

An Introduction to Uncertainty and the Use of Statistics in Biology

Defining Uncertainty in Science

The term uncertainty is not a well-defined scientific term. In fact, it may mean different things to different types of scientists. To the engineer, uncertainty could refer to the resolution, accuracy or precision of a device. To the chemist or physicist, it could refer to the accuracy or precision of an instrument or experiment. To the biologist, the term uncertainty usually refers to the uncertainty in an experiment and/or the validity of an experiment as judged by the statistical significance of the results.

Experimental uncertainty is influenced by multiple factors, including elements having little to do with the sensor being used, such as experimental design. This is why most scientists report the mean and standard deviation of the results of their experiment, not the accuracy of the device that they use to do an experiment. In biology, we are dealing with complex processes that cannot be reduced to simple equations. For this reason, experimental uncertainty is much more important than sensor uncertainty.

In biology, experimental uncertainty can be determined by repeating an experiment several times and then computing a few descriptive statistics, such as the mean and standard deviation of the results. Statistical tests can then be used to determine if the results are significantly different from each other. In science, for something to be significant, we have to be 95% sure that what we are seeing is not due to random chance.

Minimizing Uncertainty in the Classroom

In a classroom setting, the class average and standard deviation for an experiment can serve as an estimate of experimental uncertainty, but it is not ideal. This is because the greatest amount of experimental uncertainty can come from the students performing the experiment. We cannot assume that every student is performing an experiment in the exact same way. The ideal way to minimize experimental uncertainty in the classroom is to have each student or group of students repeat their experiment several times using the same instrument or sensor. For example, a biologist will repeat the same experiment several times using the same instrument following a very specific protocol. In addition, most biologists will also calibrate or verify the accuracy and precision of their instrument before performing an experiment. This is all done to reduce experimental uncertainty.

A Brief Introduction to Statistics

Sample Size: The sample size refers to the number of measurements that you have taken. For example, if you measured the heart rate of a dozen people your sample size would be twelve. If you measured the respiration rate of a sample of yeast and then repeated this experiment ten times your sample size would be ten. The letter n is used to represent sample size.

Mean: The mean of a sample is another way of saying the average. To calculate the arithmetic mean of a data set, add the individual samples (e.g., measurements) and then divide by the sample size. The symbol \bar{X} is used to represent the mean.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

Standard Deviation: The standard deviation tells us how far away a measurement is from the mean or average. If the data set is normally distributed (e.g., resembles a bell-shaped curve) then approximately 68% of the samples will fall within one standard deviation of the mean. In addition, almost all of the samples (~95%) will fall within two standard deviations of the mean. To calculate the standard deviation of a data set, take the mean and subtract it from each value (e.g., sample, measurement). Next, square each result and then sum them all. Then, find the mean of these values. This is the variance of your data set. To find the standard deviation, take the square root of the variance. The standard deviation has the same units as the sample or measurement. The letters *sd* are used to denote standard deviation.

$$sd = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2}$$

Standard Error of the Mean: In biology, error bars on graphs such as bar charts are typically expressed as the standard error of the mean. The standard error of the mean is an estimate of how close the mean of the data set is to the mean of the entire population being studied. For example, if you determined the mean resting heart rate of 12 women in the state of Oregon, it would be useful to have an estimate of how close this mean was to the mean resting heart rate of all women in the state of Oregon. One would assume that the larger the sample size the closer your mean would be to the real mean of the population. The standard error of the mean incorporates sample size, and is calculated by taking the standard deviation of the mean and dividing it by the square root of the sample size. The letters *se* are used to denote standard error of the mean. See the formula below for an expression on how to calculate the standard error of the mean.

