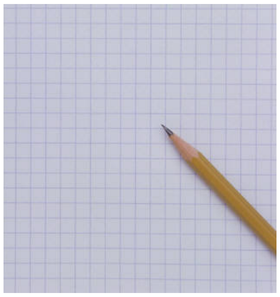
USING MATH:

CALCULATIONS, GRAPHS, & STATISTICS



CONTENTS

Page

Using Math to Think About Plants

Sometimes the best way to describe and understand plants – or anything else in the biological world – is to use mathematics. In this section, you will find a variety of ways that math can help you set up and analyze results from your investigations.

Counts:

Sometimes you can collect data without using any measuring tools. What form would your data take if you were collecting information on how many butterflies visit a plant? Is it different from the type of data for how many hairs are found on a leaf or tree species found in a plot?

Many possible kinds of data can be collected by **counts**. Counts are exactly what they sound like – *the total number of some kind of object in a sample*. For example, you may be interested in the influences on plant growth and development. For an experiment related to this topic, you might take a count of the number of leaves on a plant of a given age is a count. This is because the number of leaves plants produce at a given age can be affected by their genetics or local environment. Like other measurements, you can then summarize leaf counts from multiple plants within a treatment as an *average* leaf count and calculate how *variable* leaf counts are across the plants you sampled (see *Descriptive Statistics*).

Counting the number of seeds in a single pea pod is pretty straightforward. However, counting the total number of seeds a *Brassica* plant produces at the end of its life cycle is more difficult. When counts will involve numbers greater than about 25 for a single sample, you can use a **hand counter** can help keep track of the total. Simply press the counter button to record each item in the sample; when you are done counting, a meter will show the sample's final count value.

Total counts are not always practical or necessary. Instead, you can carry out a count on a pre-selected portion of the sample area or volume. This is called **subsampling**. For example, it might be physically possible to count the total number mature red oak trees in a forest or the number of stomata on a single red oak leaf, but this could take quite a long time! Instead, you might want to count the number of mature red oak trees in one hectare or the number of stomata in one square centimeter of a leaf. To estimate the total count for a full sample, you can then multiply the subsample count by the number of subsamples needed to make up the full sample. The ***Power of Sunlight*** and the ***Celery Challenge*** both include experiments where you might want determine the total number of stomata on a leaf. You could do this as follows:

$$\text{Total stomata on a leaf} = \text{Count of the stomata in } 1\ \text{cm}^2 \times \text{Area of leaf in cm}^2$$

See *Sizes and Shapes* for ideas on how to figure out the area of a leaf.

Proportions, Ratios, and Percentages:

In many cases, division can be helpful in thinking about your data. For example, you may have collected data about the insects that visit sunflowers. You may have seen 12 bees visit Sunflower A and 14 bees visit Sunflower B. Are bees the only insects that visited the sunflowers? If not, you may not be seeing the big picture if you only present results about bees. Instead, you could point out that 12 out of 30 insects that visited Sunflower A and 14 out of 22 insects that visited Sunflower B were bees. Can this data be simplified further?

One way that you could present such data is as a **proportion**, in which the number of bees is divided by the total number of insects. In this case, Sunflower A had a proportion of bees of 12 divided by 30, or 0.40. Meanwhile, Sunflower B had a proportion of bees of 14 divided by 22, or about 0.64. Not only did Sunflower B have more total bees visit it, its proportion of bees was also much higher than for Sunflower A. This is partly because fewer total insects visited Sunflower B.

Another way you could present the same kind of data is as a **ratio**. In a ratio, the numbers are presented as a reduced fraction or as a reduced pair of numbers separated by a colon. Sunflower A had a ratio of bees to total insects of 12/30, which can be reduced to 2/5 or 2:5. You could also say that Sunflower A had a ratio of bees to non-bee insects of 12/18, which can be reduced to 2/3 or 2:3.

> ***Thought Exercise:*** *What is the ratio of bees to total insects that visited Sunflower B? What is this sunflower's ratio of bees to non-bees? Remember to put the final ratios into a reduced form!*

A third way that division can be helpful in thinking about your data is in the use of **percentages**. Suppose you have a packet of thirty bean seeds that you will be studying for ***The Wonder of Seeds***. Five of the seeds have brown seed coats, while the rest are tan with purple speckles. The percentage of seeds with brown seed coats is calculated as:

$$\% \text{ brown seeds} = (5/30) \times 100\% = 0.1667 \times 100\% = 16.67\%$$

As you can see, calculating a percentage is very similar to calculating a proportion or ratio. The main difference is that you multiply the result by 100% after carrying out the division. Therefore, you can think of a percentage as being a specific type of ratio describing the number of units present in a total of one hundred units. What the units are in each case is not fixed. In the bean example, only thirty total beans are present, so a "unit" is clearly smaller than a single bean! The choice of one hundred units is simply to provide a good basis for comparison across a range of different sample sizes.

> ***Thought Exercise:*** *If you present your data as a proportion, ratio, or percentage, will your audience know how many items were in your total sample? If so, how? If not, how could you let them know?*

Proportions and percentages are often used in making or describing chemical solutions. Suppose you want to make a bleach solution to sterilize the surface of some seeds before germinating them. A 1:9 bleach:water solution is often used for this. How would you make the solution? You would mix 1 part bleach with 9 parts water. If you decide that one part is 5 mL, you would mix 5 mL bleach and 9 x 5 mL = 45 mL water, for a total of 50 mL of 1:9 bleach:water solution. This can also be described as a 10% bleach solution, because it has 5 mL of bleach out of a total of 50 mL:

5 mL bleach/50 mL solution x 100% = 0.1 x 100% = 10% bleach solution

If you will be working with chemicals, ***Research in the Lab*** can help you understand how to make and describe solutions in more detail.

Rates:

In many biology experiments, time is an important factor. Would you be surprised to learn that a tomato plant grew to be one meter tall? Would it be more surprising if the tomato plant had grown this much in twelve weeks or in two weeks?

Whenever you take measurements to figure out how fast a process happens, you will include time as part of the measurement. For instance, the amount of carbon dioxide a plant consumes in an hour can tell us about photosynthesis in ***The Power of Sunlight***. How quickly a celery stalk takes up water can help us measure transpiration in ***The Celery Challenge***. Measurements that consider the amount of time required for a change to occur are called **rates**. The units describe the amount of time involved, from seconds to minutes, or hours to days. Some ecological rates are described on a per year basis!

The amount of new growth a plant produces per day is a common rate used in biology. To calculate plant growth rate we can use the equation:

Growth Rate = Change in Height/Change in Time

Because rates are *measures of change*, you will need to make the same measurement at least twice to calculate its associated rate. Here, the *change in height* involves subtracting an earlier plant height measurement from a later measurement of the same plant. The *change in time* is the total amount of time that passed between the two height measurements. For example, if you measured a plant that was 3 cm tall on Day 4 and 10 cm tall on Day 6, you would have:

Growth Rate = (10 cm – 3 cm)/(6 days – 4 days) = 7 cm/2 days = 3.5 cm/day.

The unit of time here is days. For any type of rate equation, a change in time will always be used. The other unit involved in this example is cm of plant height. Many possible types of measurements could be used instead of height to find different types of rates, as long as the same measurement type is used at both time points.

You might have noticed that the growth rate is only known for the overall time interval – you have no information about plant height on Day 5. The rate may have changed at one or more points over the two day interval, and it could also be very different outside that interval. For this reason, rate calculations are often referred to as *average rates* within the interval measured.

