Unit 8: Everything Varies

Biology in a Box

A science education outreach program brought to you by a partnership
between The University of Tennessee and the National Institute for Mathematical and Biological Synthesis

Visit us on the web at http://biologyinabox.utk.edu

***Biology in a Box* Team**
Program Director/Author/Lead Editor: Dr. Susan E. Riechert (University of Tennessee)
Science Collaborators: Dr. Thomas C. Jones (East Tennessee State University), Dr. Stan Guffey (University of Tennessee)
Mathematics Collaborators/Editors: Dr. Suzanne Lenhart (NIMBioS), Kelly Sturner (NIMBioS), Lu Howard (University of Tennessee)
Outreach Coordinator: Dr. Lynn Champion (University of Tennessee)
Workshop Coordinators: Kathy DeWein (Austin Peay State University), Gale Stanley (Jacksboro Middle School)
Production/Assistant Editor: J.R. Jones (University of Tennessee)
Previous Contributors: Sarah Duncan, Communications (formerly with NIMBioS), Rachel Leander, Math Collaborator (formerly with NIMBioS)

This unit revised December 2012

Unit 8: Everything Varies
Materials List

- Container A, which contains
    - 35 unique Leaves

- Container B, which contains
    - 30 leaves of species A
    - 30 leaves of species B

- Container C, which contains
    - 30 mollusk shells with red dot on each

- Container D, which contains
    - 30 mollusk shells with blue dot on each

- Container E, which contains
    - A “mystery shell”

- Bag containing
    - 6 individual bags, each with
        - 10 assorted snail shells

- 6 rulers

- 6 protractors

- 1 container with a cloth bag of beads of 2 colors

Unit 8: Everything Varies
Table of Contents

Introduction

There are approximately 1.8 million *named* species. However, scientists estimate that the Earth may actually be home to over 10 million species. Why are there so many different kinds of living things? Or, to put it another way, why is there so much variety? To answer this question, we will consider the ways in which Earth itself varies. To start with, Earth has a lot of geological variation, that is, it has many different types of physical structures. Mountains, plains, and oceans are a few examples. Also, the Earth's shape and the way it orbits the sun create variation in Earth's climate. For example, some areas are warm and rainy year round, while other areas are warm but dry year round. Some areas are cold year round, while others, like Tennessee, have warm summers and cold winters. Climatic and geographical variability create a variety of habitats, or homes, in which organisms can live. So, why are there so many different kinds of living organisms? Organisms vary because everything else varies, and each species has adapted traits that help it to survive in the place where it lives.

In this series of exercises, you will explore the variation of life by examining leaves and seashells. Following is a brief summary of each of the exercises in this unit, and how each of them will help you explore the concept of variation.

Exercise 1: Recognizing Individuals as Unique tests your memory and powers of observation as you learn to identify a particular leaf from a group of leaves. Your leaf will vary a lot from many of the leaves but only slightly from some!

Exercise 2: Matching Leaves is another easy game in which you match each leaf in the box to a leaf on a leaf key.

Exercise 3: Using Mathematics to Describe a Population teaches you how to quantify how much variation there is in a sample of leaves, as well as to compare the amount of variation between leaves of different species.

Exercises 4 - 7 involve seashells, which protect the soft bodies of snails, clams, and their relatives:

Exercise 4: Simple Shell Match is a matching game that uses shells instead of leaves.

Exercise 5: Making Sense of Variation: The Matching Game, you will group organisms of like type together on the basis of the characteristics they share.

Exercise 6: Finding Species allows you to play the role of a taxonomist, deciding how to group shells into species by repeatedly splitting a sample of individuals into groups of similar shells.

Exercise 7: Mystery Shell tests your thinking skills as you attempt to determine the identity of a very unique shell.

Exercise 8: Mechanisms Underlying Variation explores the sources of trait variation in organisms, and illustrates the importance of variation in various contexts, again going back over leaves and shells as good examples.

Exercise 1: Recognizing Individuals as Unique (*Grades 3-12*)

A first run before discussion

- Locate the container marked A, which holds 35 laminated tree and shrub leaves. Each leaf has a unique number on its back.
- Each student should select one leaf from this batch. She or he should carefully examine the upper surface of this leaf (the side without the number) for a few minutes with the objective of committing its appearance to memory.
- After studying the leaf, the student should turn it over and write down the number on its back. *(Alternatively, the teacher might also keep track of the numbers for the students).*
- Return the leaves to the bag.
- Spread the leaves face up on a table or some other surface and instruct each student to find his or her 'special leaf' without looking at the number on its back.
- Check the number on each student's leaf to see if it matches the number from the leaf that he or she originally examined.
- Count the number of students that were successful in finding their leaves. What was the class success rate?

Rate can be expressed in different ways. It can be expressed as a fraction, a decimal, or a percent. For this exercise, you will find the group success rate using a percent. How do you do that?

> **Example:** If 10 individuals out of 30 found their leaves, the success rate can be expressed as a fraction (ratio), which can then be converted to a decimal, and then converted to a percent.
>
> $$\frac{number\ correct}{total\ number} = \frac{10}{30} = 10 \div 30 = 0.333\ldots$$
>
> For this example, we will round the repeating decimal 0.333… to 0.333.

Once you have found your decimal by dividing the numerator by the denominator, you will then multiply the decimal number by 100 to find your percent.

$$class\ success\ rate = \frac{\#\ of\ students\ who\ found\ their\ leaf}{\#\ of\ students\ in\ the\ class} \times 100$$

Example: 0.333 x 100 = 33.3%, which means the rate of success in finding the correct leaves is 33.3%.

Next you will make a table to summarize the outcome of this game. This table will help you to decide which traits are the best ones to use when identifying leaves.

- Discuss the characteristics that the students used to remember their leaves.
- Make a table like the one that follows. If a student used more than one characteristic, use the characteristic that they looked for first when they went up to the table.

Table 1. Data table for recording frequencies and successes of various traits in leaf identification.

Trait	Absolute Frequency	Relative Frequency	Number of Successes	Trait Success Rate
Shape				
Size				
Color				
Leaf Edge				
Damage				
Venation				

In the table you are asked to record the absolute and relative frequencies with which each trait was used in leaf identification. The **absolute frequency** of an event is simply number of times that the event occurs. For example, if you flip a coin seven times and the coin lands heads up twice, then the absolute frequency of heads is 2. The **relative frequency** of an event is the absolute frequency of the event divided by the total number of trials. Since you tossed the coin seven times in the example above, the total number of trials is 7. Therefore, the relative frequency of heads is $\frac{2}{7} \approx 0.29$. Note that the relative frequency is a fraction or a proportion.

To find the relative frequency of students using the different characteristics, you will use the total number of students in the class and compare that total to the numbers you used in your first bar graph for each characteristic. For example, if there are 28 students participating in the leaf activity, and 11 students out of the group of 28 students used leaf size to find their leaf, you have a ratio of 11:28. That ratio can also be written as a fraction with the numerator 11 and the denominator 28, or $\frac{11}{28}$. You then find the decimal equivalent of the fraction. You will want to use a calculator for this part of the activity.

Example: $\frac{11}{28} = 11 \div 28 = 0.392857\ldots$

- Since the decimal number in this example has many digits, round the decimal number to the thousandths place so that it is easier to work with.

Example: $0.392857 \approx 0.393$ ($\approx$ means *approximately* or *nearly equal to*)
Relative frequencies are related to percentages in the following way. If you multiply the decimal number by 100, this would give you the percent equivalent of your decimal number.

Example: $0.393 \times 100 = 39.3\%$

- Find the relative frequency of each characteristic for the group of students participating in the activity using the procedure described in the above example.
- Once you find the relative frequencies for each characteristic, find the sum of all your relative frequencies. The sum of all your relative frequencies should be very close to 1, or the sum may be exactly 1. (The sum may not be exactly 1 if you rounded some of the decimals as was done in the above example.)

Why should the sum be 1, or very close to 1? When you are working with relative frequencies, 1 indicates *1 whole*. Another way to think about this is to think about the relationship between relative frequency and percentages. A relative frequency of 1 is equal to 100%. If you eat 100% of a pizza, you eat the whole pizza. If you get a 100% on a math test, all of your answers were correct! Your relative frequencies should have a sum of 1 (or close) to indicate the whole group of students participating in the activity.

The relative frequency is a useful statistic because it can be used to compare the outcomes of experiments with different numbers of trials. For example, suppose that Mrs. Brooks' fifth grade class has 30 students, and Mrs. Tichnour's fifth grade class has 20 students. Both classes take the same science exam. Seven students in Mrs. Brook's make an A+, while six students in Mrs. Tichnour's class make an A+. However, since Mrs. Tichnour's class had fewer students than Mrs. Brook's class, this does not necessarily mean that Mrs. Brooks' class did better on the exam than Mrs. Tichnour's class.

Q1. Calculate the relative frequency of an A+ in each class. Which class did better on the exam?

- In your table, record the number of successes that belong to each trait.

Q2. Which trait(s) yielded the most successes? Are the traits with more successes necessarily the best ones to use when identifying leaves? Why or why not? How could we change the game to make the number of successes that belong to a trait a better measure of how effectively a trait can be used to identify leaves?

- Calculate the success rate of each trait to see how the success rate varied with the trait used for identification. Record the trait specific success rates in your table.

$$trait\ success\ rate = \frac{\#\ of\ successes\ that\ belong\ to\ trait}{absolute\ frequency\ of\ trait} \times 100$$

Q3. Which trait(s) yielded the highest success rates? Are the traits with higher success rates necessarily the best ones to use when identifying leaves? Why or why not? How could we change the game to make the trait success rate a better measure of how effectively a trait can be used to identify leaves?

- You are now going to use the data your teacher collected to make a bar graph indicating the leaf characteristics students used to remember their leaves.
- Your bar graph will display a visual comparison of the datawhich should be understood by anyone who might look at the graph. For example, someone should be able to look at the graph and be able to answer the following questions:
  - Which characteristics were used by more students to find their leaves?
  - Which characteristics were not used by many students?

- Now, **you** should answer the following questions, based on your class data:
    - Which characteristic was used by the *most* students?
    - How many students used that characteristic to find their leaves?
- You will use this number to determine the range of your vertical scale for your graph.
- Your vertical scale will begin at 0 and increase at equal intervals as you go up the vertical axis. Your largest number will be greater than or equal to the number of students in the group of the 'most-used' characteristic.
- Determine your interval size so that your graph clearly displays the data and the related labeling.

For example, suppose 17 students used leaf *shape* as the primary characteristic to find their leaves. You might decide to count by 2's on your vertical scale, and since the largest number you need to show is 17, your last number on the vertical scale might be 18. There is really no need to go much beyond 18 on the graph since your largest number is 17. If the largest number you need to show is 18 and you are using an interval of 2, you can stop at 18 or you might go one interval beyond 18 and stop at 20.

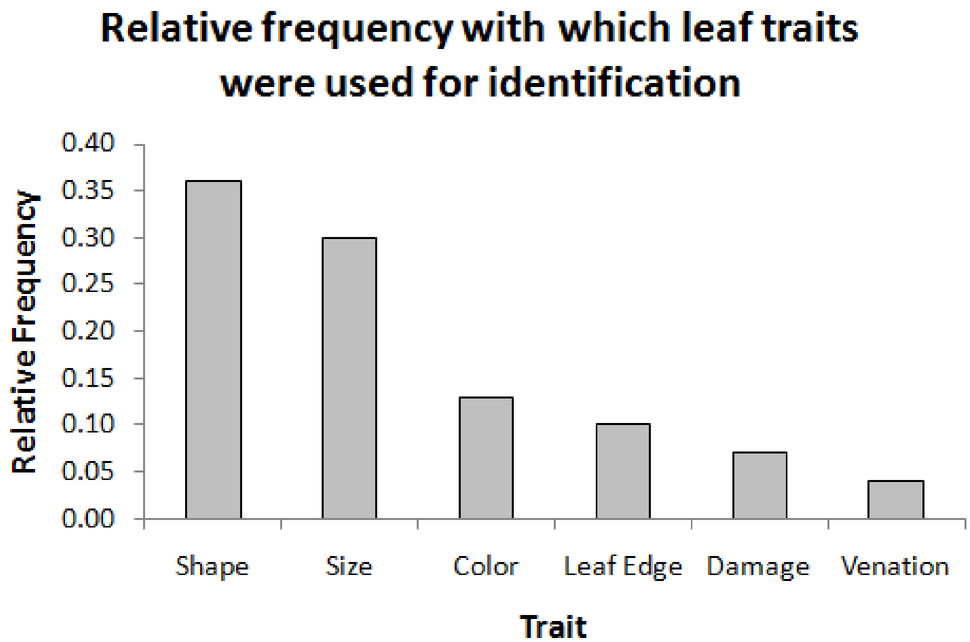
- Whatever interval you choose, be sure your graph is large enough to clearly indicate comparisons of your data, but also fit on your paper.
- Look at the examples in Figure 1 on page 8.
- You will also make a second graph using the same data set, but this time you are going to compare the relative frequencies of students using certain characteristics rather than the numbers of students.
- Using the relative frequency data you have just found, construct a bar graph of your data. The primary difference in this graph and your first graph will be your vertical scale, which will indicate relative frequencies rather than numbers of students as in your first graph.

What was the largest relative frequency of any one characteristic? That largest relative frequency tells you how far up your vertical scale should go. For example, if your largest percentage is 0.393, you will not need to go any higher (or much higher) than that number on your vertical scale (Example: if your interval is 0.10, then you might stop at 0.40 if the largest relative frequency is 0.393).

- Determine your intervals for your vertical scale so that your graph is large enough to clearly indicate comparisons of your data, but also fit on your paper.

- Look at the examples in Figure 1.

Figure 1: Bar graph comparing the relative frequencies with which various characteristics were used to distinguish leaves.

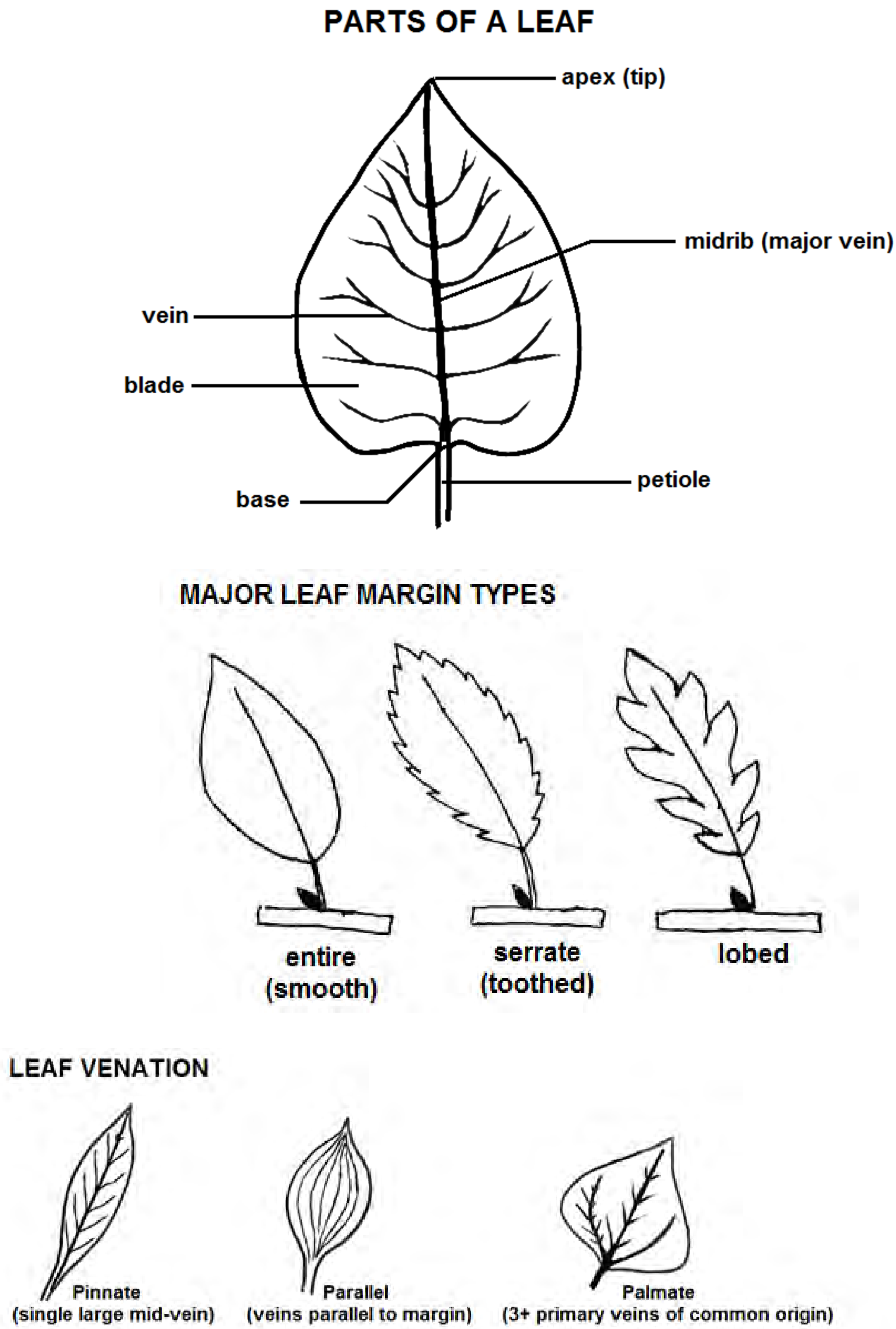


- Use Figure 2 on the next page to identify other leaf traits that might vary among species and individuals.

- Have each student pick a new leaf and play the game again.

- Calculate the new class success rate to see if it increased.

Figure 2: Leaf characteristics



Exercise 2: Leaf Match (*Grades K-12*)

Materials: Leaves from container A
Picture guide to tree leaves (found near the end of this book)

How to play: The leaf guide has pictures of all of the leaves that may be found in container A of this unit. Your goal is to find the picture that matches each leaf. Below each picture you will find the common name and the scientific name of the tree species from which the leaf came. The scientific name consists of two parts: the genus (close relatives will all have this name) and the species (only individual trees that might possibly interbreed share this name).