Calculate standard error of the mean: $se = \frac{sd}{\sqrt{n}}$

In science, there is a standard format that is used to report descriptive statistics. The mean is reported followed by the standard deviation or standard error of the mean that carries both negative and positive signs. The sample size is usually reported as well. For example, if the average heart rate of 12 women was 64 beats per minute and the standard deviation was 8 beats per minute, we would report that the mean heart rate was 64 ± 8 bpm, $n=12$. In biology, we often want to determine if there is a significant difference between two data sets or experiments. For example, we might want to determine if men and women have different resting heart rates. As a general rule, if the standard error of two means overlap then the results are probably not significantly different from one another. However, statistical tests should be performed to determine if this is the case.

A Brief Introduction to Statistical Tests

The following primer introduces you and your students to two common statistical tests used by biologists. The formulas and tables in this appendix are designed so your students can perform these tests by hand. Each test is designed to test two sample means. You should not use either of these tests to compare three or more means. These tests should not be used to compare proportions or data based on a ranking scale (e.g., Likert scale). As a general rule for classes, each mean should be based on at least five independent samples or measurements. For example, if you wanted to see if there was a significant difference between the heart rates of men and women you would need to sample at least five men and five women. For college courses, we strongly encourage a minimum sample size of ten for each mean. Many graphing calculators, spreadsheet programs, and online resources can be used to perform these tests.

The Unpaired Student's *t*-test

If you want to compare the means from two sets of data (e.g., the mean heart rate of a group of men and a group of women) we can use Student's *t*-test to determine if the means are significantly different from each other. In science we define significance as being 95% sure that what we are seeing is not due to random chance. This is just a fancy way of saying that we only consider a result significant if there is a 5% chance or less of these two samples NOT being different from each other. The *t*-test compares the means of two independent groups or samples. The formula below can be used to calculate the *t*-statistic for this case.

Calculate the *t*-statistic using

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(sp^2 / n_1) + (sp^2 / n_2)}}$$

where \bar{X}_1 and \bar{X}_2 are the means of the first and second data set, respectively, and n_1 and n_2 refer to the sample size of data set 1 and 2, respectively. The letters sp^2 refer to the pooled variance. The square root of sp^2 is the pooled standard deviation. The pooled variance can be calculated using the formula below.

Calculate pooled standard variance using

$$sp^2 = \frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{df}$$

where sd_1 and sd_2 are the standard deviation for data set 1 and 2, respectively, and *df* represents the *degrees of freedom* of the experiment. We can calculate the degrees of freedom by summing the sample size of both data sets and then subtracting 2. An in-depth discussion of degrees of freedom is beyond the scope of this primer.

Calculate degrees of freedom.

$$df = n_1 + n_2 - 2$$

Once you have performed these calculations you can determine if there is a significant difference between the two samples using a *t*-table. A *t*-table is located at the end of this appendix for this purpose (see Table 1).

Using the *t*-table, find the row that corresponds to the degrees of freedom for your experiment. Look across the table and find where your *t*-statistic fits on the table. Use the absolute value of your *t*-statistic. If your *t*-statistic falls between two columns, choose the column that is less than or equal to the *t*-statistic that you calculated.

Each column has a p value associated with it. The p value refers to the probability that the difference observed is due to random chance. Remember that for a result to be significant we need to be 95% certain that what we are seeing is NOT due to random chance. For example, if the t -statistic you calculated falls in the column labeled 0.05, then the results are deemed significant as there is only a 5% chance that difference observed is due to random chance.

In science, there is a standard format for reporting p values. If your t -statistic was found between the column for 0.05 and 0.025 you would report your p value as $p \leq 0.05$. If your t -statistic was found between the column for 0.01 and 0.005 you would report your p value as $p \leq 0.01$. If your t -statistic is found in any column to the left of the 0.05 column, the results are not considered significant and the p value can be reported as $p \geq 0.05$.

The Paired Student's t -test

If you have before and after data you can use a special form of Student's t -test to compare two sample means. This test is called the paired Student's t -test. Calculating the t -statistic for a paired t -test is much simpler than calculating a t -statistic for an unpaired t -test.

Calculate the t -statistic for a paired t -test.

$$t = \frac{\overline{X}_D}{sd_d / \sqrt{n}}$$

\overline{X}_D represents the mean difference between the before and after data and sd_d is the standard deviation of the mean difference, as described in more detail hereafter. The letter n in this case simply refers to the sample size, which is the same for each sample since we have before and after data.