Suppose another member of your team measured the height of the same plant on Day 5 and Day 7, finding heights of 6 cm and 15 cm, respectively. You can now calculate three different daily growth rates:

Growth Rate$_{\text{day 4-5}}$ = (6 cm – 3 cm)/(5 d - 4 d) = 3 cm/d
Growth Rate$_{\text{day 5-6}}$ = (10 cm – 6 cm)/(6 d – 5 d) = 4 cm/d
Growth Rate$_{\text{day 6-7}}$ = (15 cm – 10 cm)/(7 d – 6 d) = 5 cm/d

Here, the growth rate was fastest from Day 6 to Day 7 and slowest from Day 4 to Day 5. The overall average growth rate from Day 4 to Day 7 was (15 cm – 3 cm)/(7 d – 4 d) = 12 cm/3 d = 4 cm/d. Notice that this is the same as the growth rate from Day 5 to Day 6, but greater than the growth rate from Day 4 to Day 6.

Thought Exercise: *Suppose that your team planted the seed for this plant on Day 0. What was the overall growth rate for the first week? How does it compare to the average growth rate from Day 4 to Day 7? The rate from Day 4 to Day 5?*

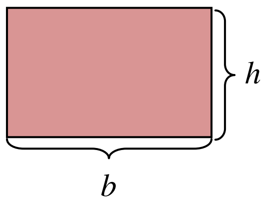
Sizes and Shapes:

Figuring out how large an object is can be an important part of planning and carrying out research. Does an oval leaf take up as much space as a triangular one of the same length? How big is the prairie preserve where you will be collecting seed? How much liquid will you need to later wash your seeds with 10% bleach?

Area is an important calculation in biology. It describes the size of an object in two dimensions, and it can be used both in setting up an experiment and in producing data. Field researchers must know the size of the area being investigated to calculate measurements like species density or population size. Very small sample areas are not helpful for estimating the numbers of large organisms present, and very large sample areas make it hard to count all of the small species within it. Therefore, planning how large an area to study is often a part of a field researcher's experimental plan.

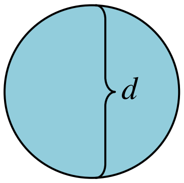
On a smaller scale, you may want to carry out counts on a subsample of a known area. For example, you might want to determine stomatal density, or the number of stomata per cm^2 of leaf surface, as you work to understand photosynthesis in ***The Power of Sunlight***. You could define a specific area in which you count the stomata, then divide the number of stomata by the total surface area to determine the stomatal density.

For the above examples, it is usually easiest to choose regular shapes for sampling. The areas of regular shapes can be determined using relatively simple formulas. You might find rectangles or squares useful in measuring out a field plot or defining the part of a leaf where you will count stomata. You can calculate the area of a rectangle from the length of its base (b) and its height (h):



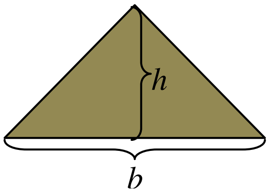
$$A_r = b \times h$$

You may also encounter circles in your work, such as when you are figuring out the area of the field of view in a microscope. Circular field plots can also be easier to lay out than perfectly square ones. You can place a stake at the center of where you would like the plot to be, then cut a string that is as long as half the desired plot diameter. Tie one end of the string to the stake, and pull the other end so that it is taut. By walking all the way around the stake with the taut string, you will trace the outer edge of a circular plot. The area of any circle, such as this field plot, is described in terms of its radius (r), which is half of its diameter (d), and the number pi (π), which is about equal to 3.14:



$$A_c = \pi \times r^2$$
$$\text{where } r = d/2$$

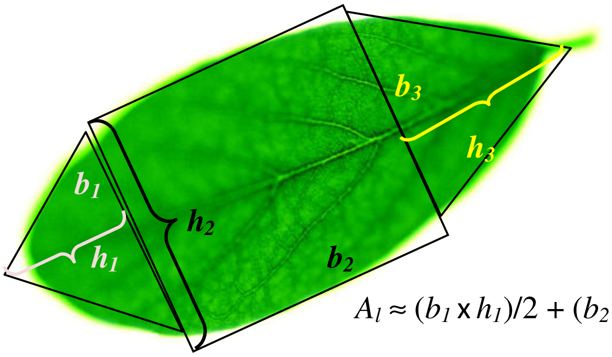
Finally, you may encounter triangular shapes in your work. This could happen when you are subsampling a larger, rectangular plot, or you may find, for example, that you can estimate the area of a plant's prickles because they have a triangular shape. The area of a triangle is also described in terms of its base and height:



$$A_t = (b \times h)/2$$

The area of an organism's body structures can provide important clues to their functions. In the previous example, the size of the plant's prickles may tell you something about the size of the herbivores the plant is able to fend off. As a second example, the area of a leaf tells us how much sunlight it can capture for photosynthesis. Leaf surface area is a common calculation in plant research.

There is a catch with this second example, however. Are most leaves shaped like rectangles, circles, or triangles? Many leaf shapes are fairly regular, but they can be much more difficult to measure than simple geometric shapes! One approach is to estimate the area of a leaf shape by imagining it is made up of two or three simpler shapes. An elongated leaf might be broken up into a rectangle and two triangles, for instance (Figure 1).



$$A_l \approx (b_1 \times h_1)/2 + (b_2 \times h_2) + (b_3 \times h_3)/2$$

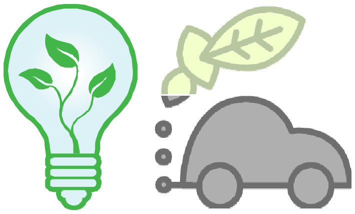
Figure 1. Estimating a leaf's area by breaking it into three shapes.

This approach works best when all the objects to be measured have the same overall shape and are broken into few "parts." It may not be as helpful in cases where the shapes differ among objects or have lots of parts, such as for lobed oak leaves. In this situation, you might instead trace the shape onto a piece of graph paper. Next, measure the area of one square in the graph paper. Count the number of squares the shape covers. To estimate the object's area, simply multiply the area of one square by the total number of squares.

Finally, you can also use software for estimating the area of irregularly-shaped objects. You may be able to directly scan leaves. If the leaves have lots of color variation, you may want to photocopy them in black and white, then scan the photocopied sheet instead. Objects in the saved, scanned image can be analyzed using a program such as ImageJ.

Mathematical Relationships:

Scientific research typically has a goal of finding connections between variables or the causes of our observations. How does the amount of light we provide affect the increase in a plant's mass? What happens when we add different amounts of nitrogen fertilizer to a set of plants? Does a pollutant limit a plant's seed production?



Some scientific ideas are tested as yes-or-no questions: "Is there a difference between a treatment and a control?" Others are tested or described as mathematical relationships: "If I add x amount of a substance, how much does y change?" A *dose-response curve* showing the effect of zero, one, two, or three grams of nitrogen fertilizer on plant growth is a good example of a mathematical relationship between an independent variable (the "dose" of fertilizer) and a dependent variable (the "response" of, for example, plant mass). Scientists usually write these relationships as equations linking the two variables. In a controlled experiment, the *independent variable* is assumed to directly or indirectly cause the change in the *dependent variable* if a relationship exists between them. Mathematical relationships come in many varieties and can be incredibly complex. Fortunately, a handful of equations can describe a broad range of biological observations.

Many mathematical relationships in biology are **linear**: *the change in the dependent variable with each unit of increase in the independent variable is the same throughout the data*. In ***The Power of Sunlight***, you may find that your plants produce ten additional grams of mass with each additional 100 μmol/m^2/s of light. The light intensity is the independent variable (y), and the plant mass is the dependent variable (x). All linear relationships have the form $y = mx + b$, where m is the *slope* of the line and b is where the line crosses the y-axis. In this example, m is 10 g/100 μmol/m^2/s, or 0.1. The amount of mass the plants produce without light would be the same as b, because $y = 0.1$ x $0 + b$. Linear relationships can be **positive**, where *one variable increases as the other increases*, or **negative**, meaning that one *variable decreases as the other increases*.

