Statistics in A Level Biology

Statistics is an aspect of Biology which can be intimidating for students who find maths challenging. This need not be the case if you concentrate on what you need to know, rather than allowing yourself to get bogged down in mathematical detail. What follows is, therefore, an attempt to help A Level Biologists become more comfortable with what they need to know about this subject.

On this page you can find useful hints on:

- Why we use statistics

- How we measure probability

- Which stats tests to use – when and why

- How to interpret your stats results

Why do we use statistics?

When we carry out a biological investigation, we will often end up with a lot of figures which we need to be able to interpret.

 For example, we might have measured the heights of a sample of crop plants from two fields which had received different doses of fertilizer, to see if the fertilizer increases the growth of the plant. We might well find that the average height from one field is greater than the other, but there is likely to be some overlap in individual plants from across the two fields. Very rarely will we find that every measurement from one sample is greater than every measurement from the other. There are two possible explanations for the difference in average height that we have found:

1.   There may be a real difference between the two crops.

2.   The crops may in fact be the same size, and the difference we have found is due to chance since we have only taken a sample from each field – we just so happened to only measure the tallest crops in one field and the smallest crops in the other.

We need to be able to distinguish between these two possibilities, and this is where statistical tests come to our aid.

Statistical tests are used to tell us how likely it is that our results are due to chance.

Stats tests enable us to calculate the **probability** that our results are real or due to chance.

How do I measure probability?

Probability is normally measured on a scale from 0 – 1, where 0 represents impossibility and 1 represents certainty. The probability that our results are due to chance could take any value between 0 and 1. We need to agree on what level of probability we are going to accept before we decide that the difference is real. For most purposes, and certainly for A Level Biology, the level which is conventionally chosen is that the probability that our result is due to chance should be no more than 0.05. This means that there is only a 1 in 20, or 5% probability that the difference we have seen is due to chance

We can be 95% confident that our results are real – there is a high chance that something really is   going on!

At this point, we say that the difference is **statistically significant**. We have not proved that one crop is taller than the other, but we will now proceed on the assumption that this is the case.

How do I know which test to use?

Different statistical tests are used in different circumstances. There are a few steps to go through to tell you what test you should be using.

- 1. First you need to decide whether you are looking for **differences** or **associations** between sets of data. In the crop measuring example above you want to know if the average height of the plants is different in the two samples. If you had measured several samples at intervals, for example from the top of a slope to the bottom, you would want to know whether there is an association between the height of the plants and the distance down the slope.

2. Now you have decided what you are testing your data for, here's what to do for each situation.

a) Testing for *Differences:*

The test you will need here depends on your exam board.

For AQA you will need the **Standard Error**.

For Edexcel and OCR there are two possibilities, so you have another decision to make:

You need to decide whether your data fit a **normal distribution**. This is a symmetrical distribution, with the greatest number of readings being in the central range, and progressively fewer readings as you move away from the mean in either direction. You can find the normality of the distribution of your data by plotting it as a histogram. When in this form, normal distribution takes the form of a bell shaped curve.

If your data roughly fits a bell shape curve and is normally distributed, you need the **T test.**

If not, try the **Mann Whitney U test**.

In practice, if you have measured any aspect of an organism you are likely to find a normal distribution. If you have counted numbers of organisms it is unlikely that you will have a normal distribution.

b) Testing for *Associations*:

The test you will need here depends on the type of data you have measured.

3. You need to decide whether your data consist of either:

Continuous variables (measurable on a scale) e.g. distance, height OR **Discrete categories** e.g. colour, gender.

Spearman Rank should be used when you are looking for associations between two continuous variables.

The Chi Squared association test is the one to choose if looking for associations between categorical data.

(Be careful with this last test. You may come across a slightly different version of this test in the context of genetics. The chi squared formula is the same, but there is a particular way to calculate the expected figures).

How do I interpret the result from my stats test?

Once you have chosen the correct test, and put your figures into the appropriate formula, you will arrive at a figure known as the **calculated test statistic**. This is not the probability figure referred to earlier, and on its own it means nothing. It needs to be compared with a **critical value** which varies according to your sample size and the level of probability you are demanding. If you are doing the test manually, you will obtain this critical value from a table in a statistics book. In an exam, you would be provided with an extract of such a table, from which you might be expected to extract the relevant critical value. If using a computer programme, you will automatically be given the critical value.

(There is no need to worry about how these critical values have been calculated. For our purposes, we can trust the clever statisticians who have worked these out for us).

You then need to compare your calculated test statistic with the critical value. In most cases the calculated test statistic needs to exceed the critical value before we can say that our result is significant:

Calculated value > Critical value = Significant result

 An exception to this rule however is when using the Mann Whitney U test where your calculated value needs to be less than the critical value before your result is significant:

  Calculated value < Critical value = Significant result

Instructions on how to interpret the result from each test can be found on the particular statistical test pages in the rest of the Statistics section. However, below is a brief summary on what is good practice when writing up statistical analysis:

When writing up statistical analysis for coursework or an exam question, there are a few things to remember to include in your answer:

- Which test you used and why (using the circumstances in the flow diagram that helped your decision in the first place is a good way of doing this).

- A statement of your null hypothesis

- Whether your calculated value was > or < the critical value

- The confidence level that you have used to decide at what point to accept that the difference is real (always at a 95% confidence level for A level Biology)

- Whether you have therefore accepted or rejected your null hypothesis