To calculate \overline{X}_D and sd_d , simply subtract each "before value" from its paired "after value." You should then have a column of differences. Compute the mean and standard deviation for this column of differences to get \overline{X}_D and sd_d . Then compute the t -statistic using the formula above.

You will also need to calculate the *degrees of freedom* (df) for your experiment. In this case, we are comparing paired data. We can calculate the degrees of freedom by subtracting 1 from the sample size. An in-depth discussion of degrees of freedom is beyond the scope of this primer.

Calculate degrees of freedom for a paired t -test.

$$df = n - 1$$

Once you have calculated these parameters, you can determine if there is a significant difference between your before and after data using a t -table. A t -table is located at the end of this appendix for this purpose (see Table 1).

Using the t -table, find the row that corresponds to the degrees of freedom for your experiment. Look across the table and find where your t -statistic fits on the table. Use the absolute value of your t -statistic. If your t -statistic falls between two columns, choose the column that is less than or equal to the t -statistic that you calculated.

Each column has a p value associated with it. The p value refers to the probability that the difference observed is due to random chance. Remember that for a result to be significant we need to be 95% certain that what we are seeing is NOT due to random chance. For example, if

the t -statistic you calculated falls in the column labeled 0.05 then the results are deemed significant as there is only a 5% chance that difference observed is due to random chance.

In science, there is a standard format for reporting p values. If your t -statistic was found between the column for 0.05 and 0.025 you would report your p value as $p \leq 0.05$. If your t -statistic was found between the column for 0.01 and 0.005 you would report your p value as $p \leq 0.01$. If your t -statistic is found in any column to the left of the 0.05 column, the results are not considered significant and the p value can be reported as $p \geq 0.05$.

Choosing the Right Statistical Test for Your Experiment

Most biologists will design their experiment around the statistical test that they plan to use. The paired t -test is more robust for small sample sizes, but you must use paired data. This means that the same samples or subjects must be tested under two different conditions. If your experiment uses different samples or subjects, you must use the unpaired t -test. Please see the examples below.

Example 1: A student decides to compare the respiration rate of germinating and non-germinating peas. They measure the respiration rate of ten germinating peas using a Nalgene bottle and a CO₂ Gas Sensor. They repeat this experiment six times using different sets of germinating peas each time. Then, they measure the respiration rate of ten non-germinating peas. They repeat this experiment using different sets of non-germinating peas six times. Which statistical test should they use to compare their results?

Answer 1: The student has collected different samples. They do not have before and after data for the same subjects or samples. They must use the unpaired t -test.

Example 2: A student decides to compare the respiration rate of germinating and non-germinating peas. They measure the respiration rate of ten non-germinating peas using a Nalgene bottle and a CO₂ Gas Sensor. They repeat this experiment six times using different sets of non-germinating peas. Then, they put each set of non-germinating peas in water and allow them to sit overnight so they begin to germinate. They repeat the experiment using the same sets of peas the next day. Which statistical test should they use to compare their results?

Answer 2: The student has collected paired data. They have before and after data for the same sets of peas. They can use the paired t -test.

Example 3: A student decides to compare the respiration rate of germinating and non-germinating peas. They measure the respiration rate of ten germinating peas using a biochamber and a CO₂ Gas Sensor. They repeat this experiment six times using the same ten germinating peas each time. Then, they repeat this experiment six times using a different set of ten non-germinating peas. Which statistical test should they use to compare their results?

Answer 3: The student has measured the same ten peas repeatedly resulting in pseudo-replication. The samples are not independent. The student should not use statistics to compare the results of this experiment.

Examples from Experiments in the Classroom

Comparing Mean Cellular Respiration of Germinating and Non-Germinating Peas

A class performed an experiment very similar to Example 1 above. Each group measured the respiration rate of a set of germinating and non-germinating peas. There were five groups of students performing the experiment. The respiration rate was divided by the mass of the peas used in each case.