- Match each leaf from container A to a picture in the leaf guide!

Exercise 3: Using Mathematics to Describe a Population (*Grades 3-12*)

When scientists study traits within populations, they are generally interested in knowing several **parameters** about the traits, such as the "most typical" value of the traits within those populations, or how much overall variation is present in the traits of interest in those populations. One way to do this is by measuring or categorizing the trait(s) of interest for each individual within the populations. Unfortunately many populations are so large, or difficult to access, that it would be impossible to examine and describe each individual in the population. For example, imagine trying to measure the length of every ant in an ant hill, or the weight of every cardinal that passes through your yard. Since scientists often can't describe every individual in a population, they settle for describing a subset of the population. The set of individuals that are described is called the **sample**. Then, the trait(s) of each individual within the sample can be described by numbers or categories, and this information can be used to calculate **statistics** that summarize the data and help make inferences about the population as a whole. Scientists use statistics to describe the population sample, and then, they make a very important assumption. They assume that the sample is representative of the population as a whole. In order for this assumption to be valid, scientists must be very careful when they select the sample individuals.

IMPORTANT NOTES TO REMEMBER: Parameters are characteristics of **populations**, and are often of interest to scientists. However, knowing the exact parameters of a population is often impossible due to population size (or availability of time and/or funding!), so scientists *estimate* population parameters from **sample statistics**. For sample statistics to be more accurate estimates of population parameters, it is important that the sample(s) used are **representative** of the population(s) from which it is/they are taken!

Q1. Suppose that you wanted to find out what brand of pants are most often worn by fifth grade boys in a small town. Which of the following samples do you think would best answer this question? Why? What is wrong with the other samples?

a) The boys in Mrs. Brooke's fifth grade class
b) The boys whose last names begin with the letters A-D or M-P.
c) The boys that ride bus 23
d) The boys that sit together in the far right corner of the cafeteria

Q2. What if instead you wanted to know the average height of a fourth grade boy? Could any of the samples that you rejected before be used to answer this new question?

In the following exercises, you will collect data on the leaves provided in this unit, and learn how scientists use data from samples, and the mathematics that they use to estimate parameters of entire populations based on those samples.

For exercises 3a & 3b, you will use all of the leaves in Container A just to get you familiar with these mathematical operations and terms. (**NOTE:** The leaves in Container A come from many different tree species, and do not particularly constitute a good representative sample of a particular population, though conducting this exercise will give you a good understanding of the procedures used in determining sample statistics as estimates of population parameters.)

In Exercise 3c, you will learn how graphs can illustrate the measures of central tendency and spread of a data set (and thus provide an idea about the distribution of a trait within the population reflected by the sample).

Finally, in Exercise 3d, you will use the knowledge gathered from Exercises 3a and 3b to learn how to scientifically compare two populations based on samples drawn from two populations that more accurately reflect representative samples from those populations.

- Before starting the following exercises, decide, as a class, a particular trait to measure on your leaves. For example, you might measure the width of the leaf blade at its widest point, or the length of the blade from its tip to its base at the attachment of the petiole (Figure 1 from Exercise 3a.) You could also make a rough estimate of the leaf's blade area by multiplying the leaf blade width times its length. ***NOTE: Petiole length may not a good measurement to use, as some of the petioles of your leaves may have been cut or broken, and thus may not accurately reflect actual petiole lengths of the leaves. However, if you wish to use petiole length just to get familiar with the calculation of sample statistics, that is okay!*** You will use this trait to calculate the measures of central tendency and spread for that trait throughout Exercises 3a-3c.

- Have each student select a leaf and take the measurements that were chosen in step 1. Round off each measurement to the nearest centimeter.

- Record all of the measurements, and then order you data in a list from least to greatest. See List 1 on page 14 for an example of ordered tree leaf width data.

Exercise 3a: Measures of central tendency (*Grades 3-12*)

Exercise 3a.1: Calculating measures of central tendency (*Grades 3-12*)

Trait values vary among individuals, but in a sample of individuals, some values are more typical than others. For instance, during the summer, tree leaves are usually green, while in the fall when leaves are changing colors, some trees, such as willows, have leaves that are primarily yellow in color, while other trees, such as oaks, have leaves that are primarily brown in color. According to our visual observations, these are the most common leaf colors for willows and oaks in the fall. However, someone else might disagree with us and say that oak leaves are primarily red in the fall. Based on observation alone, there's no way to know for certain who is correct. If, however, we can measure the value of a trait, then we can assign a numerical value to each individual in our sample and use mathematical statistics called **measures of central tendency**, to decide which value is the most typical within the population as a whole.

- Using the data that you have already collected on the leaves from Container A, you will now calculate three measures of central tendency of this "sample".
- Once you have ordered your measurements you can find the median of your data set. The **median** of a data set is the value that is in the middle when the data is ordered from least to greatest. However, if you have an even number of data entries, there will be *two* numbers in the middle. In this case, the median is the average of the two numbers in the middle. The median of the example leaf width data is $\frac{6+6}{2} = 6$.
- Draw a table with two columns on the board. Label the first column 'Values Observed' and the second column 'Number of Observations'. In the first column list the values that appear in your ordered list from least to greatest, eliminating any repeat values. In the second column list the number of times that each value from the first column appears in your ordered list. Title your table appropriately. See Table 2 for an example.

List 1. Ordered Leaf Width Measurements

Leaf widths (cm)
3
3
3
5
5
5
5
5
5
5
5
5
5
6
6
6
6
8
8
9
9
9
9
9
9
9
9
9
9
19

Table 1. Observed Leaf Widths

Value	Absolute Frequency
3	3
5	10
6	4
8	2
9	10
19	1

- Now you are ready to find the mode of your data set. The **mode** is the value that appears the most often. There are two modes for the example leaf width data set, because 5 cm and 9 cm both appear 10 times, more times than any other data values. The mode is another measure of central tendency.

- You can also find the mean of your data set. The **mean** of a data set is the average of the data entries. So, to find the mean of your data set, add all of your data entries together and then divide by the total number of data entries. You can save time by using your table to help you compute the mean. Instead of adding up all of the numbers individually like this:

$$3+3+3+5+5+5+5+5+5+5+5+5+5+\cdots$$

You can add them up in groups, like this:

$$\mathbf{3(3)+10(5)+4(6)+2(8)+10(9)+1(19)=9+50+24+16+90+19=208}$$

In the expression above, the numbers in parentheses are observed values from Table 1, and the numbers before the parentheses are the number of times those values appear in List 1. Since there are 30 data entries in all, the mean of this data set (rounded to the nearest centimeter) is

$$\frac{208}{30}=6.93\text{ cm }\approx 7\text{cm}$$

The mean, like the mode and the median, is also a measure of central tendency. Since there are multiple measures of central tendency, it is up to you to decide which measure (the median, the mode, or the mean) best represents the most typical trait value. In the example leaf width data set, the modes are probably the best measure of central tendency because over half of the leaves in this sample have leaf widths that are equal to one of the two modes.

Q3: Compute the mean, median, and mode for each of the following data sets. Which measure of central tendency best describes the data? Why?

a)

Value	Absolute Frequency
3	2
5	2
6	5
8	2
9	1

b)

Value	Absolute Frequency
5	3
6	4
8	2
25	1

Exercise 3a.2: Sigma notation (*Grades 9-12*)

At some point, you are likely to run across an expression for calculating the mean of a set of data (as well as other sample statistics or other quantities) which uses the symbol **Σ**, which is the upper-case Greek letter **sigma**. Don't be intimidated by this perhaps unfamiliar symbol, however, as its meaning in mathematical expressions is actually quite simple. In mathematical expressions, the upper-case Greek sigma (Σ) simply means to "sum up" some quantity. For this reason, **sigma notation** is also sometimes known as **summation notation**. To get you a little more comfortable with sigma notation, let's first look at a few examples.

$$\sum_{i=1}^{5} i$$

Look at the example above. We already know, since we see a Σ, that we will be summing up some values. But what values are we summing up? To know the answer to this, we have to look at the other parts of the notation. For example, note the "i" to the right of the Σ. This tells us that we will be adding together values of *i*. However, *i* could take on any number of values, unless it has been predefined, or we are given some sort of criteria for which values of *i* we should add up. This is where the expressions below and above the Σ come into play.

Note the expression "*i*=1" below the Σ. Usually in sigma notation, you will see an expression like this below the sigma. In this expression, the "*i*=" stands for the "**index of summation**," which is just a fancy way of saying with which number you should start. In this case, we will start with a value of 1 for *i*. Though sigma notation often uses a lower bound of 1, any other integer (or zero) can be used as this starting point. Now note the "5" on top of the Σ. This tells us the upper bound of the summation, or with which value we should stop. In other words, the sigma notation above tells us to add up the integer values of *i* from 1 to 5:

$$\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5 = 15$$

To further illustrate how sigma notation is used, let's try another example:

$$\sum_{i=3}^{7} (i^2 - 1)$$

In the above example, the sigma notation is telling us to sum the quantity (i^2-1) for all integer values of *i* between 3 and 7:

$$\sum_{i=3}^{7} (i^2 - 1) = (3^2 - 1) + (4^2 - 1) + (5^2 - 1) + (6^2 - 1) + (7^2 - 1)$$

$$\sum_{i=3}^{7} (i^2 - 1) = (9 - 1) + (16 - 1) + (25 - 1) + (36 - 1) + (49 - 1)$$

$$\sum_{i=3}^{7} (i^2 - 1) = 8 + 15 + 24 + 35 + 47 = 129$$

To check your understanding of sigma notation, try to see if you can answer the following questions:

Q4: Evaluate the following expression:

$$\sum_{i=6}^{9}(17-2i)$$

Q5: Evaluate the following expression:

$$\sum_{i=0}^{4}2^{i}$$

Q6: Evaluate the following expression:

$$\sum_{i=3}^{6}(-1)^{i+1}(3i+1)$$

Q7 (Super Solver Problem): How could you write "the sum of all multiples of 3 from 6 to 27" in sigma notation?

Check your answers in the answer section at the end of this workbook before moving on!

Sigma notation can also be used to denote items in a list. For example, when calculating the mean of a data set of measurements of a particular quantity (*x*), one must first add up all of the values in the data set:

$$\sum_{i=1}^{n}x_i$$

Notice that in this expression, there is a new convention used here. The *x* to the right of the Σ lets us know that we will be adding up values of *x*, our variable of interest in our data set. However, also note that the *x* has a subscript, *i*. This means that we will be adding up the "*i*th" values of *x* between our lower and upper bounds. Our lower bound (the index of summation) in this case is "*i*=1", so we will start with the *first* value of *x* in our data set. The upper bound in this case is *n*, which represents our sample size. In other words, this notation is telling us that we add up every "*i*th" value of *x* in our data set, from the first value, all the way up to the *n*th value, which would be the final value in a sample size of *n*. Or, to explain this mathematically:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 \ldots + x_n$$

Remember, however, to calculate the sample mean (usually expressed as $\bar{x}$, read as "x-bar"), we have to divide the total of the sum of all of our values by the sample size. Thus, the equation for calculating the sample mean in sigma notation would be given as follows:

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

Essentially, this notation tells us to take the sum of each of our data values in our sample, just as before, but the $\frac{1}{n}$ in front of the Σ lets us know that we have to multiply the resulting sum by $\frac{1}{n}$ (which is the same thing as dividing by *n*, or our sample size).

Exercise 3b: Measures of spread

Exercise 3b.1: Calculating measures of spread (*Grades 5-12*)

In addition to determining the most typical value of a trait, we are also interested in knowing how much variation there is within a group of individuals. Mathematical statistics called **measures of spread** can help us determine how much variation there is within a sample (and thus use that information to estimate the amount of variation present in the entire population from which the sample was drawn).

- Compute the range of your data set. The **range** is the *largest data entry minus the smallest data entry*. For example, the range of the data from List 1 is $19 - 3 = 16$cm. The range is a measure of spread, indicating the difference in size between the largest and smallest individuals in a sample. Larger values of the range indicate more spread in the data.

Another measure of spread of data is the **variance**. Variance is slightly more complicated than the range, but is still essentially a way of quantifying the variability present in a population or sample. It is important to note that there are two ways of calculating the variance of a population or sample. They are very similar, with one small exception. First, let's learn how you would calculate the variance of a population, if the population were sufficiently small enough to gather data on every individual within it. For simplicity, first assume that the "sample" of leaves in Container A actually represents a sample of every individual in the simulated "population". You are now going to calculate the **population variance** (usually represented as σ^2, read as "sigma squared", since the expression uses lower case Greek letter "sigma" squared) for your leaf widths using these leaves. Follow the steps below to see how the population variance is calculated.

- First subtract each leaf width from the mean value. If you have 30 leaf widths, you will subtract each of the 30 leaf widths from the mean.
- Find the square of each subtraction. For example, if one of your leaf widths was 4 and the mean leaf width was 8 cm, you subtract 4 from 8:

$$8 - 4 = 4$$

- Now square this value (multiply it by itself). For example, if you square 4:

$$4^2 = 4 \text{ x } 4 = 16$$

- Do this same procedure for each of your measurements.
- When you have found all your subtraction values, add all those values together.
- When you have your sum of subtraction values, divide that sum by the total number of measurements. If there are 30 leaf widths in your data set, then you will divide the sum you found by 30.

You have just learned how to calculate the *<u>population variance</u>* for a particular trait. The larger the value for variance as it compares to the mean of your data, the greater the variability in that trait among the population. Below is an example of calculating the population variance of leaf widths in the example data from List 1 on page 14.

$$\frac{3(3-7)^2 + 10(5-7)^2 + 4(6-7)^2 + 2(8-7)^2 + 10(9-7)^2 + 1(19-7)^2}{30}$$

$$= \frac{3(-4)^2 + 10(-2)^2 + 4(-1)^2 + 2(1)^2 + 10(2)^2 + 1(12)^2}{30}$$

$$= \frac{3(16) + 10(4) + 4(1) + 2(1) + 10(4) + 1(144)}{30}$$

$$= \frac{48 + 40 + 4 + 2 + 40 + 144}{30} = \frac{278}{30} \approx 9$$

- In the previous steps, we computed the population variance of the leaves in Container A, under the pretense that the leaves in Container A reflected *all* of the individuals in a population. However, as mentioned before, it is almost always impossible for scientists to collect data on every individual within a population, and they usually must settle for collecting data on a sample from that population in order to estimate the typical values and variability within the entire population. For example, you could think of all of the leaves in your class's Container A as a sample of the population consisting of all leaves used in every existing copy of *Biology in a Box* Unit 8. As it turns out, the method of calculating variance as presented earlier is not always a very good estimate of the *true population variance* if one is using a sample that does not consist of the entire population. To estimate the population variance from the sample data, the **sample variance** (usually represented as $\boldsymbol{\sigma}$, the lower case Greek letter sigma) is used. To calculate the sample variance, proceed as if you were calculating the population variance, but in the final step divide by the ***sample size minus 1***. For example, the sample variance of the leaf width data (which we would use to estimate the variance of the population from which the leaf width data was drawn) would be

$$\frac{278}{29} \approx 10$$

The **standard deviation** is simply the square root of the variance (at both the population and sample levels). For example, the **population standard deviation** (represented by $\boldsymbol{\sigma}$, the lower case Greek letter "sigma") is the square root of the population variance (calculated by dividing the sum of the squared differences from the mean by *N*, or the population size), and the **sample standard deviation** (usually represented by $\boldsymbol{s}$) is the square root of the sample variance (calculated by dividing the sum of the squared differences from the mean by the quantity of one less than the sample size, or *n*-1). For example, both the "population" standard deviation and "sample" standard deviation of the leaf width data are approximately

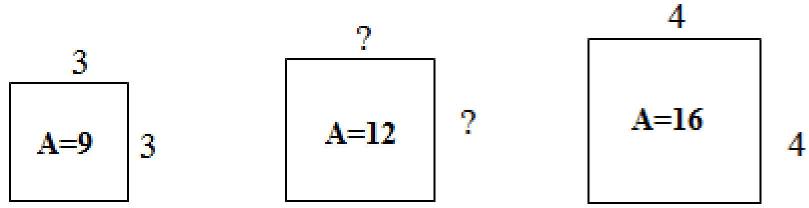
3 (if rounding to the nearest cm, though the sample variance is actually larger, and likely a better estimate of the *true* population variance).

Like the range, the variance and the standard deviation are also measures of spread. In particular, they measure how your data is spread about the mean. Larger values of the sample variance and standard deviation indicate more spread in the data.

Q8 (Super Solver Problem): If you completed Exercise 3a.2, how would you write the formulas for population variance and sample variance using sigma (summation) notation?

Mini-challenge: Estimating square roots (*Grades 5-12*)

- Look at the three squares below and notice the side lengths that are labeled on the first and last squares.



The first square has a side length of 3, and the last square has a side length of 4. The areas of all three squares are also labeled. The first square has an area of 9 because 3 x 3 = 9, and the last square has an area of 16 because 4 x 4 = 16. What will the side lengths be for the middle square if the area of the square is 12? Remember that the definition of a square tells us all sides must have equal length.

The square in the middle does not have its side length labeled so you are going to determine what the side length must be to have an area of 12. If you know the smallest square has a side length of 3 and the largest square has a side length of 4, what can you say about the length of the middle square? How does the size of the middle square relate to the other two squares? Both 3 and 4 are whole numbers, but will the missing number be a whole number? Why, or why not?

How do you think you might find the side length of that middle square? Use your calculator to help you find the missing side length, but do not use the square root key (*that would be cheating!*). Use your estimation skills to help you begin choosing side lengths to test to find the actual side length. When you believe you have the side length, test your answer by multiplying the number by itself (squaring the number). How close is the square of your number to 12? Continue testing numbers until you have a number to the hundredths place that gets as close as possible to 12 when squared.

Exercise 3b.2: Sample variance vs population variance: What a difference "1" makes (*Grades 9-12*)

This exercise is modified from an exercise presented by Dr. Harley Weston from the Department of Mathematics and Statistics at the University of Regina.