Thought Exercise: *What does a graph of a positive linear relationship look like? How is it different from a graph of a negative linear relationship? Does the plant mass/leaf number example describe a positive or negative relationship?*

Many relationships are nonlinear. For example, in ***The Wonder of Seeds***, you may find that nitrogen improves plant growth at lower concentrations but becomes toxic at very high concentrations. The overall graph of nitrogen and plant growth in this situation looks like an upside-down "U." **Quadratic relationships** are described in terms of the square of the independent variable, x^2 (e.g., Figure 2). The characteristic form of a quadratic equation is $y = ax^2 + bx + c$, where a, b, and c are constants.

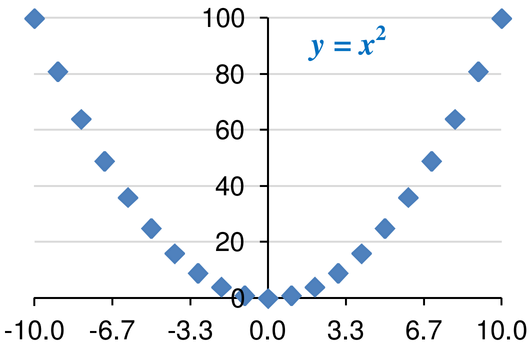


Figure 2. Example of a quadratic equation.

Suppose you add a small sample of algae to a large flask of nutrient solution. If you estimate the algal population over time using a microscope, you may discover the population *increases more quickly* over time, making the graph nonlinear. This third type of relationship seen in biology is an **exponential relationship**. The general equation for an exponential relationship is = $ae^{x+b} + c$, where $e \approx 2.718$ (e.g., Figure 3a). The constants b and c can shift the curve in the left-right or the up-down direction on the graph, respectively. In contrast, a will influence how quickly the curve changes and whether the change is upwards (for $a > 0$) or downwards (for $a < 0$).

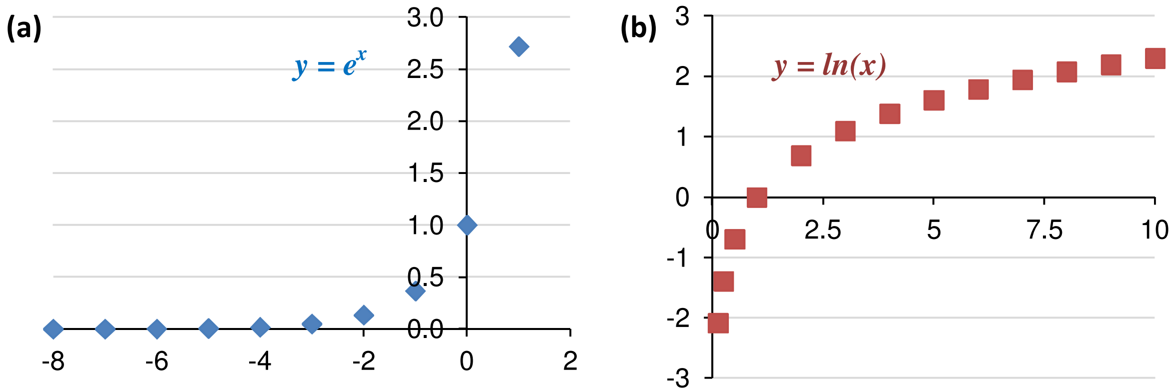


Figure 3. Examples of (a) an exponential curve and (b) a logarithmic curve.

If you measure photosynthetic rate in a leaf under different light intensities in ***The Power of Sunlight***, you may find that the rate does not increase as much for brighter light as it does for dim light. When *y changes at a decreasing rate with an increase in x*, the corresponding graph shows a **logarithmic relationship**. The general logarithmic equation is $= a\,ln(x + b) + c$, where $ln(x)$ is the **natural log** of x, which is defined as $e^{\ln(x)} = x$.

Thought Exercise: *How will the constants a, b, and c shift the curve in Figure 3b?*

Mathematical Patterns:

Have you ever looked carefully at a pineapple fruit, sunflower, or tree branches? If you have, you may have noticed interesting patterns. These patterns help account for the beauty of many plants and plant parts, and you can often describe them in mathematical terms.

Most plants and plant organs are built in a modular fashion. For example, new stems or branches are spaced fairly evenly with old ones, each producing similarly shaped leaves or flowers. This creates overall plant forms with regular, repeating patterns. The underlying math for these patterns typically arises because (1) plants grow from a limited number of points on their bodies and (2) plants that optimize their capacity for harvesting light, water, and soil nutrients tend to survive and reproduce better than those that do not.

Spirals are patterns appearing widely across the plant kingdom. Spirals may lie in two dimensions, as in the seed head of a sunflower, or in three dimensions, as in a pineapple or in the pattern of leaves along a stem. You can often see plant spirals in both a clockwise and the counterclockwise direction; in such cases, the number of spirals in each direction differs. The difference usually relates to a sequence of integers called the **Fibonacci sequence**. The sequence is created by starting with the number one, repeating it, then adding the previous two numbers for every new item in the series; that is:

$$1, 1, 2, 3, 5, 8, 13, \ldots, n_{i-1}, n_i, n_{i+1} = n_{i-1} + n_i$$

In other words, a pine cone that forms thirteen clockwise spirals when viewed from its base may form eight counterclockwise spirals from the same perspective.

You might have noticed that Fibonacci numbers also relate to the number of petals or sepals in many flowers. Many monocots, such as trilliums, have flowers with three petals and three sepals. Flowers in the rose family often have five petals and five sepals, and passionflowers or maypops have three sepals and two whorls of five petals. Careful counters have noticed that asters usually have 21 petals, and sunflowers usually have 55 or 89 – all Fibonacci numbers! Similarly, branching patterns and the number of leaves in a whorl when viewed from above often consist of Fibonacci numbers.

Fibonacci numbers are not the only pattern seen in plants. The **Lucas sequence** begins differently, but later numbers are formed using the same math as in the Fibonacci sequence to give a different list:

$$2, 1, 3, 4, 7, 11, 18, \ldots, n_{i-1}, n_i, n_{i+1} = n_{i-1} + n_i$$

Cruciferous plants like mustard, broccoli, and cauliflower produce individual flowers with four petals and four sepals. This Lucas number is present even though broccoli and cauliflower both produce flower heads with a number of spirals found in the Fibonacci sequence!

Thought Exercise: *Sometimes when a plant's early development is altered, it may form Lucas numbers instead of Fibonacci numbers. Can you think of a plant that usually has three leaves, but sometimes produces four?*

Image: Wikimedia Commons

Spirals are a specific, simple type of geometric pattern known as **fractals**. Fractals are usually mathematically described in simple equations that, when graphed, turn out to have the same form regardless of the scale at which they are viewed. In nature, fractals are usually observed as repeating patterns that decrease in size. Romanesco is an excellent example, as you can see in the photo at left: it is a type of cauliflower that shows repeating, diminishing patterns of cones in each head of flowers.

Using Statistics

Statistics are important to scientists for three reasons:

- They quantitatively describe and summarize data.
- They help researchers draw valid conclusions based on relatively small sets of data.
- Differences and relationships between many kinds of data can be objectively analyzed.

In this section you will learn about some types of statistics biologists use. Ways to calculate statistics with the help a computer are described in the section **Data Analysis Using Spreadsheets**.

What is a Typical Value?

Living things are, by their nature, variable. No single individual, population, or community will be identical to any other. To describe any group of living things, **descriptive statistics**, or *summary measures derived from sample data*, become necessary.