The mean respiration rate for germinating peas was found to be 0.258 ppm CO₂/s/g with a standard deviation of ±0.0283 ppm CO₂/s/g. The mean respiration rate for non-germinating peas was found to be 0.0192 ppm CO₂/s/g with standard deviation of ±0.008 ppm CO₂/s/g. The standard error of the mean for each data set was calculated by dividing the standard deviation by the square root of 5 ($n=5$ for each data set). The resulting *se* for germinating and non-germinating peas was found to be ±0.0126 and ±0.0036 ppm CO₂/s/g respectively. The data is not paired so we have to use the unpaired *t*-test to compare the two means.

Before we can compute the *t*-statistic, we need to calculate the degrees of freedom. In this case, the degrees of freedom is the sample size of each data set minus 2, which is 5+5–2 or 8. We can now calculate the pooled variance as shown below.

$$sp^2 = \frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{df} = \frac{(5 - 1)(0.0283)^2 + (5 - 1)(0.008)^2}{8} = 0.000432$$

Now we can calculate the *t*-statistic as shown below.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(sp^2/n_1) + (sp^2/n_2)}} = \frac{0.258 - 0.0192}{\sqrt{(0.000432/5) + (0.000432/5)}} = \frac{0.2388}{0.0131} = 18.23$$

Using the *t*-table, we now find the row that corresponds to the degrees of freedom for the experiment which in this case is 8. Looking across the table we find where the *t*-statistic fits on the table. The *t*-statistic is greater than the last value located in the row for 8 degrees of freedom. This means that our *p*-value is less than 0.01. This *p*-value indicates that there is less than a 1% chance that difference in means observed was due to random chance. In this case $p \leq 0.05$, so the results are significantly different from each other.

Comparing Mean Cellular Respiration of Germinating Peas at Two Temperatures

During the next period, the class performed an experiment very similar to Example 2 above. Each group measured the respiration rate of a set of germinating peas at two temperatures. There were five groups of students performing the experiment. The resulting respiration rate in each condition was divided by the mass of the peas being tested.

The mean respiration rate for germinating peas at room temperature (22°C) was found to be 0.258 ppm CO₂/s/g with a standard deviation of ±0.283 ppm CO₂/s/g. The mean respiration rate for germinating peas at a colder temperature (12°C) was found to be 0.152 ppm CO₂/s/g with standard deviation of ±0.0239 ppm CO₂/s/g. The standard error of the mean for each data set can be calculated by dividing the standard deviation by the square root of 5 since $n=5$ for each data set. The resulting *se* for germinating peas at 22°C and 12°C was found to be ±0.0126 ppm CO₂/s/g and ±0.0106 ppm CO₂/s/g respectively. The data is paired, so we can use the paired *t*-test to compare the two means.

Before we can compute the paired t -statistic, we need to calculate the degrees of freedom, the mean difference, and standard deviation of the mean difference. In this case, the degrees of freedom is 5 (the sample size) minus 1, which is 4. The mean difference (\bar{X}_D) was found to be 0.106 ppm CO₂/s/g. The standard deviation of the mean difference (sd_d) was found to be ± 0.0512 . We can now calculate the paired t -statistic as shown below.

$$t = \frac{\bar{X}_D}{sd_d / \sqrt{n}} = \frac{0.106}{0.0512 / \sqrt{5}} = 4.63$$

Using the t -table, we now find the row that corresponds to the degrees of freedom for the experiment, which is 4. Looking across the table we find where the t -statistic fits on the table. The t -statistic is greater than the last value located in the row for 8 degrees of freedom. This means that our p -value is less than 0.01. This p -value indicates that there is less than a 1% chance that difference in means observed was due to random chance. In this case $p \leq 0.05$, so the results are significantly different from each other.