In the previous exercise, you learned how to calculate various measures of spread within a sample, which are used to make inferences about the variation present in the entire population from which the sample is taken. You also learned that there is a slight difference in calculating the variance of a sample versus the variance of a population (which can only be calculated if every individual within the population has been measured), in that the denominator in the equation for sample variance is equal to (n-1), where n is the number of individuals in the sample, as opposed to a denominator of N (total number of individuals in the population) when calculating population variance (which, if you remember, can only be done if you could measure all individuals in the population). This exercise will illustrate why a denominator of (n-1) is used for calculating the sample variance as an estimate of the true population variance.

For this exercise, your teacher should locate the container marked "Exercise 3b.2," which contains a bag of beads of the same size and shape, but of two different colors. All of the beads in this bag represent a population, with each bead representing an individual within the population. Your teacher knows the total number of beads, as well as the number of beads of each color in the bag, but will not reveal this information to the class until you have all conducted the rest of the exercise.

NOTE FOR TEACHERS: Note that there is a greater proportion of beads of one color in the bag compared to the other. Assign this color bead a value of 1, and the beads of the less common color a value of 5, and let students know the value represented by each color before continuing with the steps below.

- Your teacher will shake up the beads in the bag, and then approach a student in the class.

- That student should reach into the bag, and withdraw a sample of 3 beads (n=3), recording on a piece of paper the color of each, and the value assigned to each of the two colors.
- Those beads should be placed back into the bag, and the bag shaken again.
- This procedure should be repeated until every student has taken a sample of 3 individuals from the "population".
- Each student should now make a table like the one below, and calculate the appropriate values for their sample, rounding their calculated values to the nearest hundredth.

Sample Values			**Sample Mean**	**Sample Variance (s^2) Using *n* as the denominator**	**Sample Variance (s^2) Using (*n-1*) as the denominator**

- After all students have calculated their values for the means and variances (using both methods) for their samples, your teacher should make a table with three columns ("Mean", "Variance Using n", and "Variance Using (n-1)") on the board, and compile students' results.

NOTE TO TEACHERS: There will only be four possible values for sample means, and two possible values for each column of variances when using a sample size of 3, so you may just wish to record each of these values as they are presented by a student obtaining those results, and use tally marks to keep track of how many samples resulted in those values under each of the columns.

- After all students' results are compiled on the board, you should calculate the average variance for all samples when using both methods (n versus n-1 in the denominator).
- Your teacher will now reveal the true population variance (σ^2) to the class.
- Compare this value to the average sample variances obtained when using both n and (n-1) in the denominator.
- Now answer the following question:

Q9: What do you notice about the average values of sample variance (s^2) using n versus (n-1) in the denominator of the equation when compared to the true population variance (σ^2)?

As a further exercise relating back to measures of central tendency, you may also wish to present students with the true population mean (μ), and have them consider the following:

- Look at the values under the "Mean" column on the board.
- Using the tally marks representing the number of samples in which each mean was calculated for the samples, and calculate the mean of all of the sample means.
- Compare this value to the true population mean (μ), and answer the following questions:

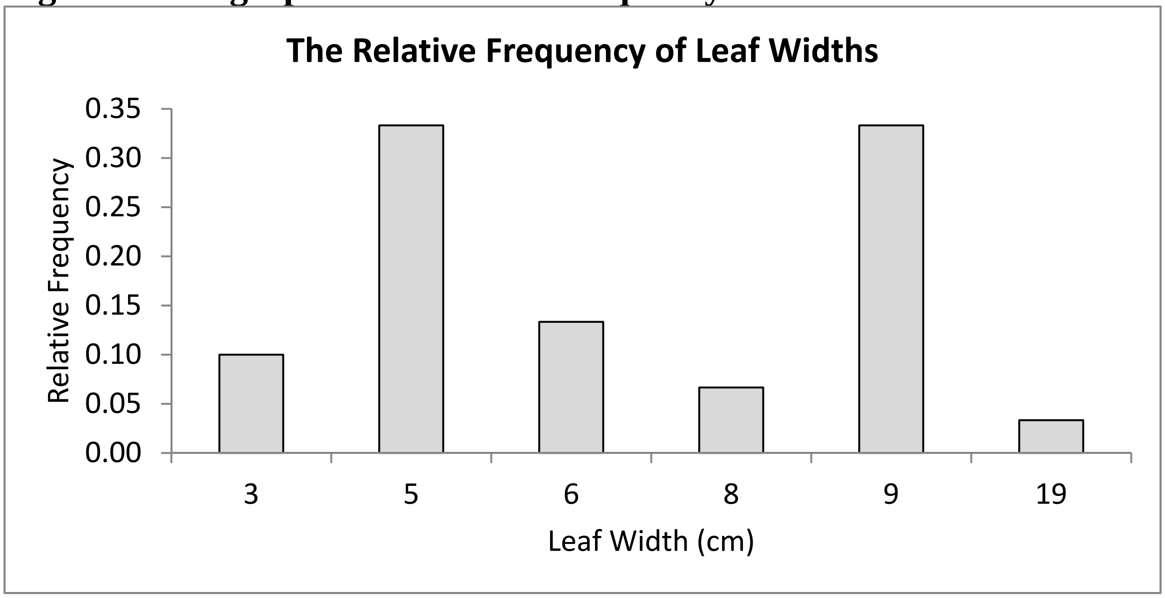
Q10: In effect, what are you doing when you are calculating the mean of all students' sample means?

Q11: Compare your calculated value of the mean of all students' sample means, as well as each of the individual sample means obtained, to the true population mean. Which value is closest to the true population mean? Considering your answer to **Q10**, what does this tell you?

Exercise 3c: Visualizing measures of central tendency and spread (*Grades 5-12*)

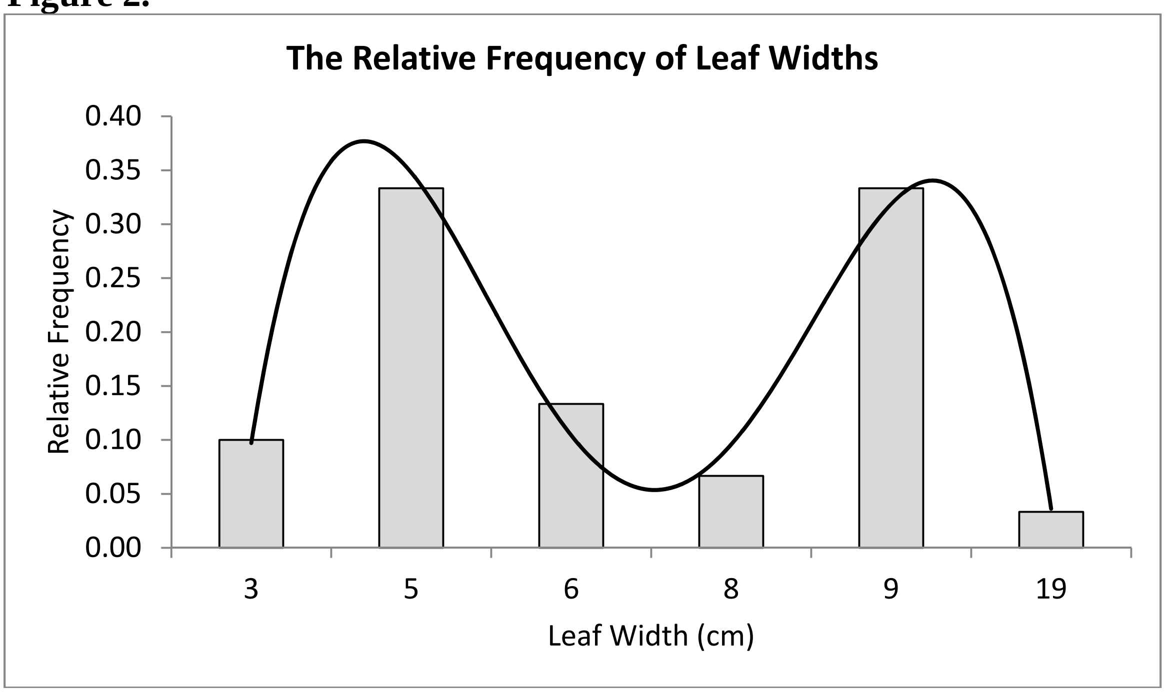
A **relative frequency bar graph** can help you to visualize how your data is distributed. For example, by looking at a bar graph, you can see how spread out the data is, and estimate the central tendency of the data. You can also use a relative frequency bar graph to find the probability that a data entry belongs to a particular category. Figure 1 is a relative frequency bar graph of the leaf width data. The leaf widths are listed along the horizontal or x-axis of the graph. The relative frequency of each leaf width is plotted along the y-axis. Figure 1 on the following page clearly shows that this data is **bimodal** (has two modes) about which the data values are clustered. We can also see that the relative frequency of 3 cm is 0.1. This means that if we select a leaf at random from our sample, then probability that the leaf's width is 3 cm is 0.1.

Figure 1: Bar graph of the relative frequency of leaf widths



Now imagine that we wanted to draw a smooth curve to describe the shape of our data. In the case of the leaf width bar graph the curve might look like the one pictured in Figure 2 below.

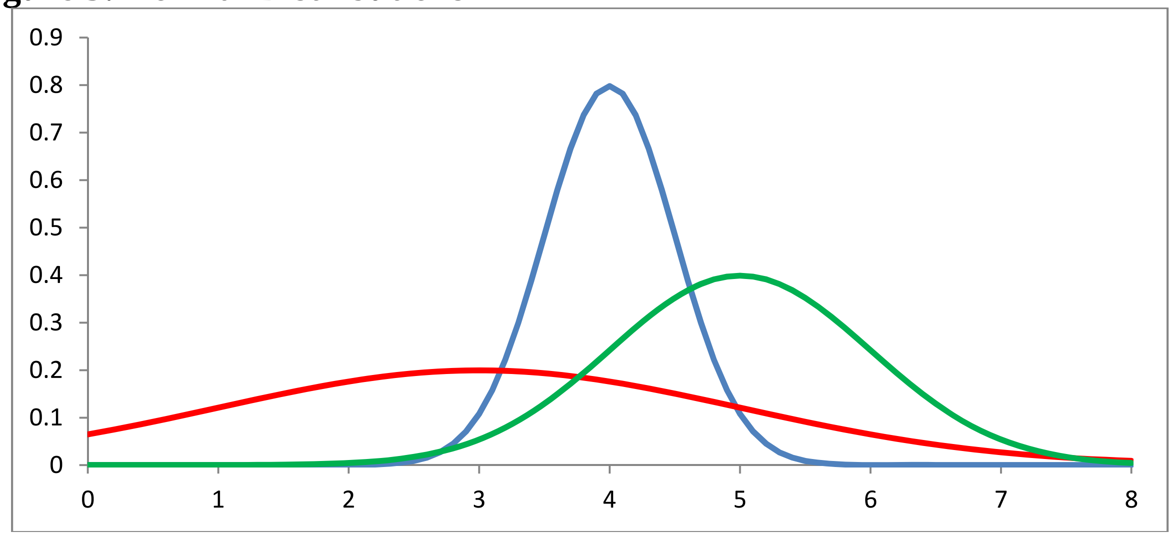
Figure 2.



Often scientists use smooth curves called **probability distributions** to describe the shape of a bar graph. One type of probability distribution is called a **normal**

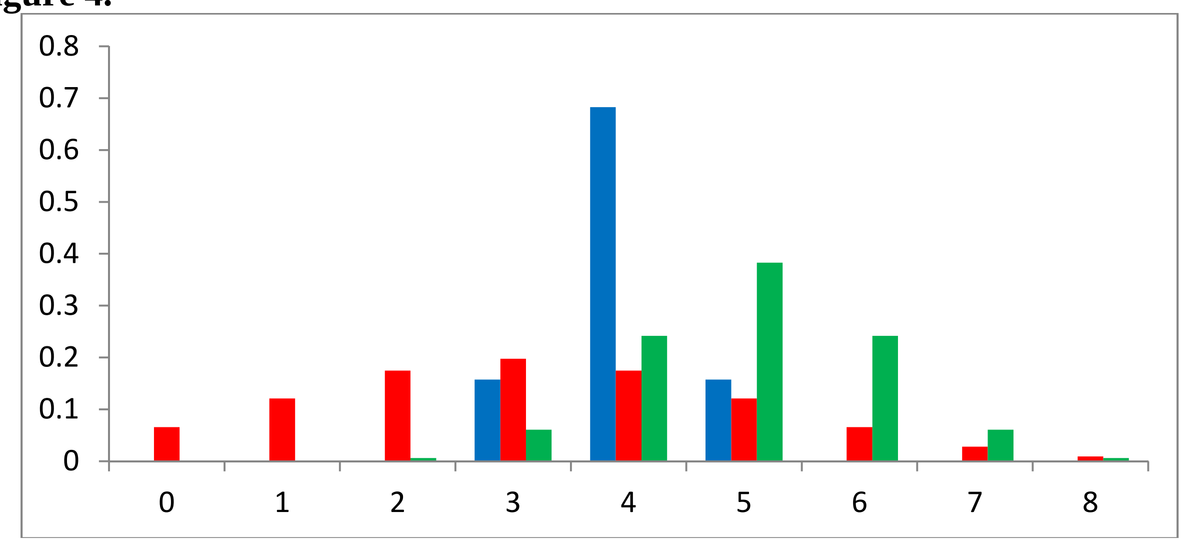
distribution. Normal distributions are very important because, as the name implies, they are considered to be the most standard type of distribution. When scientists examine data, one of the first questions they ask is whether the data are normally distributed. Figure 3 shows a plot of several normal distributions. Note that although these distributions vary, they all have the same basic shape.

Figure 3: Normal Distributions



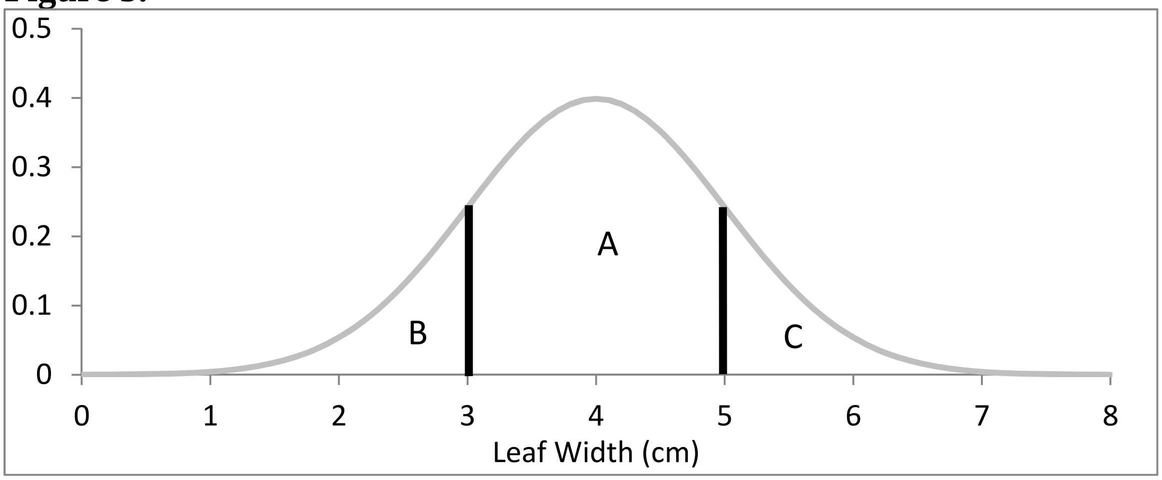
Q12: Figure 4 below shows three data sets that are normally distributed. How do their distributions differ? Do they have the same mean? Do they have the same variance? Which distribution has the greatest mean? Which distribution has the smallest variance?

Figure 4.



Probability distributions can also be used to find the probability that an individual has a range of trait values. The probability that an individual's trait value falls within an interval is equal to the area of the region bounded by the interval on the x-axis and the probability distribution curve. For example, assume that a trait is distributed according to the normal distribution pictured in Figure 5 below. The probability that an individual in the population has a trait value between 3 and 5 is equal to the area bounded by the two black line segments and the gray probability distribution curve (region A).

Figure 5.



Q13: How can you interpret the area of region C (in Figure 5) in terms of probabilities?

Q14: Now suppose that we select a leaf at random from the population. Use Figure 5 to rank the following events from most to least probable.

- **a)** The leaf's width is between 0 cm and 3 cm.
- **b)** The leaf's width is between 3 cm and 8 cm.
- **c)** The leaf's width is between 3 cm and 5 cm.

We can also use graphs to see if the data are skewed. We will say that the data are skewed to the right if the right tail of the graph is longer than the left tail of the graph (see Figure 6 for an example). Similarly, we will say that the data are skewed to the left if the left tail of the graph is longer than the right tail of the graph (see Figure 7).