One important question in describing your data is "What number can be used to best represent the most samples?" That is, you need to figure out the **central tendency** of the data. One common way of doing this is by using the **median**, or the exact middle point of all your data when they are listed from smallest to largest. Suppose you collected the following height data for a group of five control plants for a project in ***The Wonder of Seeds***: 11.7, 12.1, 10.1, 11.4, and 13.3 cm. To find the median, first list these values in order. Next, eliminate the two smallest and two largest values, leaving the middle one:

~~10.1~~, ~~11.4~~, 11.7, ~~12.1~~, ~~13.3~~

Therefore, the median of the data set is 11.7 cm. If you had six samples instead of five, you would add the third and fourth values, then divide the sum by two.

The **mean**, or *average* ($\overline{x}$), is the most widely-used measure of central tendency. To calculate the mean, simply add up all of the data you have collected, then divide the sum by the total number of samples. For the previous example, you collected data for five samples. You would therefore calculate:

$$\overline{x} = (10.1 + 11.4 + 11.7 + 12.1 + 13.3\text{ cm})/5 = 58.6\text{ cm}/5 = 11.72\text{ cm}$$

The mean height of the control plants was 11.72 cm. This is close to the median, but it's not the same! Also notice that the mean is one significant figure more precise than the original data. Each plant was measured to the nearest 0.1 cm. Therefore, you can round the mean to the nearest 0.01 cm.

How Much Does My Data Vary?

Suppose you are carrying out a leaf disc floatation trial for ***The Power of Sunlight*** and record how many minutes it takes for five leaf discs to float in a control and in an experimental treatment. If the control samples take 10, 11, 11, 12, and 11 min to float, what are the mean and the median for this set? If the experimental samples take 1, 6, 11, 21, and 16 min to float, what are their mean and median?

You will probably agree that central tendency is not enough to meaningfully describe these two data sets. Means and medians do not tell how much **variation** exists within the data. Therefore, scientists also need some way to describe how far the samples are from the mean or median. Two common measures of variability are **standard deviation (SD)** and **standard error of the mean (SEM)**. These two measurements are somewhat related to each other, as you will see.

To calculate both SD and SEM, the first step is to calculate the **sum of squares (SS)**. Subtract the mean from each data point, then square each difference. Finally, add up all of the squared differences. For the control group leaf floatation data:

$$\text{SS} = (10 - 11)^2 + \cancel{(11 - 11)^2} + \cancel{(11 - 11)^2} + (12 - 11)^2 + \cancel{(11 - 11)^2} = 2\text{ min}^2.$$

A simpler way to calculate SS is to add up all of the squares of the data (a), then subtract from it the sum of the data (b) squared divided by the number of samples (n).

$$\text{SS} = a - b^2/n.$$

For the experimental group leaf floatation data, we have:

$$\text{SS} = (1^2 + 6^2 + 11^2 + 21^2 + 16^2) - (1 + 6 + 11 + 21 + 16)^2/5 = 855 - 55^2/5 = 855 - 605 = 250\text{ min}^2.$$

Once you know the sum of squares, you can calculate the **standard deviation** is by dividing SS by the degrees of freedom. This is one less than the total number of data points for the set, or $n - 1$. For the control data set, we have:

$$\text{SD} = \sqrt{(2\text{ min}^2)/(5-1)} = \sqrt{0.5\text{ min}^2} = 0.7\text{ min}$$

The **standard error of the mean** is calculated based on SD and the number of samples, by using the formula **SEM** = SD/$\sqrt{n}$. For the control data set, this yields:

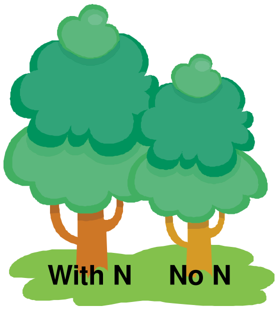
$$\text{SEM} = 0.7\text{ cm}/\sqrt{5} = 0.3\text{ cm}$$

Thought Exercise: *How would you use the sum of squares for the experimental leaf discs to calculate SD and SEM for that data set?*

You can accurately summarize a data set by stating either the mean and SD ($\overline{x}$ = 11.0 min, SD = 0.7 min) or the mean and SEM ($\overline{x}$ = 11.0 min, SE = 0.3 min). Notice that, like for mean, SD and SEM are described with a precision one decimal point greater than the original measurements. Furthermore, because SD and SEM are so similar, it is critical to tell which one you are using in a table or a graph's error bars.

You may have noticed that sample variation decreases as sample size increases, because you must divide by n - 1 to find SD. This means you can get more precise information about biological variation, and overall better results, if you collect more data. However, collecting more data takes more time. You will need to weigh these costs and benefits to choose a good sample size for your experiment!

Are My Treatments Different?



Scientists carry out experiments to test hypotheses. Because organisms are variable, biological data will rarely match exactly what our hypotheses predict. It is therefore helpful to test whether or not a difference exists between data from control and experimental treatments. *When a difference is larger than expected due to chance*, it is said to be **statistically significant**. Science limits the use of "significance" to cases where a statistical test has shown clear, unexpected differences between two means, or between observed and expected results.

In some cases, the difference between two sets of data might be very large and obvious, but in other cases, the means and variation may be quite similar. For example, suppose you wanted to compare the curvature of celery sticks soaked at two different salt water concentrations for ***The Celery Challenge***. If you soaked three sticks in each treatment and found bending angles of 15^o, 20^o, and 17^o for 5% salt water, would this be significantly different from celery sticks that bent 20^o, 17^o, and 25^o in the 10% salt water? Statistical tests provide an objective method to determine the amount of difference or similarity when differences are hard to determine.

The **Student's t-test** is a statistical test often used to compare two means. In this test, the *null hypothesis* is H_o: $\overline{x}_1 = \overline{x}_2$; that is, both means are the same and variation is due mainly to chance. The null hypothesis may or may not be rejected, depending on the size of the *t*-statistic. The *t*-statistic is somewhat complicated to calculate:

$$t = \frac{|\,\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}}.$$

The numerator is the absolute value of the difference between the means for the two data sets. For the celery bending example, the means are 17.3^{o} for the 5% salt water group and 20.7^{o} and for the 10% group. The absolute value of $17.3^{o} – 20.7^{o}$ is 3.4^{o}.

Since a data set's SD is the square root of SS divided by *df* (n – 1), SD^2 is simply SS/(n-1). For the celery data, $SS_{5\%} = (-2.3^{o})^2 + (-0.3^{o})^2 + (2.7^{o})^2 = 12.7^{oo}$, and $SS_{10\%} =(-0.7^{o})^2 + (-3.7^{o})^2 + (4.3^{o})^2 = 32.7^{oo}$, while *df* for both groups is 3-1, or 2. Therefore, $SD_{5\%}{}^2 = (12.7^{oo}/2) = 6.4^{oo}$, and $SD_{10\%}{}^2 = (32.7^{oo}/2) = 16.4^{oo}$. Finally, we can calculate the denominator:

$$\sqrt{\frac{SD_1^2}{n_1}+\frac{SD_2^2}{n_2}} = \sqrt{\frac{6.4^{oo}}{3}+\frac{16.4^{oo}}{3}} = \sqrt{7.6^{oo}} = 2.8^{o}.$$

Overall, then, for the celery data we have $t = 3.4^{o}/2.8^{o} = 1.2$.

Thought Exercise: *What is the value of* t *for the leaf floatation data sets used in the earlier example?*

Is this value of *t* large or small? To find out, we can compare it to one member in a table of **critical values** (Table 1). To find the correct critical value, you first need to choose the appropriate degrees of freedom. The overall *df* in the table is the sum of the *df* for each sample set: $df = (n_1\text{-}1) + (n_2\text{-}1) = n_1 + n_2 - 2$. For the celery data, this is 3 + 3 – 2 = 4.