Comparing Respiration Rates of Yeast to Glucose and Fructose

A group of students performed Experiment 10, which looks at the metabolism of sugars by yeast using a CO₂ Gas Sensor. The students measured the respiration rate of baker's yeast after exposure to 0.3M glucose. They repeated this experiment eight times using a different sample of yeast for each run. Then, they measured the respiration rate of baker's yeast after exposure to 0.3M fructose. They repeated this experiment six times using a different sample of yeast for each run. The mean respiration rate for yeast exposed to glucose was found to be 498 ± 34.78 ppm CO₂/min, $n=8$. The mean respiration rate for yeast exposed to fructose was found to be 524 ± 50.49 ppm CO₂/s/g ± 50.49 ppm, $n=6$. The standard error of the mean for each data set can be calculated by dividing the standard deviation by the square root of the number of runs for each data set, which is 8 for glucose and 6 for fructose. The resulting se for yeast exposed to glucose and fructose was found to be ± 12.33 and ± 20.62 ppm CO₂/min respectively. The data is not paired so we have to use the unpaired t -test to compare the two means.

Before we can compute the t -statistic we need to calculate the degrees of freedom. In this case, the degrees of freedom is the sample size of each data set minus 2, which is $8+6-2$ or 12. We can now calculate the pooled variance as shown below.

$$sp^2 = \frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{df} = \frac{(8 - 1)(34.77)^2 + (6 - 1)(50.49)^2}{12} = 1767.41$$

Now we can calculate the t -statistic as shown below.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(sp^2 / n_1) + (sp^2 / n_2)}} = \frac{498 - 524}{\sqrt{(1767.41 / 8) + (1767.41 / 6)}} = \frac{-26}{22.7} = -1.14$$

Using the t -table, we find the row that corresponds to the degrees of freedom for the experiment, which is 12. Using the absolute value of the t -statistic (1.24) we look across the row until we find where the t -statistic fits on the table. The t -statistic is smaller than the first value located in the row for 12 degrees of freedom. This means that our p -value is greater than 0.2. This p -value

indicates that there is more than a 20% chance that difference in means observed was due to random chance. In this case $p \geq 0.05$, so the results are not significantly different from each other.

Using Logger Pro with this Book

Uncertainty in Linear and Other Curve Fits

After you perform a linear fit in Logger Pro a helper object containing the slope, y-intercept, and correlation coefficient of the regression line, will be added to the graph. The proper name for the correlation coefficient is the Pearson product-moment correlation coefficient, which is commonly referred to as the r value. The correlation coefficient is a decent measure of uncertainty, but it is not a measure of statistical significance. The correlation coefficient is a number between -1 and 1 which measures to what extent two variables are linearly related. A perfect correlation with a positive slope is shown by a $+1$ correlation, while a negative slope is shown by a -1 correlation. A correlation of 0 means that there is no linear relationship.

After performing a curve fit in Logger Pro, a helper object containing the parameters of the fit including the Root Mean Square Error will be added to the graph. The Root Mean Square Error (RMSE) is a good measure of uncertainty. The Root Mean Square Error is a measure of how far away, on average, the data points are from the fitted curve. RMSE is in the units of the y-axis.

A tutorial covering curve fitting in general is available in Logger Pro. To access it, choose **Open** from the **File** menu. Open the **Experiments** folder, then choose the **Tutorials** folder.

Creating Bar Graphs

1. To create a bar graph, double-click the graph of interest and select Bar Graph from the Graph Options window.
2. If you want to enter data into Logger Pro to make a bar graph,
 - a. Double-click the X column heading and name it appropriately. This column should represent an independent variable that you manipulated (e.g., pH, temperature, salinity). It could also represent a category (e.g., glucose, sucrose, fructose). This column needs to contain the value that you measured for each condition or category. Select OK.
 - b. Double-click the Y column heading and give it an appropriate name based on the data for your experiment (e.g., respiration rate, reaction rate, abs/min). If you are entering categorical data select Treat As Text for the Data Type. Select OK.
 - c. Enter your data in the data table. As you enter values, your data should be displayed on the graph.
 - d. When you are done entering data, double-click the graph and select Bar Graph in the Graph Options dialog.