Figure 6

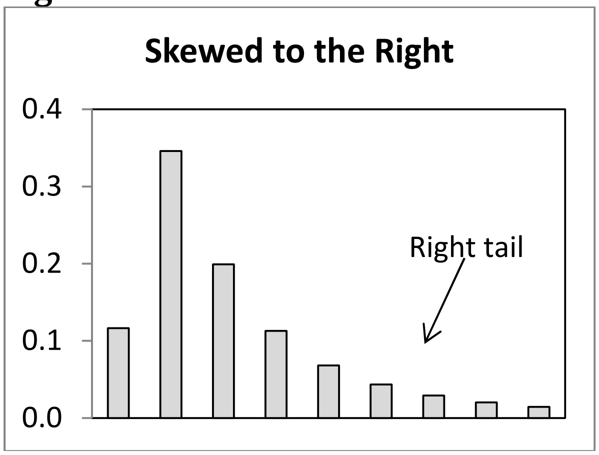
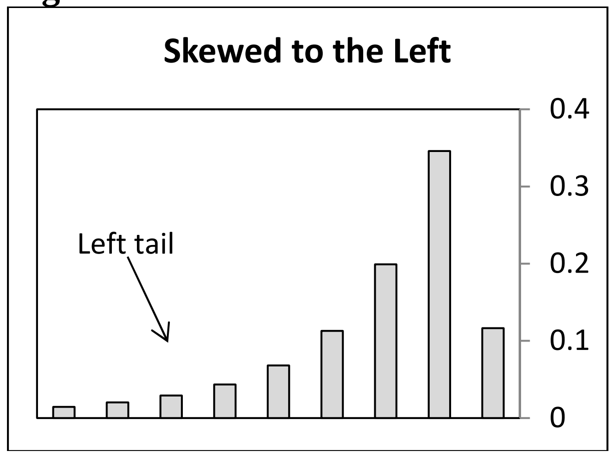


Figure 7



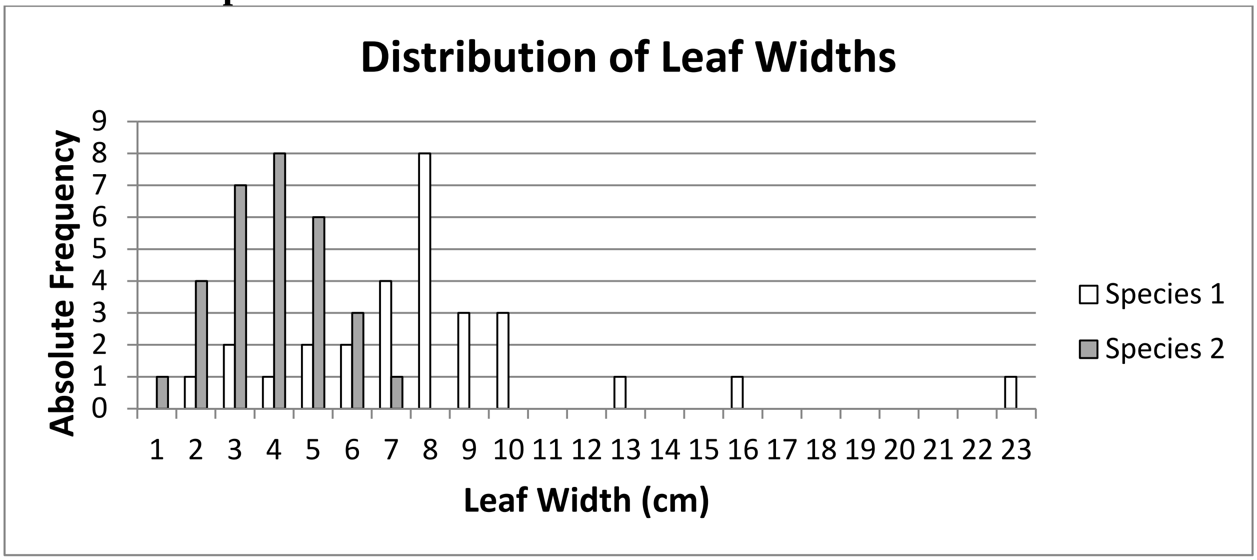
If a data set is normally distributed then it is not skewed, and it has only one mode, which is equal to its mean and its median.

- Make a bar graph of your leaf data. Is your graph skewed to the left or the right? Are your data normally distributed?

Exercise 3d: Comparing leaves from two samples (*Grades 9-12*)

- Find Container B, which contains two samples of leaves that were collected from two different species of trees or shrubs.
- Separate the leaves according to which sample they belong to (species A or species B).
- Have each student select one leaf from each sample and measure the selected trait.
- For each sample, order the measurements and construct a frequency table as you did in Exercise 3a. Compute the mean, median, mode, range, and variance for both samples and make a double bar graph to help you compare how the different samples are distributed.

Figure 8: A bar graph showing the distribution of leaf width data sampled from two species.



Q15: Examine Figure 8. Which species appears to have narrower leaves? Which species appears to have a broader leaf width distribution? Check your answers under Exercise 2a in the answer book.

Q16: Use your bar graph to answer the following questions about species A and species B. Do you think that one of these trees has larger leaves on average? Why or why not? Do you think that one of these trees has a broader distribution of leaf widths? Why or why not? Now use your statistics to answer the same questions again. Did your answers change?

- Add the samples' means, medians, and modes to the label of the x-axis of your graph.

Comparing populations through the use of samples

Often scientists compare populations to see if they are different. In particular, scientists are often interested in comparing the means of two populations. This is no easy task, because, as we have already mentioned, it is not usually possible to measure a population as a whole. Usually scientists are only able to measure a sample of the population. So, instead of comparing the population means scientists are forced to compare sample means. Because samples are incomplete and imperfect representations of the populations from which they are drawn, the sample mean is only an estimate of the population mean. Therefore, sometimes it happens that two samples may have different means, while the populations from which they were drawn have the same mean. Alternatively, the sample means may be the same, though the population means may actually be different.

Statistical Tests: Evaluating Hypotheses

You are probably already familiar with the scientific definition of the word **hypothesis** (an "educated guess" based on observations). In the scientific method, one of the steps is to evaluate or test one's hypothesis. One way that scientists do this is with **statistical tests**, which provide support to (or sometimes fail to support) their hypotheses. When testing these hypotheses statistically, however, one is actually comparing two different-statistical hypotheses: the **null hypothesis** and the **alternative hypothesis**. *However, it is important to note that statistical hypotheses are still slightly different from scientific hypotheses in that statistical hypotheses simply examine whether there is a difference between groups, or a pattern relating particular variables, but not the WHY behind them.*

Let's say, for example, that you think there is a difference in average height of American high school baseball players and high school soccer players.
The **alternative hypothesis** (often represented as $\mathbf{H_A}$ or $\mathbf{H_1}$) is simply the original hypothesis formulated by the scientist. For example, there is a difference between the average heights of American high school baseball players and high school soccer players. This would be your alternative hypothesis in this case.

The **null hypothesis** (often represented as $\mathbf{H_0}$) is essentially a hypothesis that is one that opposes your alternative hypothesis. In our example, the null hypothesis in this case would be that there is NO difference in the average heights of American high school baseball players and high school soccer players.

Null and alternative hypotheses can also have directionality. For example, if you think that baseball players are TALLER than soccer players, the hypotheses you would be testing would be as follows:

> $\mathbf{H_0}$**:** American high school baseball players are shorter than, or equal in height to American high school soccer players on average.
> $\mathbf{H_A}$**:** American high school baseball players are taller than American high school soccer players on average.

As we have already discussed, the populations of interest in this example (all American high school baseball players and soccer players) are way too large for you to measure the height of every player. So, you could collect data (ideally by randomly sampling) the populations. For example, you could travel to randomly determined high schools, and measure randomly selected players, and measure as many individuals as time and money would allow. You would then have a set of sample data, which you could use to statistically test your hypothesis.

In conducting statistical hypothesis tests, it is important to realize that what you are doing, ALWAYS, is really testing, based on the sample data, whether you should reject or "fail to reject" the NULL HYPOTHESIS. (*Scientists prefer to say that they "fail to reject" a null hypothesis when doing so is supported by their data, rather than saying they "accept" it, as accepting it implies that the null hypothesis is completely true. However, they know that this very well may not be the case. The null hypothesis may be false, but they may not have enough data to conclusively show that it is false!*)

There are many different types of statistical tests, each of which are used for various types of data, but they all follow the same basic idea:

1. Data are used to calculate a **test-statistic**.
2. The test-statistic is compared to a **critical value** (which depends on the size of the sample, as well as the degree of confidence one wishes to have regarding their decision).
3. The null hypothesis is either rejected or fails to be rejected, based on the comparison of the test statistic to the critical value.

So, how do scientists make informed decisions on population means based only on sample means? Such decisions are based around the concept of **significance**. In other words, if two sample means are different, how likely is it that the difference between those sample means actually reflects a difference in the means of the entire populations in question? Let's return to our example of the average heights of American high school baseball players and American high school soccer players, but first let's talk a little bit about the types of errors that can be made in statistical tests.

Remember, very often in science, it is impossible to be 100% certain about populations in which we are interested, simply because we cannot measure every individual in the populations. Because of this, there are two different types of errors that we could make based on our data:

1. We could incorrectly reject the null hypothesis when it is true. This is called a **Type I error**.
2. We could incorrectly fail to reject the null hypothesis when it is false, which is known as a **Type II error**.

Dealing with Type II errors is easily solved by gathering as much data as possible. If our sample sizes are large, we are more likely to detect real differences that exist between populations based on our samples. However, scientists deal with Type I errors by deciding how sure they want to be (usually 95%, sometimes 99%; remember, we can never have 100% certainty!) that they are correct in their decision to *reject* the null hypothesis. The value representing the remaining margin of uncertainty (5% or 1%, usually represented in decimal forms as probabilities of 0.05 or 0.01) is known as the **significance level** of the test. This is directly related to our earlier discussion of the meaning of significance, in that the more certain we wish to be that our decision to reject the null hypothesis is correct, the larger the differences between our samples from the populations of interest (relative to the variances) have to be in order for them to be considered **significant**.

If you find that, in your samples, the baseball players are an average of ¼" taller, is this really significant? Such a small difference may be accounted for by some of the soccer players slouching while being measured, or differences in the thickness of the soles of shoes of any of the players. Such a small difference could also be a matter of error in your measurements, as perhaps not all of your measurements were done exactly the same way. Thus, chance or error could account for such a small difference, and baseball players may not really be taller than soccer players after all! However, if your sample means indicated that baseball players are 6" taller than soccer players, this difference is likely to be a real one, because such a difference is likely too large to be accounted for by chance or error.

Just how large does a difference between sample means have to be to be significant? Well, this depends on how confident we want to be when making statements about the population means, based on our sample means, as well as the sizes of our samples. It is important to note that in science, one can never be 100% confident in the conclusions drawn from characteristics of samples. The only way we can be 100% confident about differences between the means of two different populations would be to measure the value of interest on every single individual in each of the populations we wish to compare, which is often impossible due to limitations of time, money, etc.

Let's return to our baseball/soccer player example. Again, if the difference in mean heights of each type of player was very small in our samples, our confidence in our ability to say that the population mean heights are different is lower than it would be if the difference in sample mean heights was fairly large. Also, let's say we only measured 10 baseball players and 10 soccer players. Those are fairly small samples. What if the sample of baseball players, just by chance, contained two very tall boys, that are much taller than the average baseball player, or our sample of soccer players contained one player much shorter than the average soccer player? Such a possibility also increases the chances that our results are not significant. However, if we measured 100 players of each sport, and obtained the same height difference in our sample means, we would be more confident that such a height difference between average American high school players of the two sports is significant. Nevertheless, it is true that the greater the difference between the sample means, the more likely it is that the population means are different as well.

However, the significance of the difference between sample means also depends on the variance of the samples. **In general, the larger the sample variances, the less likely it is that the difference in the sample means is significant.** In order to understand how the variance affects the significance of the difference in the means, we will look at an example.

Suppose that a scientist is studying how leaf width varies with rainfall in two plant species, fire bushes (FB) and maple trees (MT). He collects four sets of leaf samples representing the two woody plant species, each found at the same two localities that differed in annual rainfall. Both localities were suburban yards, but locality 1 (L1 in east Tennessee) receives 48 inches of rainfall a year and locality 2 (L2 in Fort Lauderdale, Florida), 64 inches of rainfall per year.

Q17: Assuming that a large number of leaves were collected from all of the trees at the respective localities, which of the following best describes the population to which MTL1 belongs?

a) All of the maple leaves in North America
b) All of the maple leaves from suburban yards
c) All of the maple leaves found at the locality L1
d) All of the maple leaves that receive approximately 48 inches of rainfall a year

Look at Figures 9 and 10 on the following page. Figure 9 shows the distribution of leaf widths in samples FBL1 and FBL2. Figure 10 shows the distribution of leaf widths in samples MTL1 and MTL2. The populations from which these samples were taken are known to be normally distributed. Also, the populations from which FBL1 and FBL2 were taken have the same variance, as do the populations from which MTL1 and MTL2 were taken.

Figure 9. Distributions of leaf widths from fire bush samples from areas of different annual rainfall.

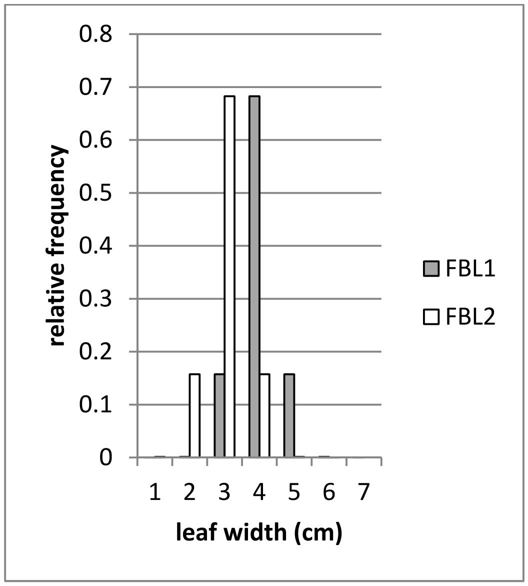
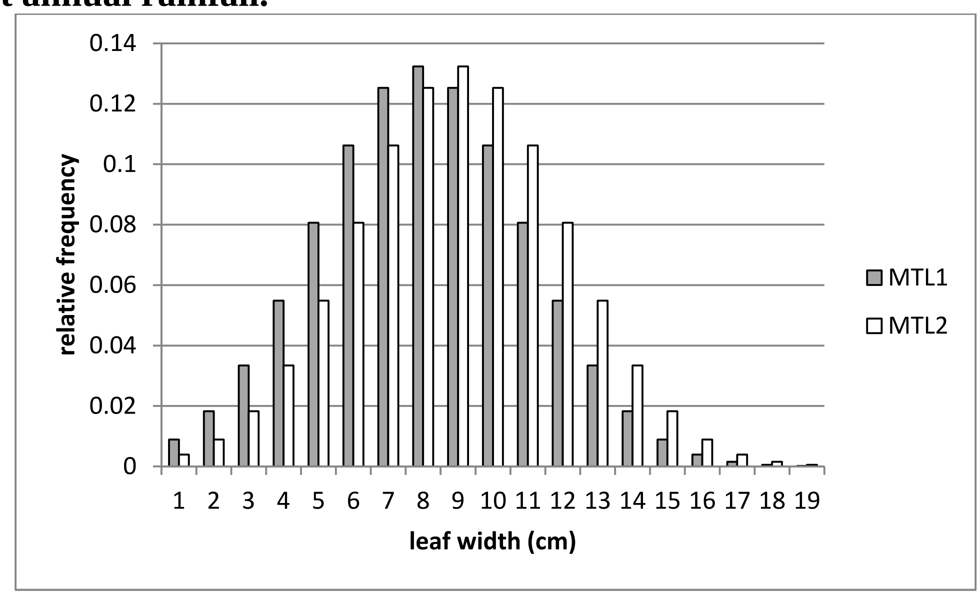


Figure 10. Distributions of leaf widths from maple tree samples from areas of different annual rainfall.



Use Figures 9 and 10 and your knowledge of normal distributions to answer the following questions.

Q18: What are the means of FBL1, FBL2, MTL1, and MTL2? Label the x-axis of Figure 9 with the means of FBL1 and FBL2. Label the x-axis of Figure 10 with the means of MTL1 and MTL2.

Q19: What is the difference between the means of FBL1 &FBL2? What is difference between the means of MTL1 &MTL2?

Q20: Compare the x-axis of Figure 9 to the x-axis of Figure 10. Which difference appears to be more significant, the difference between the means of FBL1 and FBL2 along the x-axis of Figure 9, or the difference between the means of MTL1 and MTL2 along the x-axis of Figure 10? Explain your choice.

Q21: Thinking about the influence of variance on the significance of the difference between sample means, what is another method, using a ruler, you could use to evaluate the significance of the differences in the comparisons in Figures 9 &10? (**HINT: Think about the differences between sample means relative to the sample variances!**)

Q22: Using the method you came up with in the previous question, which comparison of sample means (FBL1 vs FBL2 or MTL1 vs MTL2) appears to be more significant? Why?

Check your answers to the previous questions before moving on!

In summary, smaller sample variances makes the difference between two samples more significant. Another way of looking at this is comparing the degree of overlap between the samples in each figure. The greater the percentage of overlap of two distributions, the less significant are the differences between the two distributions' means. Naturally, for a given difference between sample means, the distributions of two samples with larger variances would overlap with one another more than the distributions of two samples with smaller variances.

Comparing population means: the t-test

In this exercise, you will actually use a statistical test, known as the **t-test**, to compare the average values of particular leaf traits of two different species of trees. You will be using data that you collect yourself, as well as calculating the test-statistic, and using this to help you make a decision regarding the appropriate null and alternative hypotheses. The **t-test** is frequently used to compare sample means to determine if the population means are likely to be different.

How the t-test works:

> **The t-test uses the *sample* means to determine if the *population* means are likely to be different. It works if the following statements are true:**

i) **The populations from which the samples were drawn are normally distributed.** *NOTE: Remember our discussion about the importance of sample size? The larger a sample size, the more likely it is to have an approximately normal distribution. Therefore, this is why it's always a good thing to make your sample sizes as large as feasibly possible!*

ii) **The populations from which the samples were drawn have the same variance.** *This is usually not a problem if the sample sizes are the same, or very close. This makes it important to have equal, or nearly equal sample sizes, in order for the t-test to be an appropriate one for your data!* (This is why we chose a baseball team and soccer team to compare heights in as opposed to comparing a baseball team and a football team: the latter has many more players than the other two teams.)

