |
| 10 | 112 | 9 | 178 | 5 | 4 | 16 |

**

rs = -0.22

We have a negative correlation coefficient, suggesting the possibility of a negative correlation between synthesis time and half-life. However, is this significant or just due to chance? We check the critical value table (ignoring the negative sign in our result)

The critical value for the Spearman’s rank correlation coefficient at p= 0.05 where n is 10 is 0.6485

The calculated rs value ( 0.22, ignoring the negative) is less than the critical value so there is no significant correlation

We have failed to reject the null hypothesis that there is no relationship between synthesis time and half-life of RNA molecules.

**Questions M1.10**

1. Here is a dataset of the cell sizes of a sample from a population of the single-celled eukaryote *Paramecium bursaria*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Size (um) | 80 | 150 | 95 | 110 | 210 | 140 | 97 | 101 | 85 | 134 |

1. State the interval that covers one standard deviation above and one standard deviation below the mean.

80.7-159.7 um

1. Any scores which lie more than three standard deviations above or below the mean could be considered extreme scores, are there any scores which fit this criterion?

No

**Questions M1.11**

1. A 10 µl pipette is guaranteed by its manufacturer as accurate to ±0.03 µl, this was tested in the lab by drawing up 10 µl of water repeatedly and verifying the volumes independently. The volumes were verified as follows:

9.9978; 10.0062; 10.0020; 10.0047; 9.9998; 9.9982; 10.0023; 10.0034; 10.0012; 9.9994

Is this pipette accurate to within the manufacturer’s specifications?

Values larger than 10.0300 or smaller than 9.9700 would contradict the specification.

There are no such values in this sample.

1. 2.5 ml of a liquid bacterial culture is transferred into a larger volume of 9 ml of fresh broth. The pipette used to transfer the culture is accurate to ±0.03 ml, and the measuring cylinder for the broth is accurate to ±0.05 ml.
2. What is the new volume and the absolute uncertainty?

11.5ml ±0.08

1. What is the relative uncertainty of the pipette measurement?

1.2%

**OCR Resources**: *the small print*OCR’s resources are provided to support the delivery of OCR qualifications, but in no way constitute an endorsed teaching method that is required by the Board, and the decision to use them lies with the individual teacher. Whilst every effort is made to ensure the accuracy of the content, OCR cannot be held responsible for any errors or omissions within these resources.
© OCR 2017 - This resource may be freely copied and distributed, as long as the OCR logo and this message remain intact and OCR is acknowledged as the originator of this work.

OCR acknowledges the use of the following content: M1.8:: B0004116 Credit Jean Wade and Linda Sharp, Wellcome Images, Copyrighted work available under Creative Commons Attribution only licence CC BY 4.0 http://creativecommons.org/licenses/by/4.0/

Please get in touch if you want to discuss the accessibility of resources we offer to support delivery of our qualifications: resources.feedback@ocr.org.uk

We’d like to know your view on the resources we produce. By clicking on ‘Like’ or ‘Dislike’ you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click ‘Send’. Thank you.

If you do not currently offer this OCR qualification but would like to do so, please complete the Expression of Interest Form which can be found here: [www.ocr.org.uk/expression-of-interest](http://www.ocr.org.uk/expression-of-interest)

Looking for a resource? There is now a quick and easy search tool to help find free resources for your qualification:
[www.ocr.org.uk/i-want-to/find-resources/](http://www.ocr.org.uk/i-want-to/find-resources/)

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