Table 1. Critical Values for Student's *t*, Based on *df* and α.

df	α = 0.10	α = 0.05	*df*	α = 0.10	α = 0.05
1	6.31	12.71	11	1.80	2.20
2	2.92	4.31	12	1.78	2.18
3	2.35	3.18	13	1.77	2.16
4	2.13	2.78	14	1.76	2.14
5	2.01	2.57	15	1.75	2.13
6	1.94	2.45	16	1.75	2.12
7	1.89	2.36	17	1.74	2.11
8	1.86	2.31	18	1.73	2.10
9	1.83	2.26	19	1.73	2.09
10	1.81	2.23	20	1.72	2.09

Next, you must set the desired **level of significance**, or α. This is *the probability of rejecting the null hypothesis when it is, in fact, true*. When a true null hypothesis is rejected, a **Type I error** has been made. The most common level of significance in biology is α = 0.05. This means the probability of falsely rejecting the null hypothesis is 5%. You can select a higher or lower value for α as desired if your table has this information. Unfortunately, by reducing the probability of making a Type I error, you increase the chance of *accepting* the null hypothesis when it is *not* true -- a **Type II error**! Biologists use α =0.05 because it provides a balance between the likelihood of Type I and Type II errors.

In the celery example, *df* = 4 and α = 0.05 corresponds to a critical value of 2.78. The calculated *t* value 1.2 is *smaller than* the critical value, which tells us the null hypothesis, H_o: $\overline{x}_1 = \overline{x}_2$ cannot be rejected. The difference in mean bending angle between the two treatments is mainly due to random variation, not the salt water concentration. However, in any situation where *t* is *greater than or equal to* the critical value, the difference between means is bigger than expected due to chance. You can reject the null hypothesis in such cases, because the alternative hypothesis of H_A: $\overline{x}_1 \neq \overline{x}_2$ is a better explanation.

Making Meaningful Tables and Graphs

To make their raw data exciting, useful, and meaningful to others, scientists figure out ways to clearly summarize, analyze, and present it. You've already seen how data can be summarized in the **Statistics** section, but what do you do with that information once you have it?

Some results are best summarized in a table. Qualitative data can be described in a written format this way; for example, you might be able to organize a list of different groups of plants, then make notes about leaf or flower colors within each group. If you have numerical data as well, you can present the mean and standard deviation of the measurements for each group in the same table.

On the other hand, most quantitative data are more easily understood in graphs. Instead of thinking about what the numbers mean, we can visually judge what we see in a graph. Your fellow scientists in the classroom can probably more quickly understand the main findings of this work if you use a graph.

Making useful graphs is much easier if you already know how to read a graph. In most cases, you will see a large square with lines or bars in it, and along the bottom and the left side, you will find numbered, labeled **axes**. The labels written on each axis tell you what measurements (and their units) or treatment groups are shown, while the numbers show the scales involved. By convention, the horizontal axis, or x-axis, is where the independent variable is plotted. The dependent variable (your measurement data) is plotted on the vertical axis, or y-axis. You might also see a key describing what the different colors or patterns in a graph indicate.

You can make a graph by hand or using a computer. In either case, be sure that you plot your data accurately and clearly. For example, if you are making a graph by hand, use graph paper and a straight-edge. The kind of graph you choose to make depends on the kind of data you collect and what you want to show. Lines, bars, histograms, scatter plots, and pie charts are all useful types of graphs. Each is best suited for presenting a different kind of information. For example, the leaf growth rate of an oat plant, or its change in leaf length over time, could be shown in a line graph. If you want to compare the percentage of seeds that germinated in control and test treatments, try a bar chart. In the following sections, you can learn about what each type of graph does best.

Line Graphs:

A **line graph** is used when the *dependent variable is continuous, and the independent variable represents specific points that are not continuous, but are still part of a broader, continuous set of possible numbers*. Consider the earlier example of oat leaf length. Leaf length is a *continuous variable*. It was also measured at daily intervals. This interval is not continuous data, but in theory you could make a measurement at any point in between two actual sample points.

The resulting data for your oat leaf length measurements could be shown in a table (Table 2), but a graph would probably be more helpful to show how the sampling time relates this continuous data. For example, if you sampled at 1.5 days, leaf length would be somewhere between 10 and 15 cm; at 3.75 days, it would be between 18 and 24 cm.

Table 2. Example Data: Mean and SD of Oat Leaf Length Over Four Days

Sample Point (day)	Mean Length (cm)	SD of Length (cm)
1	10	1.0
2	15	2.5
3	18	2.0
4	24	0.5

Plotting the data as a line graph helps show that at intermediate times, we would expect intermediate lengths based on the line the graph's data follows (Figure 4). In a line graph, data points are **not** connected by a series of chords. Instead, a **single** curve or line of best fit is plotted with the data. The slope of this line is actually the *average value* at which change is occurring over the sampling period.

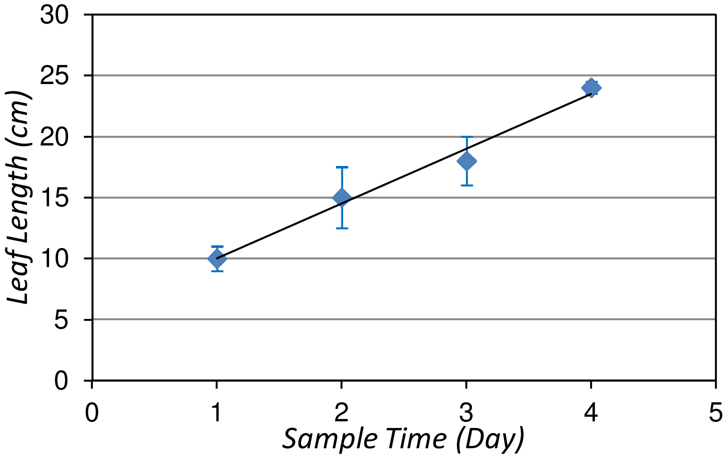


Figure 4. Example line graph using data from Table 2.

Bar Graphs:

Bar graphs are used *when the independent variable involves qualitative categories*, such as different species. The dependent variable again involves continuous data. Good examples of qualitative independent variables are treatment type and species. Table 3 shows hypothetical data for the effects of experimental and control treatments on germination in four different crop species.

Table 3. Germination of Four Plant Species in Control (Ctrl) and Experimental (Exptl) Conditions

Species	Ctrl Mean (%)	Ctrl SD (%)	Exptl Mean (%)	Exptl SD (%)
Bean	72	3.0	50	2.0
Corn	90	2.0	95	2.5
Pea	78	5.0	87	3.0
Squash	84	2.5	80	4.0

You might have noticed that two different types of independent variables are present in this data. Treatment type and species are both independent variables, and both are qualitative categories! To make a graph in this situation, each species can be listed on the x-axis, while different treatments can be shown using different colors (Figure 5). Each individual bar indicates one mean, while the error bars show the corresponding standard deviation in both directions.

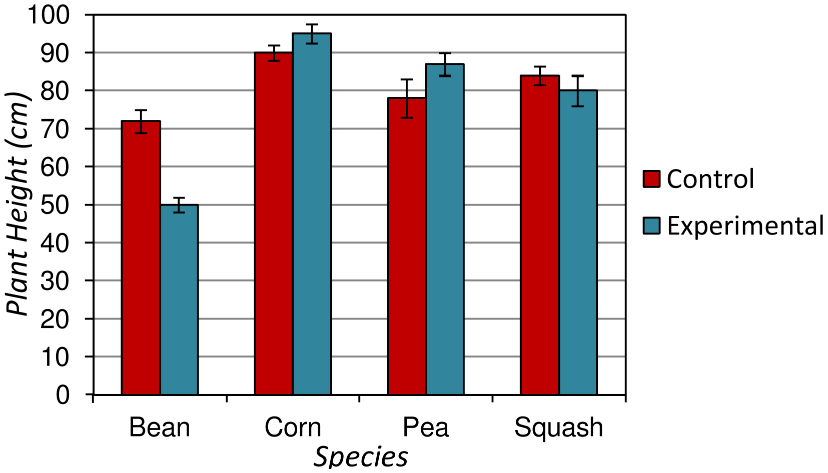


Figure 5. Example bar graph using data from Table 3.