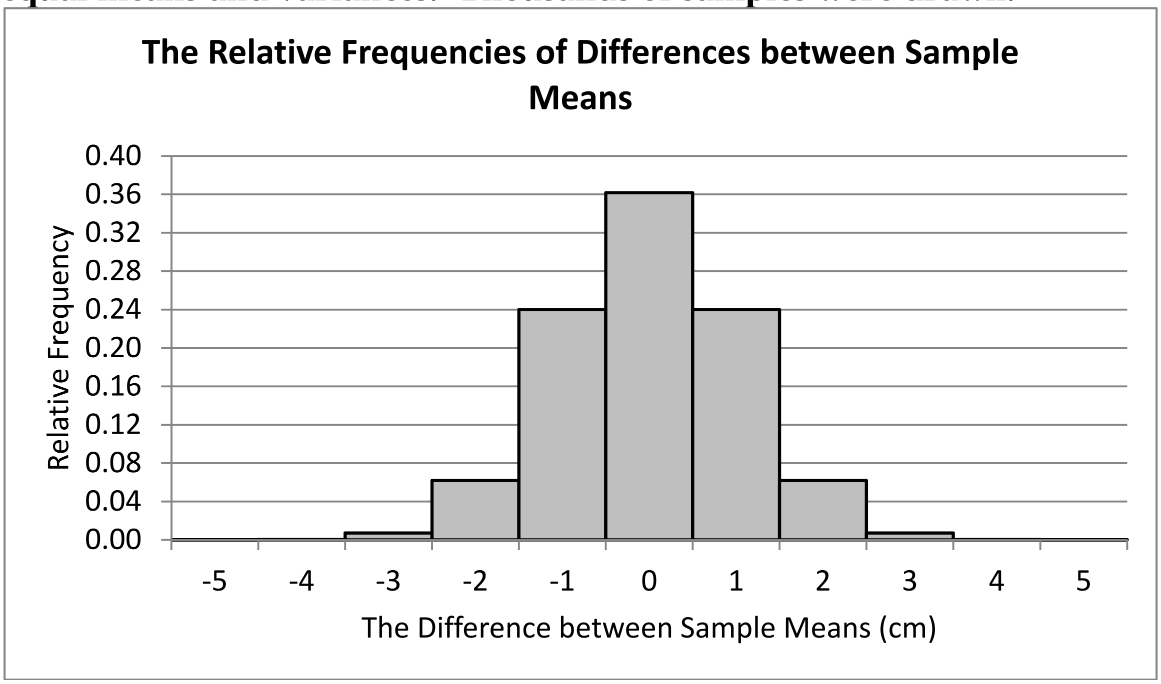
When we perform the t-test we are essentially testing the following null hypothesis:

> **H_0: The populations from which the samples were drawn have the same mean.**

This is the hypothesis that we are testing, so at the beginning of the experiment, we don't know if it is true.

Suppose that two populations are normally distributed with the same mean and variance. Imagine that instead of sampling the populations just once, we sampled the populations thousands of times, and recorded the difference between the means each time we sampled (rounded to the nearest half centimeter). From these data, we made the relative frequency bar graph that is pictured in Figure 11. This bar graph can be used to estimate the probability that the difference between sample means falls into a given range. For example, according to Figure 11, the probability that the difference between sample means is between 0.5 and 1.5 is about 0.24.

Figure 11: The relative frequencies of values for differences between sample means. The samples drawn from two normally distributed populations with equal means and variances. Thousands of samples were drawn.



Q23: What is the most likely value of the difference between sample means? Does this make sense? Why?

Q24: Estimate the probability that the difference between the sample means is greater than 2.5 cm. Is this event likely?

Q25: Suppose that you sample two normally distributed populations with the same variance as the populations that we used to make Figure 11. After computing the sample means you find that the difference between them is 3 cm. Do you think that these populations have the same mean? Why or why not?

The t-test works in this same way, only there is no need to sample the populations thousands of times and create a relative frequency bar graph because the assumptions of the t-test *determine* the probability distribution that the difference in the sample means should follow**. The shape of this probability distribution depends on the population variances.** In particular, the difference in the means is more likely to be large for populations with larger variances. In order to eliminate the dependence of the probability distribution on the variances, we divide the difference in the means by a measure of the variances. This new quantity is called the **t-statistic**. Its probability distribution is called a **student's t distribution**. The t-statistic may be positive or negative, but the larger its magnitude, the more likely it is that the difference between the means is significant, that is, the less likely it is that the population means are the same.

The student's t-test for the comparison of two means uses the ***sample sizes*** (***n***), the ***sample means*** ($\bar{x}$) you have already calculated for species A and species B, and the ***sample variances*** (s^2; remember, these are our estimates of the ***population variances***) you have already calculated for species A and species B (remember to divide by *n-1* for these variances instead of just *n*!). The t-test also uses the ***standard deviations*** (***s***) that you found by taking the **square roots** of the variances.

- List these values on a sheet of paper for each of your species as follows:

$\bar{x}_A =$	${s^2}_A =$	$n_A =$	$s_A =$
$\bar{x}_B =$	${s^2}_B =$	$n_B =$	$s_B =$

How to compute the t-statistic:

The t-statistic is a quotient. The numerator is the difference of the sample means and the denominator is a measure of the variance of the two samples. Below is an equation showing how the t-statistic is calculated when the number of individuals in each sample are the same. Let $\bar{x}_A$ denote the mean of sample A, $\bar{x}_B$ denote the mean of sample B, ${s^2}_A$ denote the estimate of the variance of population A, and ${s^2}_B$ denote the estimate of the variance of population B, and n denote the size of the samples (which is the same value for both samples). **Then the t-statistic is computed as follows:**

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{{s^2}_A + {s^2}_B}{n}}}$$

If the sample sizes are not the same, then the formula for the t-statistic is a little more complicated. In particular, we will need to calculate the **pooled standard**

deviation ($\boldsymbol{s_{AB}}$), before we can find it. If $\boldsymbol{n_A}$ is the size of sample A, and if $\boldsymbol{n_B}$ is the size of sample B then the pooled standard deviation is equal to

$$\boldsymbol{s_{AB}} = \sqrt{\frac{\boldsymbol{s^2{}_A(n_A - 1) + s^2{}_B(n_B - 1)}}{\boldsymbol{n_A + n_B - 2}}},$$

The t-statistic is then calculated as follows:

$$\boldsymbol{t} = \frac{\boldsymbol{\bar{x}_A - \bar{x}_B}}{\boldsymbol{s_{AB}}\sqrt{\frac{\boldsymbol{s^2{}_A}}{\boldsymbol{n_A}} + \frac{\boldsymbol{s^2{}_B}}{\boldsymbol{n_B}}}}$$

Q26 (Super Solver Problem): Show that the formula for the t-statistic with equal sample sizes is equivalent to the formula for the t-statistic with unequal sample sizes if $\boldsymbol{n_A = n_B}$.

Recall that the assumption that the population means are the same is called the **null hypothesis**. It is the hypothesis that we are checking. If the t-statistic is large enough, then we will reject the null hypothesis and conclude that populations have different means. However, if the t-statistic is not large enough to reject the null hypothesis, we *cannot* conclude that the populations have the same mean. In this case, all we can say is that the data do not support the conclusion that the populations have different means.

If all of the assumptions mentioned above are met, that is, if the populations are both normally distributed with the same variance, and **if the populations from which the samples were taken have the same mean**, then the t-statistic follows a student's t distribution with $\boldsymbol{n_A + n_B - 2}$ degrees of freedom. (**NOTE: The concept of "degrees of freedom" is difficult to grasp, and will not be discussed in detail here. However, the number of degrees of freedom in a statistical test is related to sample size, and also determines the critical value of the test, which will be discussed a bit later.**) Figure 12 is a plot of several student's t distributions.

- Use Figure 12 (on page 41) to answer the following questions.

Q27: How does the variance of the student's t distribution change with the degrees of freedom?

Q28: If a t-statistic has a student's t-distribution with 58 degrees of freedom, how does the probability its absolute value is greater than 3 compare to the probability that its absolute value is less than 3?

Q29: As the sample size increases, what happens to the degrees of freedom?

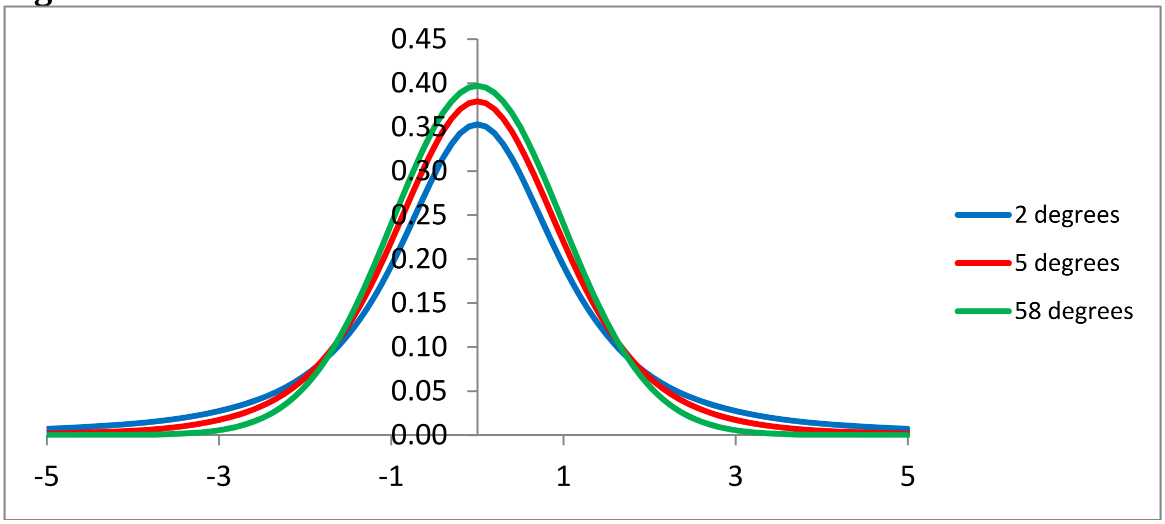
Q30: How does the probability that the absolute value of the t-statistic is bigger than 3 change as the sample size gets bigger? Does this make sense? Why or why not?

In general, if t follows a student's t distribution then the probability that $|t| \geq x$ decreases as x gets bigger. In particular, there is a number x_c, that depends on the degrees of freedom, so that the probability that $|t| \geq x_c$ is less than 0.05. **Scientists usually reject the null hypothesis if the absolute value of the t-statistic is greater than or equal to x_c. In effect we reject the null hypothesis if it implies that the chances of obtaining such a large t-statistic are less than 5%.**

In summary, if the two populations from which the samples were taken have the same mean, that is, if the null hypothesis holds, then the t-statistic should follow a student's t distribution, and so its absolute value should be less than x_c 95% of the time (again, this incorporates the 5% Type I error rate used by most scientists). If the absolute value of the t-statistic is greater than or equal to x_c, we reject the null hypothesis and conclude that the population means are not the same. If the absolute value of the t-statistic is less than x_c, we cannot reject the null hypothesis, and so the outcome of the test is indeterminate, that is, the test does not support the conclusion that the means are different, nor does it allow us to claim that the means are the same.

Q31: Why do we reject the null hypothesis if the absolute value of the t-statistic is greater than or equal to x_c?

Figure 12: Several student's t distributions.



- For a student's t-distribution with 58 degrees of freedom (and a confidence level of 95%), $x_c \approx 2.00$. Calculate the t-statistic for your data set and compare it to x_c. Is the difference between your sample means significant enough that you should conclude that the population means are different as well?

Exercise 4: Simple Shell Match (*Grades K-12*)

Materials: The shells from containers C (red dot) and D (blue dot)
The shell identification sheets, located at the end of this book

How to play:

- Use identification sheet C to match each shell from container C to its name and picture.
- Repeat the exercise with the shells from Container D.

Exercise 5: Making Sense of Variation: The Matching Game (*Grades 3-12*)

Mollusk shells come in an amazing variety of shapes, patterns and colors. There are about 100,000 species of mollusks, and many of these species have their own special shell. Although every shell is unique, individual shells may have some characteristics in common. Because of this, we can sort shells by the characteristics they share, just as we sort the dishes in our kitchens. This exercise and the next demonstrate how biologists group organisms based on the characteristics they share.

- Check to see that each shell in Container C has a red dot on it. There should be a total of 30 shells in this container.
- Become familiar with the parts of a shell and the various shell types, textures, projections, and color patterns by studying Figures 1 & 2 on page 44.
- Choose a shell from container C and examine it carefully.
- Now talk with the other students in your class and examine their shells in order to find those shells that most closely resemble your own shell. Form a group with the students whose shells most closely resemble your own.
- Once you have formed a group, devise a group name and write it on two sheets of paper.
- Identify the key characteristics that all of the shells in your group share, and record them on both sheets of paper.

- Find the black number on each shell in your group, and record these numbers on *one* of the sheets of paper.

- Give your shells to your teacher along with the sheet of paper that lists the shell characteristics but *does not* show the shell numbers.

- After all groups have turned in their shells and papers, the teacher will give each group a list of descriptive characteristics that was made by a different group.

- Now the teacher will spread the shells out on a long table.

- Collect all of the shells that have the characteristics from the list that the teacher gave your group.

- Record the shells' numbers on the sheet of paper that the teacher gave your group.

- Once all of the groups have finished, one at a time, each group should announce the group name from the top of the sheet of paper that the teacher gave them and then read the list of traits aloud while holding up the matching shells for the class to see. Also they should read the numbers on the shells so that the group that originally compiled the list can compare these numbers to the numbers on their own list.

- As a class discuss how successful each group was at finding and collecting another group's shells. If a group was unable to collect all the correct shells, discuss what characteristics could be added to the list to make the task easier.

- Consult the answer sheet to Exercise 3 to find the name and type of each shell in container C.

- Please check to see that all red dot shells are placed back in Container C.

Figure 1: Parts of a gastropod (snail) shell and bivalve (clam) shell.

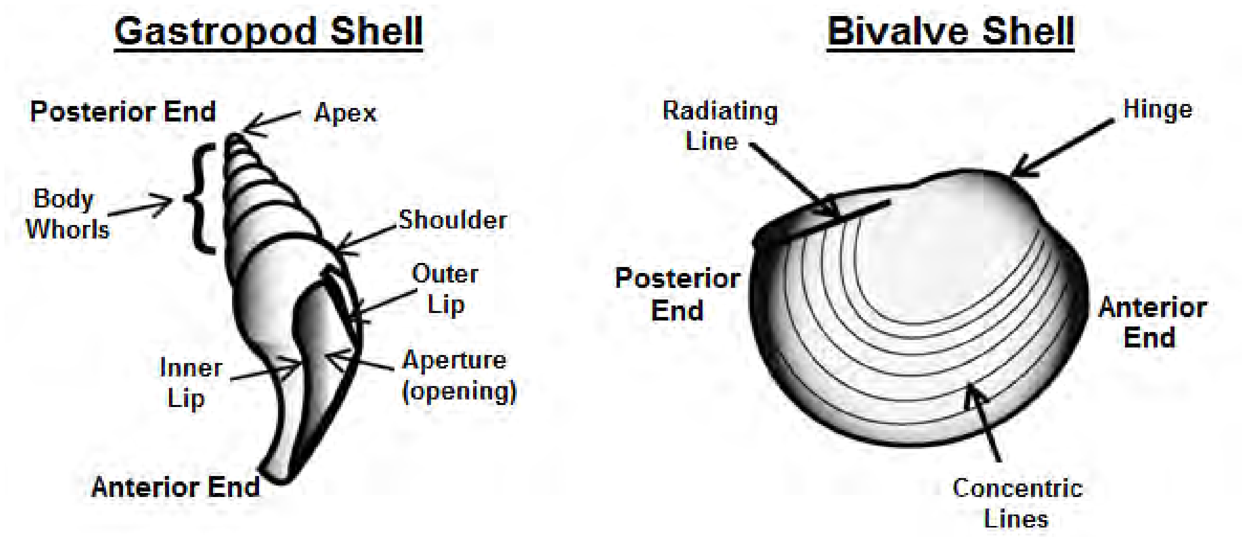
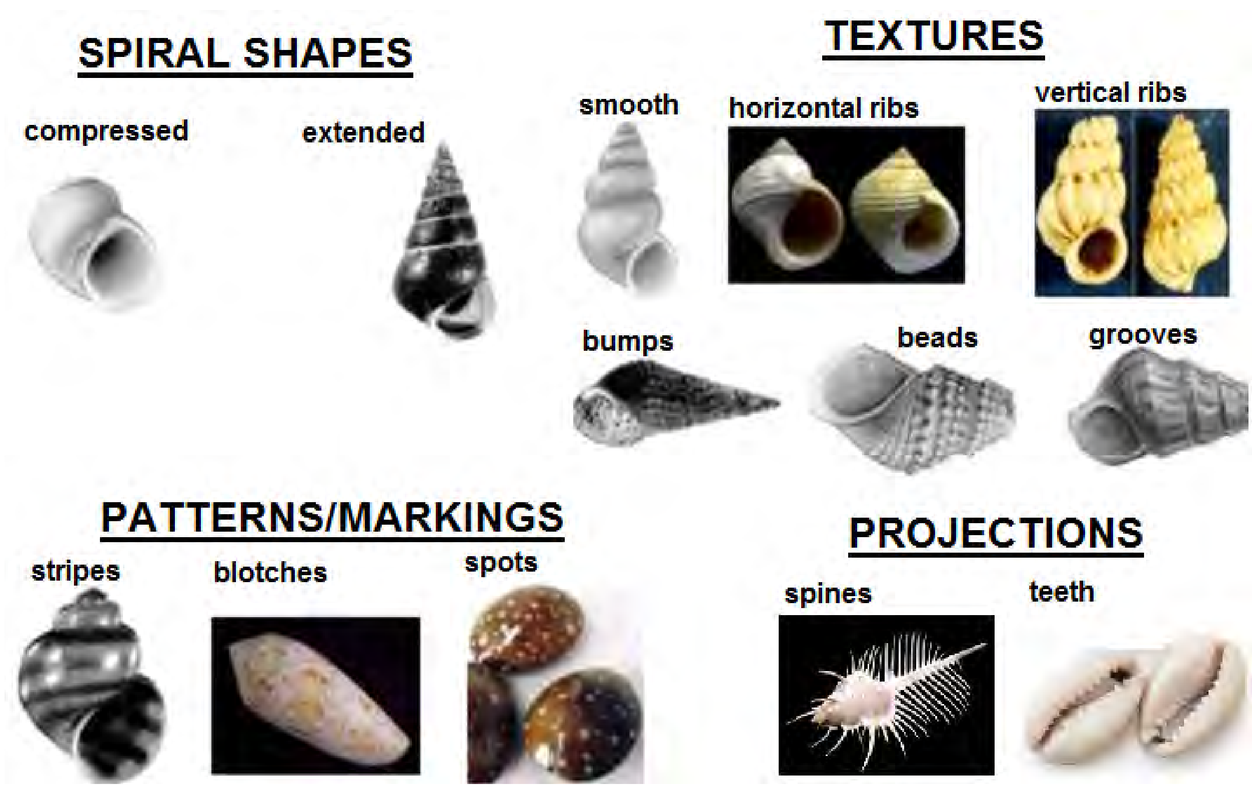


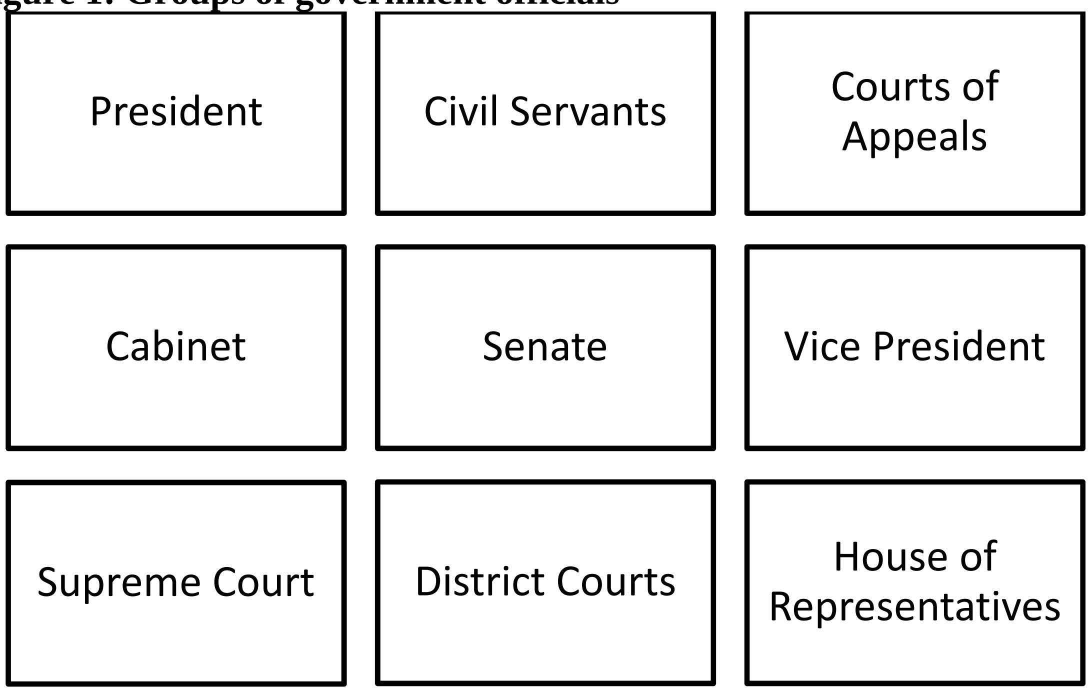
Figure 2. Example shell characteristics that can be used to sort shells.