Reporting Descriptive Statistics

1. To calculate descriptive statistics, disconnect any sensors or interfaces from your computer and start Logger Pro. A table called Data Set with X and Y columns will be displayed with a blank graph.
2. Double-click the X column heading and name it “sample” or “experiment number.” Select OK.

3. Double-click the Y column heading and give it an appropriate name based on the data for your experiment (e.g., respiration rate, reaction rate, abs/min). Select OK.
4. Enter the appropriate x, y pairs into the data table. This means that you need to enter the experiment number in the x-column followed by the appropriate value that you received for each experiment in the y-column.
5. Click the Statistics button. The mean and standard deviation of the Y values will be displayed.

Displaying Error Bars

1. Disconnect any sensors or interfaces from your computer and start *Logger Pro*. A table called Data Set with X and Y columns will be displayed with a blank graph.
2. Double-click the X column heading and name it appropriately. This column should represent an independent variable that you manipulated (e.g., pH, temperature, salinity). It could also represent a category (e.g., glucose, sucrose, fructose). Select OK.
3. Double-click the Y column heading and give it an appropriate name. This column should contain the *mean* value that you found for each condition or category. Select OK.
4. Enter the data from your experiment. You now have a set of X, Y pairs.
5. Choose New Manual Column from the Data menu. Name this column **se** (short for Standard Error of the Mean) or **sd** (short for Standard Deviation), and then enter the value that you calculated for each condition or category. Select OK.
6. Double-click the y-column heading. On the Options tab, select Error Bar Calculations, and then select Use Column. Select the column that you created for standard error from the drop down list. Select Done and then a graph should be displayed that shows your data with error bars.

Moving a Linear Fit to Make Use of Error Bars

1. Click and drag across a region of interest. The shaded area contains the selected data upon which the movable linear fit will be calculated. **Note:** If a region is not selected, the movable linear fit will be calculated for all data displayed in the graph.
2. Choose Model from the Analyze menu and select Linear as the equation type. A helper object containing the slope, y-intercept, and RMSE of the line will be displayed on the graph.

3. Double-click the helper object and select the option for Enable Line Drag. The movable linear fit can now be adjusted by dragging one of the three handles on the line. The center handle adjusts the y-intercept, the outer two handles can be used to rotate the line about the far edge. If the x-axis is scaled from zero, then the right handle will control the slope while leaving the y-intercept alone.
4. To remove the movable linear fit, click the X in the upper corner of the helper object.

Table 1: <i>t</i> -Distribution (absolute value)					
	Probability (P value)				
Degrees of Freedom	0.200	0.100	0.050	0.025	0.010
1	3.078	6.314	12.707	25.452	63.657
2	1.886	2.920	4.303	6.206	9.925
3	1.638	2.354	3.183	4.177	5.841
4	1.534	2.132	2.777	3.496	4.605
5	1.476	2.016	2.571	3.164	4.033
6	1.440	1.944	2.447	2.969	3.708
7	1.415	1.895	2.365	2.842	3.500
8	1.397	1.860	2.307	2.752	3.356
9	1.384	1.834	2.263	2.686	3.250
10	1.373	1.813	2.229	2.634	3.170
11	1.364	1.796	2.201	2.594	3.106
12	1.357	1.783	2.179	2.561	3.055
13	1.351	1.771	2.161	2.533	3.013
14	1.346	1.762	2.145	2.510	2.977
15	1.341	1.754	2.132	2.490	2.947
16	1.337	1.746	2.120	2.473	2.921
17	1.334	1.740	2.110	2.459	2.899
18	1.331	1.735	2.101	2.446	2.879
19	1.328	1.730	2.094	2.434	2.861
20	1.326	1.725	2.086	2.424	2.846

Table created using the function "T.DIST.2T(x, deg_freedom)" in Microsoft® Excel® 2010. This function returns the two-tailed Student's *t*-distribution. Excel's lookup functions were then used to find values of interest within the table, namely the *t*-values corresponding to the selected probabilities.