Histograms:

Histograms represent *frequency distributions*, in which a series of numerical ranges is the independent variable and the number of items within each range is the dependent variable. For instance, you might be interested in studying how genetics affects height for the ***Genetics in Inbred*** **Arabidopsis** module. Before growing *Arabidopsis* plants in an F1 generation, you can define size classes based on the heights you expect to see after two weeks, such as <5 cm, 5-10 cm, 10-15 cm, and >15 cm.

Table 4. *Arabidopsis* F1 Plants in Four Height Classes

Plant Height (cm)	Number of Plants
<5	4
5 – 10	15
10 – 15	13
>15	8

Table 4 is a hypothetical set of such data. You may find it easier to see the overall distribution of plant heights using the histogram in Figure 6. Notice that the continuum of size classes is indicated by placing the bars adjacent to each other. This is different from a bar graph, where the bars are separated.

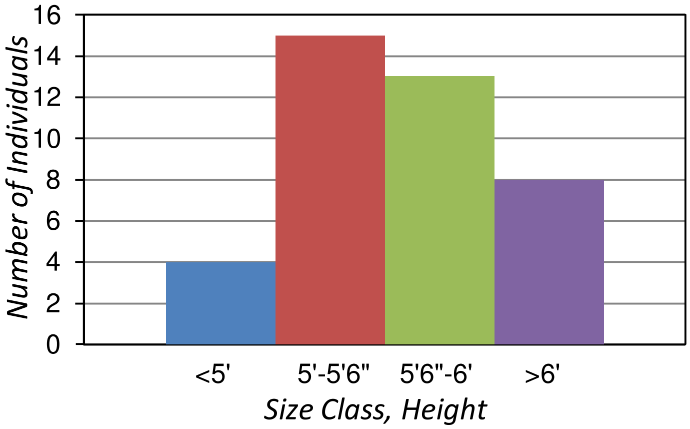


Figure 6. Example histogram using data from Table 4.

Scatter Plots:

Unlike other graphs, where mean values are often plotted, **scatter plots** include every individual data point collected. Scatter plots can help show the strength of association between two variables. For example, Figure 7 shows the relationship between leaf length and leaf width in one plant.

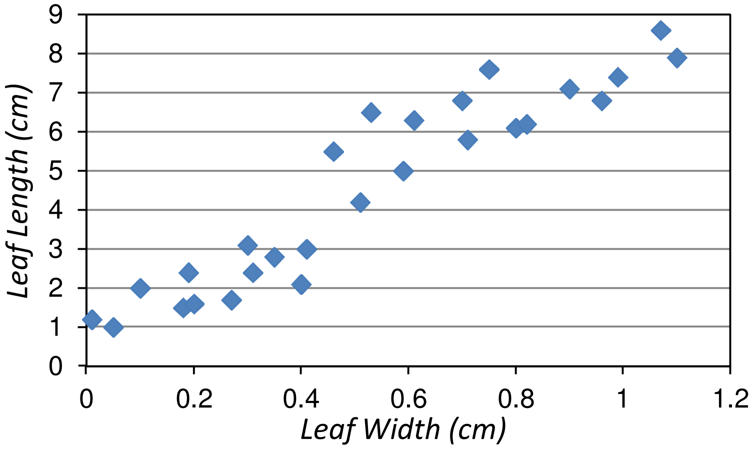


Figure 7. Example scatter plot based on hypothetical leaf data for a single plant.

Scatter plots look similar to line graphs, but they have important differences. In a line graph, the dependent variable is continuous, while the independent variable is restricted by the experiment. In a scatter plot, both variables have continuous variation. One way to think about this is to ask whether the independent variable is chosen or measured. If you use three different, set levels of fertilizer in different treatments for an experiment on plant height in ***The Wonder of Seeds***, you would use a line graph. If you carried out an experiment in which you measured the levels of fertilizer along with plant height, you would use a scatter plot.

Pie Charts:

While most types of graphs are plotted on an x-axis and a y-axis, **pie charts** plot data by dividing up a circle into “slices” of different sizes. This type of graph is great for showing what parts make up a whole, and what proportion each part contributes. That is, pie charts represent a dependent variable in both qualitative and quantitative terms. The example data set in Table 5 shows the types and amounts of pollinators that visited two different sunflowers during a field study. Which sunflower was visited is a categorical, independent variable. Pollinator type is a *qualitative* dependent variable, while the number and percentage of each type of pollinator are *quantitative* dependent variables.

Table 5. Pollinators Visiting Two Different Sunflowers

Pollinators	Sunflower A (Number)	Sunflower A (%)	Sunflower B (Number)	Sunflower B (%)
Bees	12	40.0	14	63.6
Butterflies	11	36.7	7	31.2
Others	7	23.3	1	4.5
Total	*30*	*100*	*22*	*100*

As you can see, it is a challenge to take in both numbers and percentages at the same time! A pie chart can help make this easier. In Figure 8, the three different pollinator categories are each different colors. You can see the proportion that each type of pollinator contributed to the total visits. To be even clearer, you could label each slice with the percentage or number of visits that it represents.

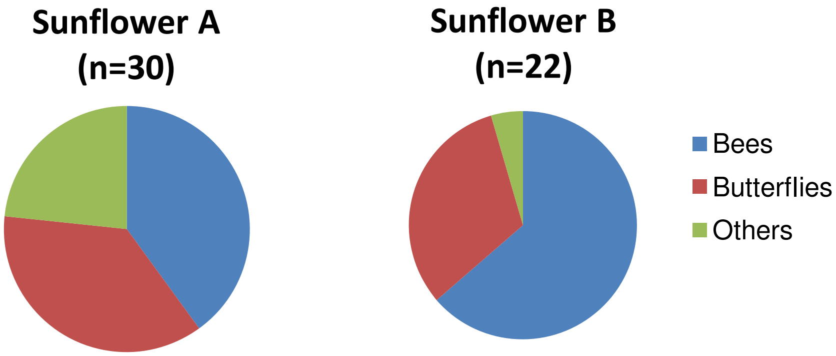


Figure 8. Example pie charts using data from Table 5.

Thought Exercise: *In what two ways is the overall number of pollinator visits for each sunflower represented in Figure 8?*