Exercise 6: Finding Species (*Grades 5-12*)

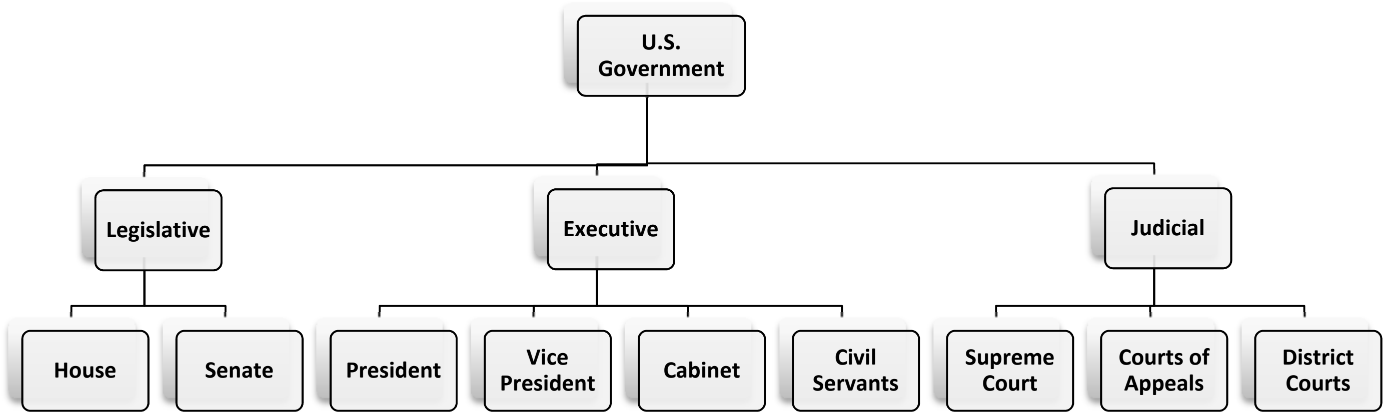
In this exercise, you will learn about **hierarchy** and **classification**. When scientists classify organisms, they do more than just separate the organisms into groups of like organisms, they also reveal how the organisms are related, or how similar they are. Suppose, for example that you wanted to classify the officials in our federal government. You might start by assigning government officials to the following groups.

Figure 1: Groups of government officials



However, this grouping doesn't tell us anything about how the government functions, or how these groups are related. In order to see how these groups of government officials are related, we need to create a tree diagram which shows the **classification hierarchy,** that is, the criteria by which these groups of officials are classified, and the order in which these criteria are considered. Figure 2 is such a diagram. This diagram shows that the government is broken into three separate branches, which are responsible for the three basic tasks of governing. Then, each branch of government is broken into groups of officials that are charged with specific duties. That is, government officials are classified first by the branch of government that they belong to, and then by their specific title.

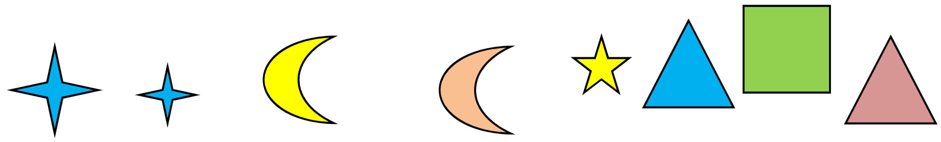
Figure 2: Tree diagram of government officials



As a second example, we can classify the students in your class by their birth place. In this example, the criteria of classification would be birth country, birth state or province, and then birth city, in that order.

- Create a tree diagram to show how the students in your class are classified based on place of birth.

- Examine the objects below and create a hierarchy by which to classify them. Draw a tree diagram to show how the objects are related through your classification hierarchy. There are many possible responses. Talk with other students in your class to see how they classified the objects.



The classification hierarchies that we have examined so far are small compared to the hierarchy by which scientists classify life. The first criteria by which scientists classify life is called the kingdom. One kingdom is the kingdom Animalia (the animals). This kingdom consists of multicellular organisms that obtain their food by eating other organisms. The kingdoms are then divided into phyla. The animal kingdom is divided into approximately 30 phyla, which share similar body plans. One phylum is the phylum Mollusca (mollusks). Mollusks are characterized by the presence of a body cavity that contains organs, a mantle cavity that provides gas exchange, and a protective shell. Each phylum is divided into classes. There are five classes in the kingdom Mollusca. The two classes of mollusks used in this exercise are the Gastropoda (snails and their relatives) and the Bivalvia (clams and their relatives). There are four additional levels of classification below the class level in this hierarchy: order, family, genus and species.

In this exercise, your goal is to play the role of a **taxonomists** (scientists who focus on classifying the broad diversity of living organisms). In this exercise, you will classify the mollusk shells in Container C and then those in Container D by successively partitioning them into two groups until each group is composed only of shells that are very similar in appearance. When you are finished, each group of shells that result from your classification hierarchy will represent a species, at least as you define the term.

- Spread the shells from container C out onto a long table. All of these shells should have a red dot. (Remove any shells that have a blue dot and return them to container D).

- Consult the diagrams in Figures 1 & 2 from the previous exercise (on page 44) to see parts and characteristics of shells that you may want to use in your classification. Can you think of any other criteria by which to classify these shells?

- Choose a characteristic by which to separate the shells into two groups. Write the two variants of the characteristic on the board. Your goal is to make a tree diagram like the one in Figure 5 (page 48).

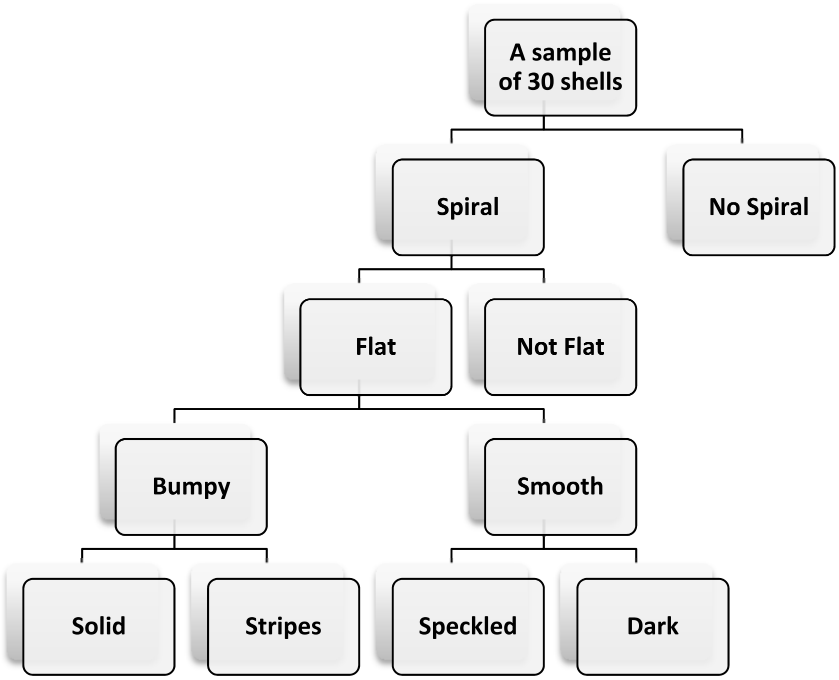
- Continue to choose criteria by which to divide your groups, and update your diagram accordingly until each group consists only of shells that are very similar in appearance. As shown in Figure 5, you may finish partitioning one group of shells before the other.

- It will be helpful if you keep track of the numbers of each shell that you place into each group at each split in your hierarchical diagram, as this will facilitate quick location/identification of shells when discussing this exercise in class.

- Label each final group on your diagram with the numbers of the shells that belong to that group. Consult the answer sheet to Exercise 6 to see that each of your groups consists only of shells from the same species.

* Repeat this exercise, using the shells from Container D.

Figure 5: A hierarchical diagram that shows how shells were partitioned into groups based on their characteristics.



In this example, the first split separates one class of mollusks from the other shells: the shells of bivalves (class Bivalvia, which includes clams and their relatives) , lack a spiral, while all the other shells were species of snails (class Gastropoda), which have a spiral. Inspection of the figure shows that five hierarchical criteria were used to separate this particular collection of shells into groups.

Species concepts & "lumpers" versus "splitters" (*Open-ended exploration, Grades 5-12*)

In the previous exercise, you played the role of taxonomists, deciding where to draw the line between *species* represented by the samples of shells in your containers. Many of you may have arrived at different decisions on which shells constituted species in the previous exercise. This should illustrate to you a fairly important point: taxonomy is a human construct.

Of course, some organisms are more closely related to one another than others, but how scientists break them into various taxonomic groups is a way that humans have devised to help organize closely related organisms together in a way that makes sense. Defining larger groups (orders, families, classes, phyla/divisions, kingdoms/domains) of organisms that share a common ancestry and traits is a work in progress, and often used as an example of how science is self-correcting, yet what defines a **species** is also a major topic of debate. The big question for many biologists is "where ***do*** we draw the line at what constitutes a species?" There are tons of different concepts that have been used in both historical and modern contexts on what defines a "species", and taxonomists use all sorts of criteria on dividing species (morphology, ecology, genetics, ability to mate and produce viable offspring etc.).

It often seems, to many biologists, that there are nearly as many different species concepts as there are types of living organisms! Ask any two biologists their definition of a species, and you may get very different answers. Most (though not all) biologists define species using the **biological species** concept. The biological species concept is one of the first more formal definitions of species, and defines a species as consisting of "***a group of interbreeding populations that are reproductively isolated from other such groups***" (Mayr, 1942). This definition is one of the more commonly used ones for a species, though again, many other biologists define species in different ways. However, there are several other major species concepts that are used by other scientists to define a species. These concepts are all different in various ways, but may sometimes agree (though they may often disagree) on when to call two organisms different species. A little later on, you will be asked to learn a little more about some other major species concepts on your own, or in teams.

How taxonomists classify organisms into species usually falls somewhere between two opposite ends of a spectrum of "lumpers" and "splitters". A "lumper" would be a taxonomist that groups (lumps) together lots of different organisms in the same species on the basis of several broad similarities that they view as being important characteristics that defines the species. "Lumpers" generally view species as biological units that may display a very broad range of inter-individual variation. On the other hand, a taxonomist who separates organisms into separate species on the basis of lots and lots of smaller details would be known as a "splitter". Splitters generally take the view that species, though they may display some important degree of inter-individual variation, are biological entities that display a lot of important (though often miniscule) highly-conserved traits, and that recognizing substantial variation beyond that is worthy of giving other organisms their own formal species names.

Lumping and splitting can both be viewed as having value in different ways, but both "schools of thought" in taxonomy have drawn some criticism. A quote from

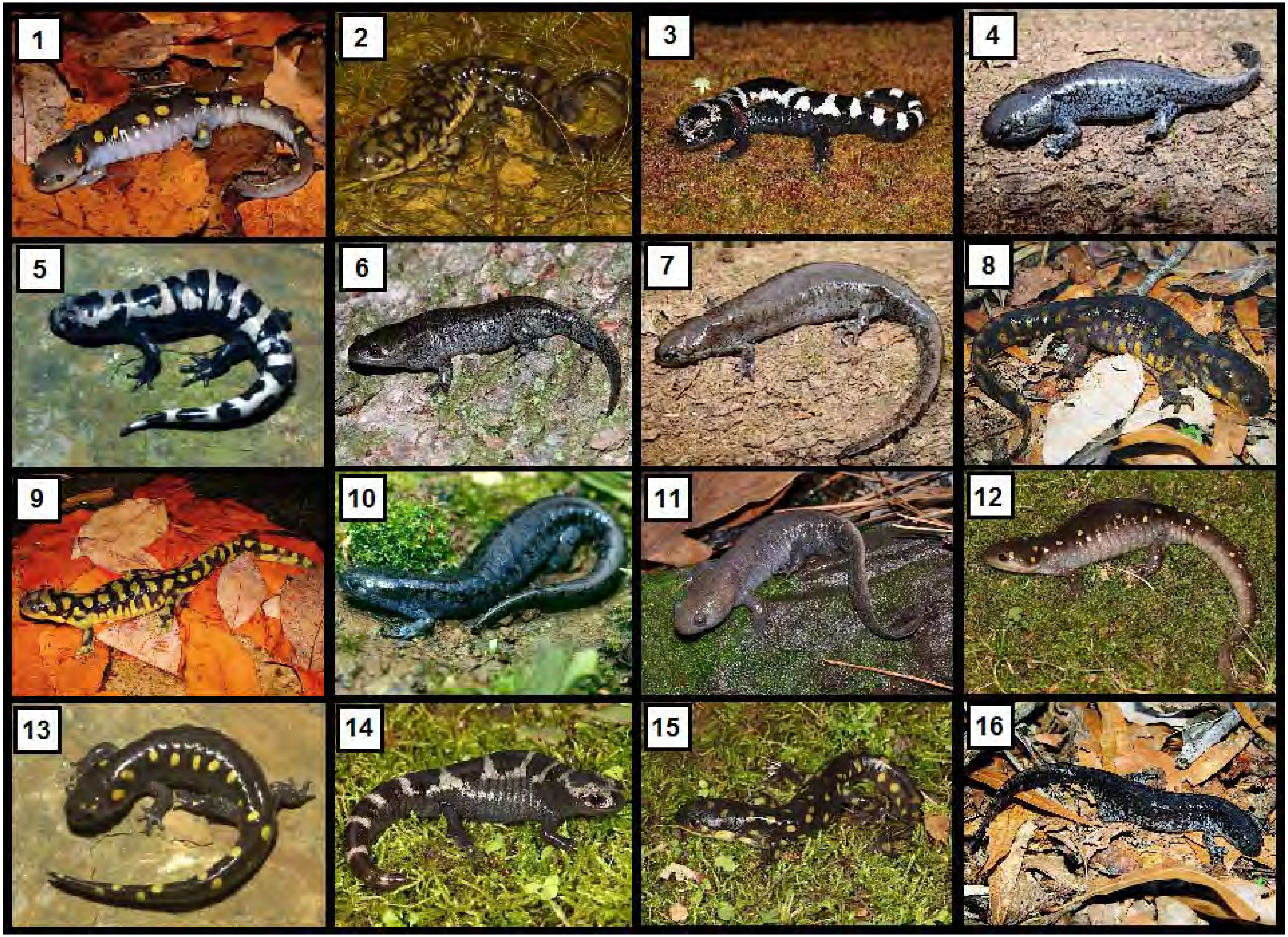
George Simpson, a famous paleontologist (a scientist that studies fossils) once provided the following tongue-in-cheek definitions of both:

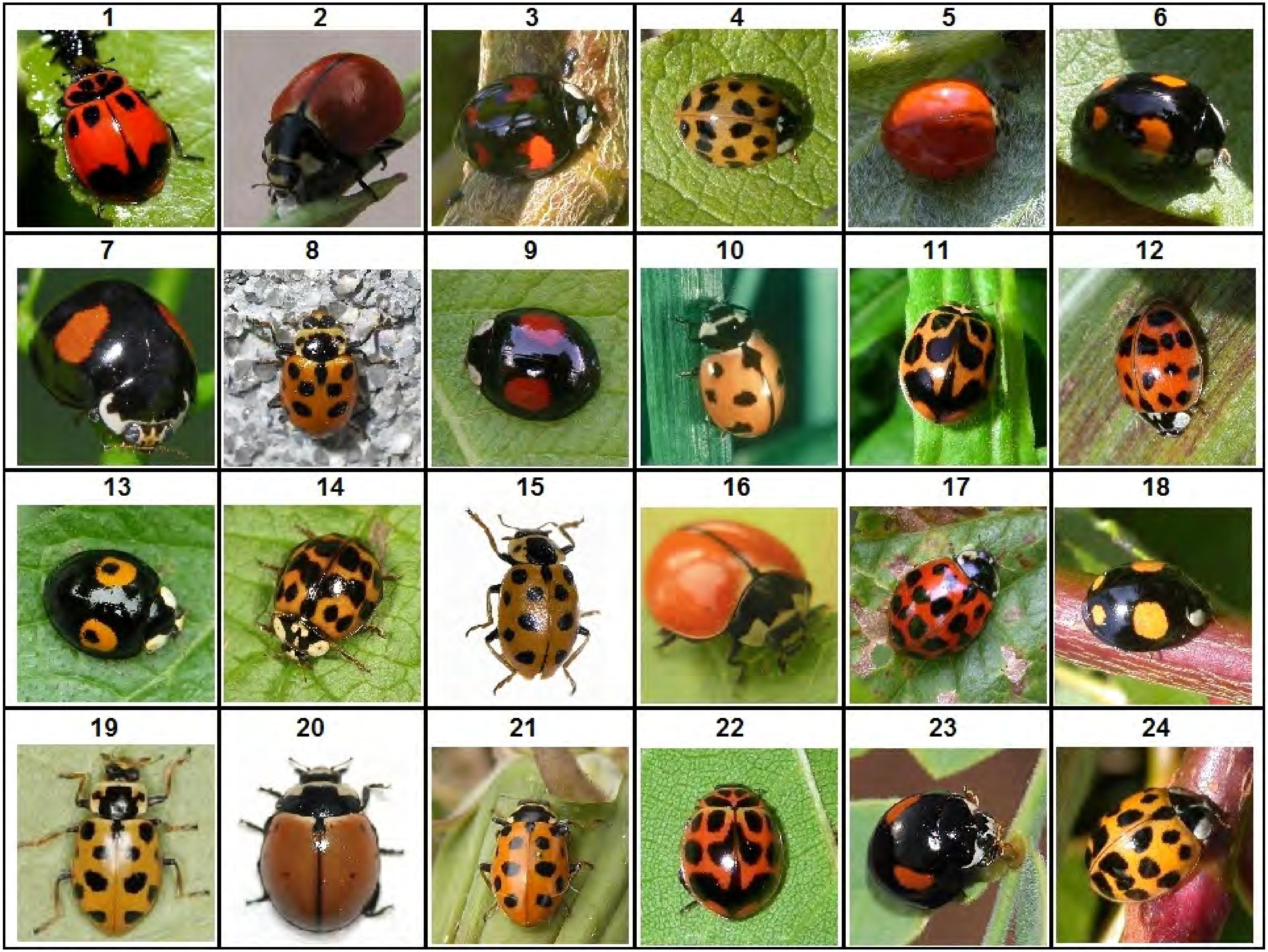
> *"Splitters see very small, highly differentiated units – their critics say that if they can tell two animals apart, they place them in different genera … and if they cannot tell them apart, they place them in different species. … Lumpers, on the other hand, see only large units – their critics say that if a carnivore is neither a dog nor a bear, they call it a cat."*

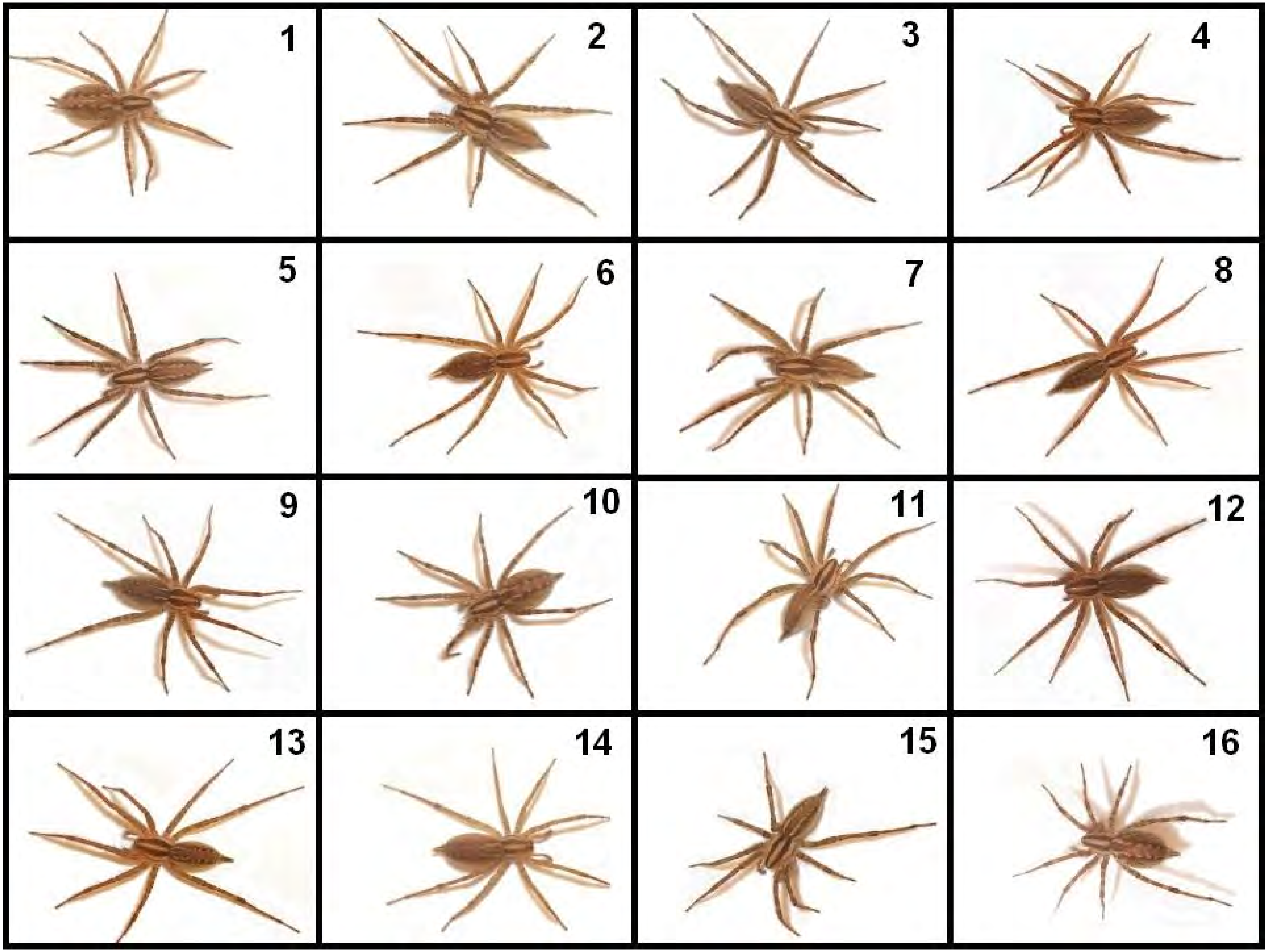
- On the following pages, you will play the role of taxonomists, using images of several groups of organisms, many of which you might even see here in Tennessee! You should work through these groups of organisms in order:
    - First, you will look at salamanders in the genus *Ambystoma*.
    - Next, you will examine lady beetles (ladybugs) in the family Coccinellidae.
    - Finally, you will examine spiders in the genus *Agelenopsis*. **NOTE:** Since spiders often exhibit **sexual dimorphism** (male and female spiders often look very different), the images of the spiders in this exercise consist of females only.
- For this exercise, you should work in groups of 3-4.
- For each group of organisms, examine the numbered pictures of individuals in that group.
- Discuss with your teammates the visible characteristics of the individuals, and see if you can come to a consensus decision on how many species are represented by that group of images.
- You may find it helpful to draw hierarchical tree diagrams as you did in the previous exercise to help you keep track of traits that you used to separate the organisms into species.
- After the entire class is finished sorting a particular group into species, get your teacher, or a scribe from your team, to write on the board how many species you think is in that group.
- Next, try to answer the following questions via class discussion.
    - Did the teams differ in the number species they identified in the three groups? If so, which teams would be considered "lumpers," which would be considered "splitters," and which teams would fall somewhere in between?
    - What proportion of the teams could be classified as lumpers, as splitters, or as 'tweeners' (in between lumpers and splitters)?

- Now organize yourselves into three larger teams (lumpers, splitters, and 'tweeners') based on the number of "species" your smaller teams identified.
- Take a few minutes to talk with your new larger team, but then discuss, as a class, your overall reasoning for the number of species that you identified in the sample (in other words, why you fell on a particular end, or the middle, of the lumper/splitter spectrum).
- After you have done this for each of the groups of images, check the answers in the back of the book to see exactly how many species (as identified by the most recent taxonomy) are represented in each group!
- When you are finished with this exercise, you should also take some time to do some library and/or internet research and find some definitions of the following species concepts:
    - **Typological species concept**
    - **Ecological species concept**
    - **Evolutionary species concept**
- See if you can find other popular species concepts and their definitions.
- Try to answer the following questions:
    - Which species concept do you think is the closest to the one you used in dividing up the images in each group into species?
    - Which species concept seems the most sensible to you, and why?
    - Why is the notion of how one defines a species important?

Some additional information on these topics is also discussed in the answer section of this book!


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Exercise 7: Mystery Shell (*Grades K-12*)

- Examine the "mystery shell" in Container E.
- See if you can figure out what kind of shell it is.
- Check your answer in the back of the book to see if you were right!

Exercise 8: Mechanisms Underlying Variation (*Grades 3-12*)

Exercise 8a: Variation in leaves and flowers

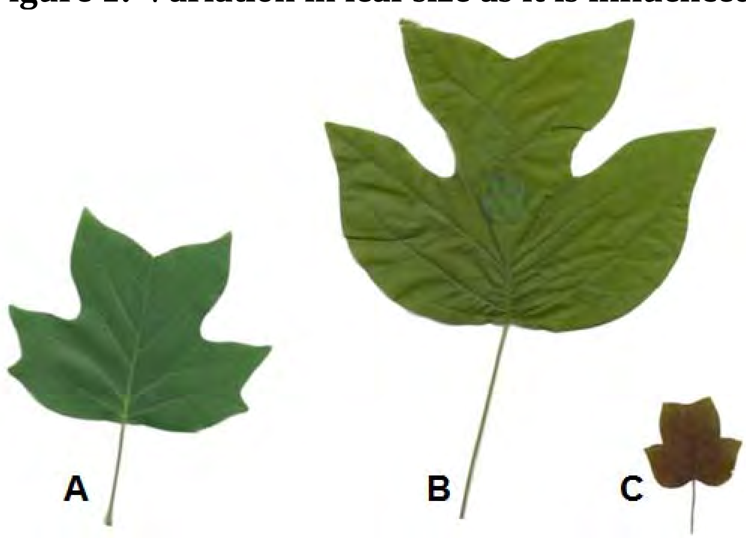
So far in our exploration of variation, we have seen that individuals vary, and we have learned how to measure variation within a population. We have also seen that populations vary, and learned to measure the significance of inter-population variation. We have also learned that individuals can be classified according to their differences via a classification hierarchy.

This exercise introduces the mechanisms that underlie trait variation. We will use leaves and flowers as examples. You have also been presented with an additional sample of gastropod shells for an exercise that allows you to measure their shells, and explore the concept of variation among species in the context of adaptations.

- Examine the leaves in Figure 1 below. These leaves were collected from the same tree. In fact, they all came from the same branch. Note that individual leaves differ in size. Within a single species, differences in size are frequently associated with differences in age or developmental stage. As a young leaf unfolds from its leaf bud, it gains room to grow larger. All it needs is time and nourishment in order to grow to the size of a typical mature leaf.

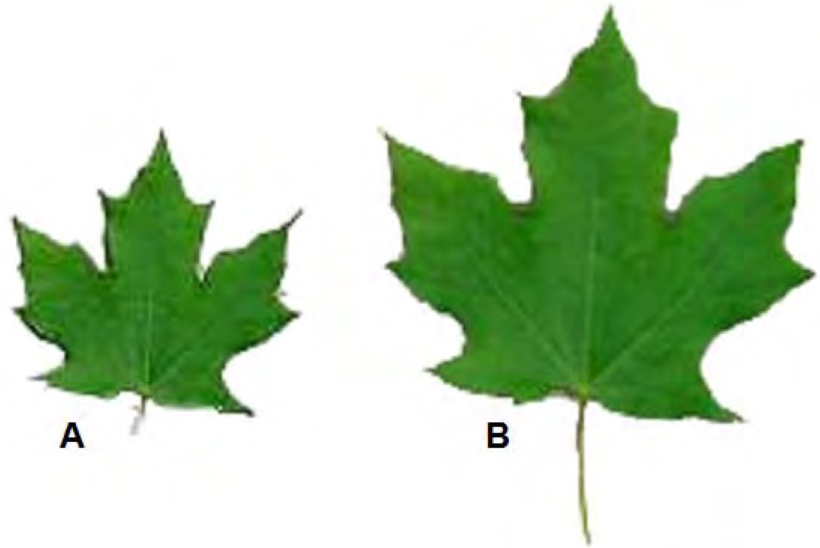
  Q1. Can you rank the leaves in Figure 1 from youngest to most mature? Check your ranking under Exercise 7.1 in the answer booklet. Which species of tree did these leaves come from? Check your answer by looking at the Leaf Guide near the end of this book.

Figure 1: Variation in leaf size as it is influenced by age.



- Examine the two leaves in Figure 2 below.

Figure 2. Two leaves from different locations on a tree.



Unlike the leaves in Figure 1, these leaves are both mature, and yet they differ markedly in size. As you can see, leaf B has a much greater surface area than leaf A. However, leaf A is actually thicker than leaf B. These leaves were collected from different parts of the same tree. One of the leaves is a **sun leaf** which was collected from a branch near the top of the tree. The other leaf is a **shade leaf** which was collected from one of the lower branches. The size of these leaves was influenced by their environment and the special needs of trees. Unlike animals that must consume food to obtain energy, trees obtain energy by making their own food when sunlight hits their leaves. In order to convert sunlight to usable chemical energy, plants need water, and lots of it. It is estimated that the typical oak tree needs about 50 gallons of water a day. Therefore, lack of water and light can severely limit the amount of energy that a tree leaf can produce. Because the top of the tree receives more intense sunlight than the bottom of the tree, leaves at the top of the tree are able to produce more food per unit surface area than leaves at the bottom of the tree. However, leaves at the top of the tree are also subject to greater water loss per unit surface area, also because they are exposed to more sun and wind.

Q2. Why does the top of the tree typically receive more sun than the bottom of the tree?

Q3. Which type of leaves would you expect to have a greater surface area: shade leaves or sun leaves? Use the information from the preceding paragraph to create a hypothesis as to why your answer is correct.

Q4. Which of the leaves in Figure 2 is the sun leaf?

Q5. How much larger is leaf B's surface area than leaf A's surface area? Come up with a way to estimate the surface area of the leaves and then use your surface area estimates to perform the following calculations. Calculate how many times larger leaf B's surface area is than leaf A's surface area. Calculate the percent difference between the surface areas of leaf A and leaf B.

- Examine the leaves in Figure 3 below.

Figure 3. Variation in sassafras leaf shape.

- Note that there are four different leaf shapes:
  - Elliptical/ovate (balloon or egg-shaped)
  - three-lobed (two side lobes and one main lobe)
  - two-lobed, left side
  - two-lobed, right side
- Also note that two of the shapes look like left-handed and right-handed mittens.
- Locate each of the four leaf shapes on the diagram.

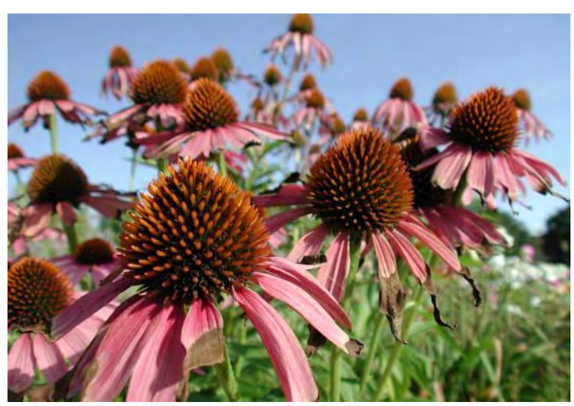
Sassafras is a tree that produces all four leaf types, often on the same branch! If a trait, such as leaf shape, varies for genetic reasons, then the distinct variants of the trait are called **morphs**, and the trait is said to be **polymorphic** (has many morphs). Thus, using this new terminology we would say that sassafras exhibits a leaf shape polymorphism with four morphs.

Q6. Can you think of other examples of polymorphisms? Check Exercise 7.3 in the answer booklet for some examples.

We have learned that leaves can vary in size and shape for a number of reasons. Examine the flowers in Figure 4. Do you see many differences between the two flowers of the red lily plant or among the many flowers of purple coneflowers? Though there are numerous different species of flowering plants, within a species, of all the parts of a plant, the flower is the least variable. We call traits that show little inter-individual variation **conserved traits**. Read on to learn why flowers show little variation within a species.

Flowers house the reproductive organs of the plant. The male organs are called **anthers**. The female organ is called a **pistil**. In order to reproduce, flowering plants need animals called **pollinators** to carry pollen produced by the anthers of one plant to the pistil of another plant. Pollinators are often species-specific, that is, they pollinate a specific set of species. For example, hummingbirds may pollinate morning glories but they do not pollinate daisies. Therefore, it is important that pollinators be able to recognize the flowers of the species they pollinate. If a pollinator visits the wrong type of flower, then any pollen that it leaves behind is wasted. Pollen is very costly to make, and is also species-specific. Seeds will not form if the wrong pollen is delivered to a pistil. Thus the flowers of a particular plant species have little variation.