Data Analysis Using Spreadsheets

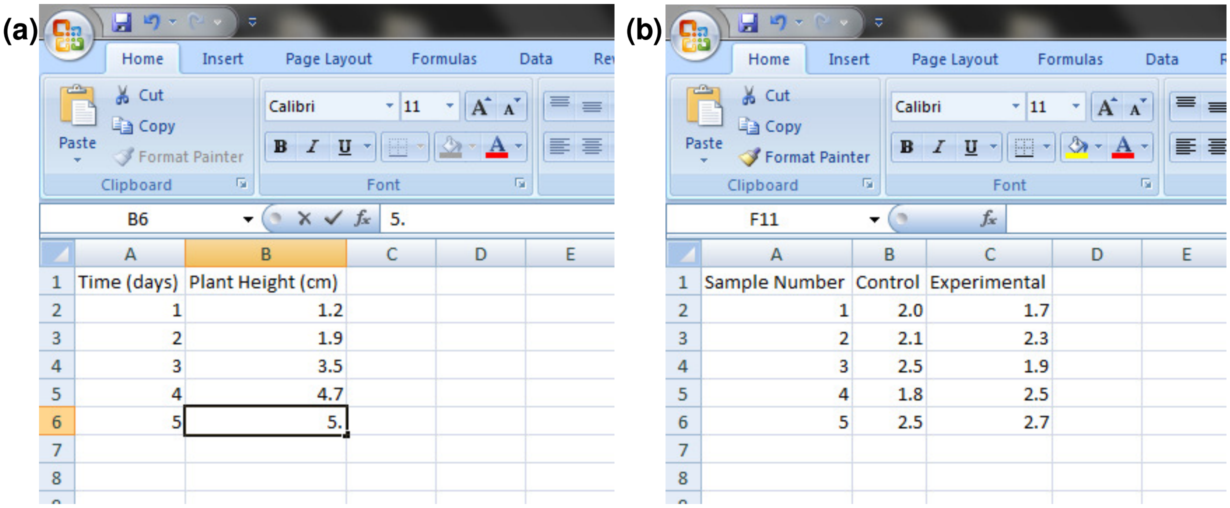
A spreadsheet is a simple but powerful tool for storing information. Most of a spreadsheet is made up of a grid of cells. To help you keep track of where you are, numbers run down the left side of the page

(1,2,3,4,…) and letters run across the top of the page (A, B, C, D,…). Each number acts as a **ROW** heading for a line of cells running across the page. Each letter acts as a **COLUMN** heading for a line of cells running down the page. Scientists use the headings to identify each cell. For example, the very first cell on a spreadsheet is A1. The next cell in the same row is B1, the third C1, and the tenth J1. The second cell in Column A is A2, the third cell is A3, and the fifteenth is A15.

Spreadsheet software can help you quickly analyze and graph the data you have collected. In this section, examples of "number crunching" are shown using MS Excel 2007. Software is updated every few years and differs by brand, so the exact procedures you use may be slightly different from those described here. Nonetheless, you can still get a general idea of how to use spreadsheets.

__Entering Data:__

When you open the spreadsheet program, a blank sheet will fill most of the window. The top part will be filled with tabs containing different kinds of tools. To enter your data, just click on a cell and type in it. Add labels and organize the information so that others can easily understand what you are showing. For example, each data set should be put in its own column or row. One approach is to list the name of each variable and the units in which it was measured in the first row. Then you can add data for the independent variable (data that you might plot on an x-axis) running down one column and data for each dependent variables down separate columns.



(a) B6 — 5.

	A	B	C	D	E
1	Time (days)	Plant Height (cm)			
2	1	1.2			
3	2	1.9			
4	3	3.5			
5	4	4.7			
6	5	5.			
7					
8					

(b) F11

	A	B	C	D	E
1	Sample Number	Control	Experimental		
2	1	2.0	1.7		
3	2	2.1	2.3		
4	3	2.5	1.9		
5	4	1.8	2.5		
6	5	2.5	2.7		
7					
8					

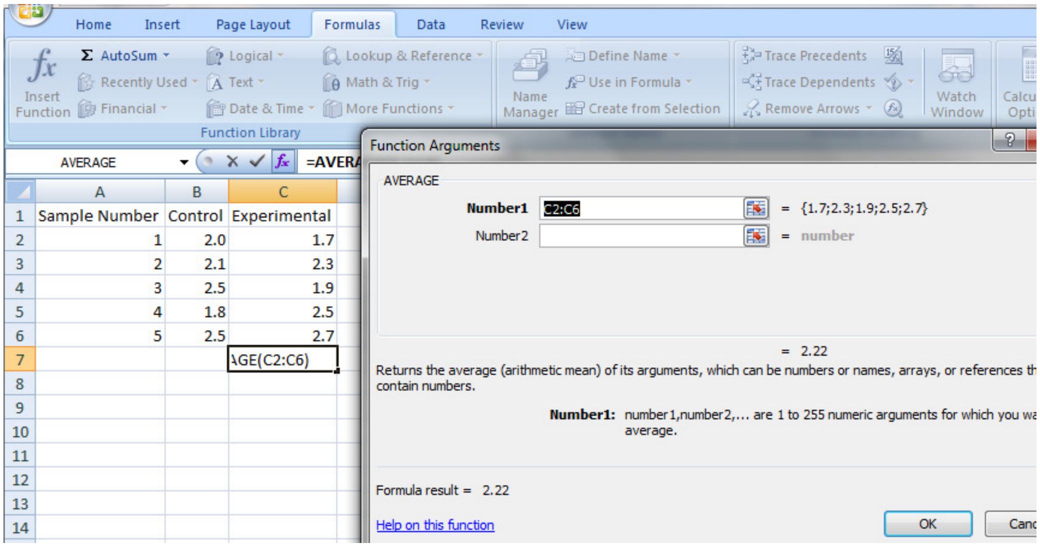
Figure 9. Data entry for (a) a time-course experiments and (b) comparing two conditions.

Suppose you have collected data for a plant growth experiment. You might label Column A as "Time (days)," and the later rows in that column would be listed sequentially as 1, 2, 3, 4, 5 – one for each day you collected data (Figure 9a). The dependent variable could be listed in B1 as "Plant Height (cm)." Each measurement would be listed in Column B alongside the day that measurement was recorded. Perhaps you have carried out an experiment to see if fertilizer affects leaf width after two weeks. You could list Column A as the sample number, while Columns B and C could list leaf width data from "Control" and "Experimental" fertilizer treatments, respectively (Figure 9b).

Descriptive Statistics:

The spreadsheet software can automatically calculate descriptive statistics such as mean, median, or standard deviation by using functions. To do this, select an empty box in the sheet where you want to put one of the descriptive statistics. For example, the first empty row after your raw data might be a good choice. In MS Excel 2007, select the **Formulas** tab in the bar above the spreadsheet, then click on the two orange books that say **More Functions**. In the drop-down menu, choose **Statistical** in the drop-down menu. The new menu that opens will have many different functions listed in alphabetical order, so you may have to scroll down to find the one you want. Choose the "AVERAGE" function to calculate a mean, "MEDIAN" to calculate that statistic, or "STDEV" to calculate standard deviation for any of your numerical data sets.

When you choose a function, a new window will open up. Inside it will be a cell describing the range of data to be used for the calculation (Figure 10). You can type or click and drag the data range you want. Be sure this input range includes ***only*** the data you want in the calculation, then click "OK." The output will now be shown in the cell you selected earlier. You can use tools in the **Font** box of the **Home** tab to highlight this cell or use a different font style or color. This will help you and others see that this cell is not part of the raw data.



AVERAGE =AVERA

	A	B	C
1	Sample Number	Control	Experimental
2	1	2.0	1.7
3	2	2.1	2.3
4	3	2.5	1.9
5	4	1.8	2.5
6	5	2.5	2.7
7			\GE(C2:C6)
8			
9			
10			
11			
12			
13			
14			

Function Arguments

AVERAGE

Number1 C2:C6 = {1.7;2.3;1.9;2.5;2.7}

Number2 = number

= 2.22

Returns the average (arithmetic mean) of its arguments, which can be numbers or names, arrays, or references th contain numbers.

Number1: number1,number2,... are 1 to 255 numeric arguments for which you wa average.

Formula result = 2.22

Help on this function

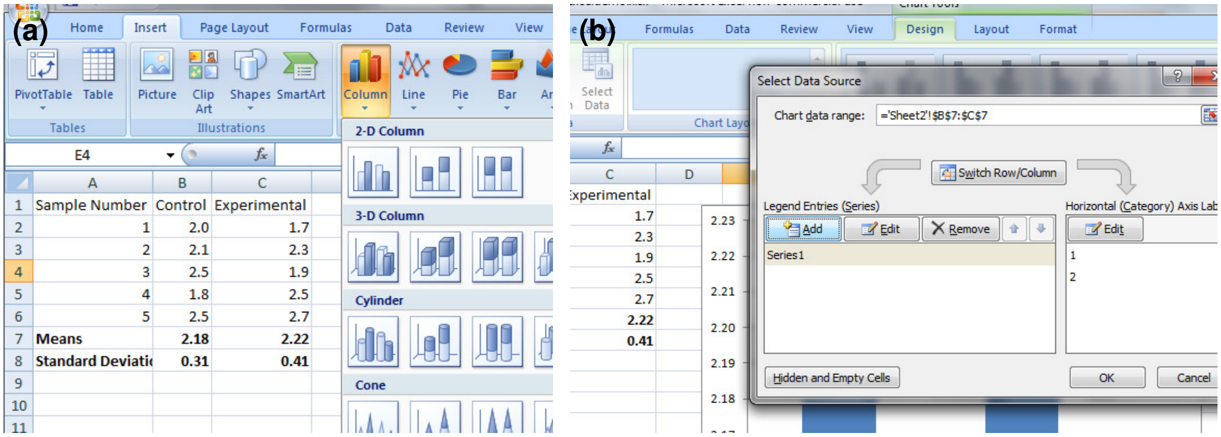
OK Canc

Figure 10. Calculating the mean of a data set using the AVERAGE function. Note that cell C7 contains the formula for the function averaging the data in cells C2 through C6. The **Formulas** tab can be seen above the "Function Arguments" box.

Making Graphs:

You can also create graphs using the data in your spreadsheet. In MS Excel 2007, select the **Insert** tab in the bar above the sheet. The **Charts** box in this tab shows several options for what kinds of graphs you can make. Simply click on the general type of graph you want, then select a specific design from the pull-down menu (Figure 11a). An empty box will be added to the spreadsheet. This might cover some of the cells you want to see, but you can move the box by clicking and dragging it with your cursor.

The next step is to add data to the graph. When you add a new chart to the spreadsheet, a new set of tabs labeled **Chart Tools** will show up at the top of the window. In the **Design** tab, click on "Select Data" to open a new window for entering the chart data. Clicking on "Add" in this window will allow you to add a new series of dependent variable data to the chart, such as a group of means for different treatments (Figure 11b). If you like, you may also add a series name to describe the variable in the new, smaller window that pops up. After the data are entered, you can also name the independent variable categories by clicking on "Edit" and adding the category names in the same order the data were added.



	A	B	C
1	Sample Number	Control	Experimental
2	1	2.0	1.7
3	2	2.1	2.3
4	3	2.5	1.9
5	4	1.8	2.5
6	5	2.5	2.7
7	Means	2.18	2.22
8	Standard Deviati	0.31	0.41

Chart data range: ='Sheet2'!$B$7:$C$7

Legend Entries (Series): Series1

Horizontal (Category) Axis Lab: 1, 2

Figure 11. Adding a graph to a spreadsheet (a) and adding data to the graph (b) in MS Excel 2007. Note the "Select Data" icon in the upper left corner of (b).

The box will now show a graph, with the data range for the dependent variable listed on the y-axis and the independent variable categories listed on the x-axis. You can change the fonts, the numerical range of the axis, the colors of the items, or even add error bars to the graph using tools from the **Layout** and **Format** tabs within the **Chart Tools** section. You can also simply right-click on the item in the graph you want to change, then choose one of the categories in the menu that pops up. Once you are happy with your graph (e.g., Figure 12), you can copy it, then paste it into a lab report or presentation file.

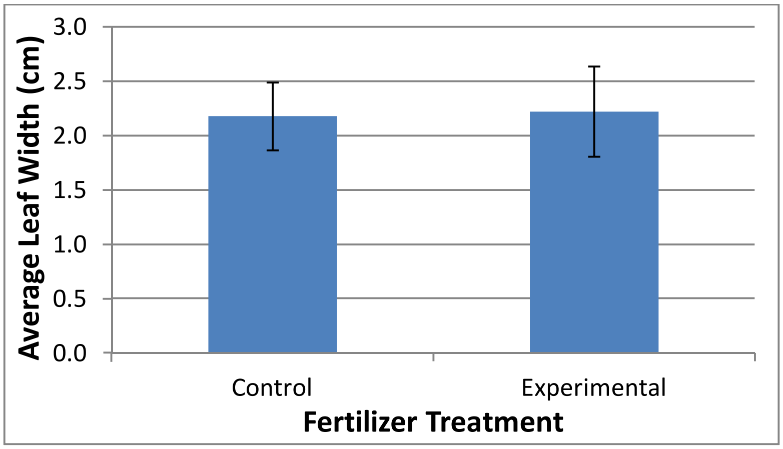


Figure 12. Example of a finished bar graph created in MS Excel 2007.

Additional Resources

Videos:

Doodling in Math: Spirals, Fibonacci, and Being a Plant, by Vi Hart. A mathemagician explores spirals in plants and on graph paper. She builds relationships between the Fibonacci and Lucas sequences, plant spirals, and doodling.
http://www.youtube.com/watch?v=ahXIMUkSXX0
http://www.youtube.com/watch?v=lOIP_Z_-0Hs
http://www.youtube.com/watch?v=14-NdQwKz9w

Web Pages:

Analysis and Presentation of Data, by Biocyclopedia. Part of a larger plant biology website, this page has many links to tips for presenting data in graphs and tables, solving math problems, and basic statistics.
http://www.eplantscience.com/botanical_biotechnology_biology_chemistry/dean/analysis_and_presentation_of_data.php

Computers, by Biocyclopedia. This page describes the many types of software used in biology research.
http://www.biocyclopedia.com/index/biotechnology_methods/tools_and_techniques_in_biological_studies/computers.php

Fibonacci Numbers and Nature, by Ron Knott. This large page is part of a math website describing Fibonacci numbers and the golden section. The link will place you at the start of a section focusing on plants. Scroll up to find more connections to biology.
http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fibnat.html#plants

ImageJ: Image Processing and Analysis in Java, by the Research Services Branch of NIMH's National Institute of Neurological Disorders and Stroke. Here you can download and learn about ImageJ, software for analyzing, displaying, saving, and printing images of up to 32 bits in many file formats.
http://rsbweb.nih.gov/ij/

Phyllotaxis: An Interactive Site for the Mathematical Study of Plant Pattern Formation, by Pau Atela and Christophe Golé from Smith College. Great photos of plant patterns are linked to mathematical modeling research and learning tools.
http://www.math.smith.edu/phyllo/

Plants "Do Maths" to Control Overnight Food Supplies, by Helen Briggs. This BBC News article discusses how plants use "molecular math" to precisely regulate their use of starch at night.
http://www.bbc.co.uk/news/health-22991838

Books and Articles:

Jelen, B. 2011. *Charts and Graphs: Microsoft Excel 2010.* Indianapolis, Indiana: Que Publishing. 496 pp.

O'Neal, M.E., Landis, D.A., and R. Isaacs. 2002. An inexpensive, accurate method for measuring leaf area and defoliation through digital image analysis. *Journal of Economic Entomology* 95(6): 1190-1194.

Robbins, N.B. 2005. *Creating More Effective Graphs.* Hoboken, New Jersey: John Wiley & Sons, Inc. 420 pp.

Walker, R.L., Burns, I.G., and J. Moorby. 2001. Repsonses of plant growth rate to nitrogen supply: a comparison of relative addition and N interruption treatments. *Journal of Experimental Botany* 52(355): 309-317.

White, J.M., Barrett, K.D., Kopp, J., Manoux, C., Johnson, K., and Y. McCullough. 2006. *Math in the Garden: Hands-On Activities that Bring Math to Life.* Williston, Vermont: National Gardening Association. 160 pp.