Figure 4: Red lily flowers (left), purple coneflowers (right)



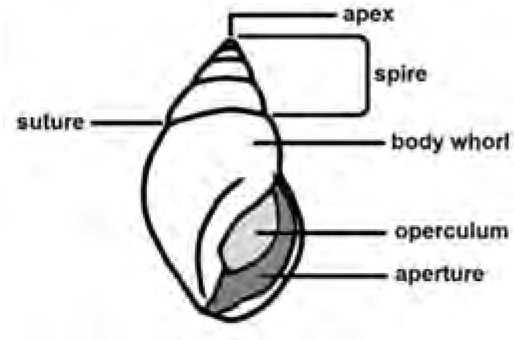
Exercise 8b: Variation in snail shells (*Grades 3-12*)

NOTE TO TEACHERS: *This exercise is duplicated in Unit 6 (Animal Kingdom), but fits well with the introduction of variation in gastropod shells in this unit. For further exercises exploring mathematics using mollusk shells, you may also wish to look into some of the exercises in Unit 6!*

Are you left-handed or right-handed? Although they don't have hands, interestingly, snails also have a form of "handedness," based on the direction of the opening of their shell. How can you tell if the spiral of a snail is left-handed or right-handed? One easy way to do this is to hold the shell so that the opening is towards you, and the apex (top of the spiral) is pointing upwards. If the opening is pointed towards your left, your snail has a left-handed spiral. If the opening is pointed towards your right, your snail has a right-handed spiral. Another way is to hold the shell in the same position, and place a finger in the groove of the spiral at the apex. Trace this groove with your finger so that you are moving along the spiral path towards the opening in the shell. If your finger moves in a clockwise path along the groove, your snail has a right-handed spiral. If the path your finger traveled was counterclockwise, the spiral of the shell is left-handed. Examine Figure 1 below for an illustration of shells with left-handed and right-handed spirals.

Figure 1. "Handedness" in snail shells. The shell on the left displays a left-handed spiral, while the shell on the right displays a right-handed spiral.

Figure 2. Illustration of basic snail shell morphology.



As a snail grows, it makes its shell "house" bigger to hold its larger body. The smallest inner coil of the shell is called the **apex.** It is what remains of the original shell that the snail formed before it even hatched. Notice the spiral groove on the snail's shell. If you place your finger in the groove at the shell's apex, and follow its path, you will see that the shell gets bigger and wider as you move along the groove. This reflects the way the snail made it shell larger and larger by adding new material as it grew.

Aside from whether they are left- or right-handed, snail shells are also differentiated by their shape and size. Look at the shells in your sample and notice their many different forms. Biologists that study shells measure their features so that they can be **quantified,** or described with numbers. Today you will play the role of a **conchologist** (pronounced con-KAH-luh-gist), or scientist that studies shells. You will measure the shells in your sample so that you can quantify their features.

Examine your sample of snail shells. You may notice lots of variation amongst these shells, in terms of size, color, texture, etc., but what about shape? Can you divide your sample of shells into several groups according to their general shape?

- Separate the collection of shells into several piles, by placing shells that share similar shapes together. Try to come up with a name to describe each shape group.
- After you have grouped the shells according to shape, examine the shells within each shape group to see how they vary in size. Position the shells within each shape group according to their apparent size from smallest to largest.
- After you have ordered the shells in each shape group according to apparent size, measure each shell's size. There is more than one way to measure a

shell's size. Choose one of the possible measures of shell size listed below, and measure this value for each shell.

- Make a table in which to record each shell's apparent size ranking (1 being the smallest), and its size measurement. See the table below for an example.

Rank	**Size measurement (number of whorls)**
1	1.25
2	2.75
3	3
4	3.5

Q7. Does your ranking according to apparent size agree with your size measurements? Why might your visual rankings differ than rankings based on your measurements?

<u>Size measurements for students of all ages:</u>

- **Number of whorls:** This is the number of full rotations of the spiral path of the shell as you move from the apex to the opening. To calculate the number of whorls place your finger on the spiral groove at the apex of the shell and trace its path. Count how many full circles you make as you follow the path to the shell's opening. Estimate partial rotations as you near the opening as fractions of a full rotation (for example, 0.25, 0.5, etc.).
- **Width and/or height of the aperture (opening)**
- **Shell height:** Hold the shell so that the opening is facing you, and the apex is facing upwards. The height of the shell is the straight-line distance from the very bottom to the top of the shell when held in this position.
- **Shell width:** With the shell held in the same position as above, measure the greatest straight line distance across the shell.
- **Greatest shell circumference:** Wrap a piece of string around the widest part of the shell, then straighten out the string and measure the part that wrapped all the way around the shell at this point.

- An alternative of measuring height, width, or circumference of the shells could be accomplished via tracing each of the shells on a piece of paper, and taking the measurements from the traced outlines. Though this will introduce some error due to the width of the instrument used to trace each shell, since this will be the same for each shell, this would still be an effective way of obtaining ranked size measures of each shell. Circumference of each shell could then be calculated from the drawings, since the greatest width of each outline would correspond to the diameter of the shell, and the known relationship ($\boldsymbol{C = \pi d}$) between circumference and diameter.

<u>Size measurements for grades 3 and up:</u>

- **Shell volume:** If you measure the shell circumference and its height, you can estimate the volume of the shell by using the equations for the volume of a cone (for pointier shells), a cylinder (for more flattened shells), or a sphere (for rounder shells). These formulas are listed below.

$$\boldsymbol{Volume\ of\ a\ cone = \frac{1}{3}\pi r^2 h}$$
$$\boldsymbol{Volume\ of\ a\ cylinder = \pi r^2 h}$$
$$\boldsymbol{Volume\ of\ a\ sphere = \frac{4}{3}\pi r^3}$$

  In these equations, $\boldsymbol{h}$ represents the shell height, $\boldsymbol{\pi}$ ("pi") is a constant that is approximately equal to 3.14, and $\boldsymbol{r}$ is the radius of a circle. In the formula for the cone $\boldsymbol{r}$ represents the radius of the circle that forms the base of the cone. In the formula for the cylinder $\boldsymbol{r}$ represents the radius of the circle that forms the base of the cylinder. In the formula for the sphere $\boldsymbol{r}$ represents the radius of the equator of the sphere. You can calculate the radius from your value of shell circumference.

- **Area of the aperture:** You can estimate this area by using the formulas listed below. Use the formula for the shape that is the most like the opening to your shell.

$$\boldsymbol{Area\ of\ a\ circle = \ \pi r^2}$$
$$\boldsymbol{Area\ of\ an\ ellipse = \frac{1}{2}\pi ab}$$

- In the formula for the area of a circle, $\boldsymbol{r}$ represents the radius of the circle. In the formula for the area of an ellipse, $\boldsymbol{a}$ and $\boldsymbol{b}$ represent the width and height of the ellipse respectively.

Q8. The equation for the circumference of a circle with radius $\boldsymbol{r}$ is $\boldsymbol{C = 2\pi r}$. Rearrange this equation to solve for $\boldsymbol{r}$ in terms of $\boldsymbol{C}$.

Q9. How is a snail's shell volume related to its body size?

- We can also compare shells to see how much bigger one shell is than another shell. In particular, we can find the **factor** by which one shell's size is greater or less than another shell's size. The factor by which one number is greater or less than a second number is the number that we have to multiply the second number by to get the first number back. For example, the value of a dime is greater than the value of a nickel by a factor of 2, because we have to multiply the value of a nickel by 2 to get back the value of a dime.

<u>Factors</u>

If y is greater or less than x by a factor of c, then $y = cx$.

If we divide both sides of this equation by x we find that $\frac{y}{x} = c$.

In other words, the factor by which y is greater/less than x is equal to $\frac{y}{x}$.

Q10. What is the factor by which the value of a nickel is less than the value of a dime?

Q11. What is the factor by which the value of a nickel is less than the value of a quarter?

Q12. What is the factor by which the value of a quarter is greater than the value of a nickel?

Q13. Complete the following sentence: If x is greater than y by a factor of $, c$ then y is less than x by a factor of ___.

- Suppose that we ranked a sample of five shells according to their whorl numbers and then organized this data in a table like the one on the following page.

Whorl number	The factor by which the whorl number increases (relative to next smaller shell)
1	
2	
2.5	
3.25	
3.5	

- To complete the table we need to find the factor by which the whorl number increases from one shell to the next. For example, to find the factor by which the whorl number increases between shell #1 and shell #2, we divide the number of whorls that shell #2 has by the number of whorls that shell #1 has.

$$\frac{\#\ of\ whorls\ on\ shell\ \#2}{\#\ of\ whorls\ on\ shell\ \#1} = \frac{2}{1} = 2$$

That is, shell #2 is twice as many whorls as shell #1. Similarly, we can find the factor by which the number of whorls increases between shell #2 and shell #3.

$$\frac{\#\ of\ whorls\ on\ shell\ \#3}{\#\ of\ whorls\ on\ shell\ \#2} = \frac{2.5}{2} = 1.25$$

That is, shell #3 has 1.25 times as many whorls as shell #2.

Whorl number	The factor by which the whorl number increases (relative to next smaller shell)
1	Can’t be calculated
2	2
2.5	1.25
3.25	
3.5	

- Complete the table above, and then construct a similar table for each of your shell shape groups to show the factor by which the size measurement that you chose increased from one shell to the next.

Exercise 8b.2: Thinking about adaptive variation (*Open-ended exploration*)

Examine the shells in each shape group. What can you learn about a snail's lifestyle (i.e. where it lives, what it eats, how fast it moves, how it defends itself against predators, etc.) by examining the shape of its shell? Does the shape of a snail's shell serve a function, help the snail to accomplish particular tasks, or is the shape of the shell the result of the environment in which the snail lives? Do other shell traits, besides shape, give you insight into a snail's life?

- Pick a particular aspect of a snail's lifestyle that interests you.
- Think of a few (at least 2-3) traits that might be related to the lifestyle aspect that you chose. Below are listed a few examples of additional traits that you could measure, in addition to the size-related traits from earlier. You may also measure any other traits that you think might be important.
- Measure these traits for each of your shells, and record this data in tables for each shape group, similar to the one below. Use mm for linear measurements.

Shell #	Trait #1	Trait #2	Trait #3

Below are some other shell traits you can measure. If necessary, ask your teacher for help making these measurements. Look back at Figure 2 if you need any help with terminology regarding parts of the shell. You are welcome to measure any other traits related to shell shape that may interest you. Just be sure to clearly define how you make such a measurement, and be consistent in the methods of your measurements from shell to shell.

<u>Other measurements:</u>

- **Color variation:** Some shells are a solid color, while others are variegated. A possible quantitative measure of the variation in shell color would be to count the number of colors on the shell.
- **Relative height:** This is simply the height of the shell divided by its width. This number could actually tell someone about the shape of a shell without

them even seeing it! Large values of this ratio indicate that a shell is long and thin, while small values of this ratio indicate that a shell is short and flat.

- **Whorl expansion rate:** This is the factor by which the diameter of the aperture increases with a single rotation.   It may be difficult to measure the whorl expansion rate if your shell is intact and the whorls overlap. However, there are many ways to estimate this rate. For example, you could estimate the whorl expansion rate by dividing the height of the visible part of a particular whorl by the height of the visible part of the whorl formed before it (closer to the apex).
- **Shell angle:** This measurement also gives some information about the shape of a shell, and can be measured as the angle formed by the apex of the shell and the most external sutures on the shell.

For an illustration of these measurements, as well as many other informative measures of shell shape, the following link provides images, as well as descriptions on how to measure these traits:

http://en.wikipedia.org/wiki/File:Gastropod_shell_measuring.png

Looking for correlations:

- After you have recorded your data, look to see if the traits you measured are correlated. If one trait is usually large when another trait is large, then there is a positive correlation between the traits. If one trait is usually small when another trait is large, then there is a negative correlation between the traits. You may wish to create graphs or plots to better visualize and evaluate whether such relationships are present.
- If relationships exist between traits, are these relationships similar or consistent among different shape groups? Why do you think this may or may not be the case?

Further open-ended exploration:

- Using the shell identification guide in the answer section of this book, find the species of snail that made each shell in your collection, and add this information in a new column in your tables.
- Use the library or the internet to research these snail species. As you research each snail species, try to answer the following questions:
    - Are the shell traits that you measured related to the lifestyle aspect that you chose to study?
    - If so, why might these traits be important?
    - How do they relate to your chosen aspect of snail lifestyle?

  - If not, can you find other aspects of the snails' lifestyles that appear to be related to the traits you chose to measure?
- Prepare a brief report of your findings to present to the rest of your class.
- **NOTE:** Teachers may wish to provide a scientific review article on the importance of various aspects of land snail shell shape (included with the teacher materials for Unit 6 on the *Biology in a Box* CD) to more advanced students to assist them. However, students should consider the following:
  - The provided article is primarily concerned with land snail species, while some of the provided shells come from aquatic species.
  - Do you think that shell shape traits may differ in importance between land and aquatic snails? If so, how?
  - What about different types of aquatic habitats (for example, freshwater versus marine dwelling snails)?

An optional web-based exercise:

- Go to *http://www.ams.org/featurecolumn/archive/shell6.html*, and play around with some of the parameters (***W***, ***D***, and ***T***), clicking the "*Draw*" button to see their effects.
  - Can you think of ways to measure your snail shells so that you can estimate these parameters?
  - Try to see if you can do so for one or more of your shells. If you think you have come up with a way to estimate ***W***, ***D***, and ***T*** for those shells, try using those parameters in the model.
  - Can you produce a shape similar to that of any of the shells in your sample?

ANSWER SHEETS

Answers for Exercise 1: Recognizing Individuals as Unique.

Q1. Calculate the relative frequency of an A+ in each class. Which class made relatively more A+'s?

The relative frequency of an A+ in Mrs. Brook's class was approximately 0.23 while the relative frequency of an A+ in Mrs. Tichnour's class was 0.3. Therefore, relatively more students made an A+ in Mrs. Tichnour's class.

Q2. Which trait(s) yielded the most successes? Are the traits with more successes necessarily the best ones to use when identifying leaves? Why or why not? How could we change the game to make the number of successes that belong to a trait a better measure of how effectively a trait can be used to identify leaves?

The traits with the most successes are not necessarily the best ones to use when identifying leaves. For example, these traits may have yielded more successes because they were more often employed. Or it could be that some traits work well for certain people but less well for others. There are many ways that we could alter the game to make the number of successes a better measure of how effectively a trait can be used to identify a leaf. If instead we assigned each trait to an equal number of students then the number of successes that each trait yielded would be a better measure of how effective the trait can be used to identify leaves.

Q3. Which trait(s) yielded the highest success rates? Are the traits with higher success rates necessarily the best ones to use when identifying leaves? Why or why not? How could we change the game to make the trait success rate a better measure of how effectively a trait can be used to identify leaves?

The traits with the highest success rates are not necessarily the best ones to use when identifying leaves. For example, if a trait was employed by a single person, and that person was successful at identifying their leaf then the trait's success rate would be a perfect 100%. However, it could be that the trait is only effective at identifying that one particular leaf, or that the person was just very good at using that trait. There are many ways that we could alter the game to make the success rate a better measure of how effectively a trait can be used to identify a leaf. For example, we could add more leaves to the game so that each student is required to identify multiple leaves.

Answers for Exercise 3: Using Mathematics to Describe a Population

Q1. Suppose that you wanted to find out what brand of pants are most often worn by fifth grade boys in a small town. Which of the following samples do you think would best answer this question? Why? What is wrong with the other samples?

e) **The boys in Mrs. Brooke's fifth grade class**
f) **The boys whose last names begin with the letters A-D or M-P.**
g) **The boys that ride bus 23**
h) **The boys that sit together in the far right corner of the cafeteria**

Sample b is probably the best sample to use when answering this question, because the boys in sample b are not likely to be linked in ways that affect the brand of pants that they wear. The boys in sample a, c, and d, see each other every day and so they might influence each other. The boys in sample c live in the same neighborhood and so they probably belong to the same socio-economic class, thus their parents may shop for the same brand of pants. The boys in sample d are likely friends and so they may be similar in many additional ways that might cause them to prefer the same brand of pants.

Q2. What if instead you wanted to know the average height of a fourth grade boy? Could any of the samples that you rejected before be used to answer this new question?

Sample a or b could be used to answer this question. We probably shouldn't use sample c or d because these boys likely belong to the same socio-economic class, and so they may have similar diets which in turn affects their growth.

Answers for Exercise 3a: Measures of central tendency

Exercise 3a.1: Calculating measures of central tendency (*Grades 3-12*)

Q3: Compute the mean, median, and mode for each of the following data sets. Which measure of central tendency best describes the data? Why?

a)

Value	Absolute Frequency
3	2
5	2
6	5
8	2
9	1

$$mean = \frac{2(3)+2(5)+5(6)+2(8)+9}{2+2+5+2+1} = \frac{6+10+30+16+9}{12} = \frac{71}{12} \approx 5.9$$

$$median = 6$$

$$mode = 6$$

The median, mean, and the mode are all good measures of central tendency for this data set, because they are all approximately the same.

b)

Value	Absolute Frequency
5	3
6	4
8	2
25	1

$$mean = \frac{3(5)+4(6)+2(8)+25}{3+4+2+1} = \frac{15+24+16+25}{10} = \frac{80}{10} = 8$$

$$median = 6$$

$$mode = 6$$

The median and the mode are both good measures of central tendency for this data set. The mean is too large, because it is heavily influenced by the outlier value of 25.

Exercise 3a.2: Sigma notation (*Grades 9-12*)

Q4: Evaluate the following expression:

$$\sum_{i=6}^{9}(17-2i)$$

$$\sum_{i=6}^{9}(17-2i) = \big(17-2(6)\big) + \big(17-2(7)\big) + \big(17-2(8)\big) + (17-2(9))$$

$$\sum_{i=6}^{9}(17-2i) = (17-12) + (17-14) + (17-16) + (17-18)$$

$$\sum_{i=6}^{9}(17-2i) = 5+3+1+(-1) = 8$$

Q5: Evaluate the following expression:

$$\sum_{i=0}^{4} 2^i$$

$$\sum_{i=0}^{4} 2^i = 2^0 + 2^1 + 2^2 + 2^3 + 2^4 = 1+2+4+8+16 = 31$$

Q6: Evaluate the following expression:

$$\sum_{i=3}^{6}(-1)^{i+1}(3i)$$

$$\sum_{i=3}^{6}(-1)^{i+1}(3i) = (-1)^{3+1}(3(3)) + (-1)^{4+1}(3(4)) + (-1)^{5+1}(3(5)) + (-1)^{6+1}(3(6))$$

$$\sum_{i=3}^{6}(-1)^{i+1}(3i) = (-1)^4(9) + (-1)^5(12) + (-1)^6(15) + (-1)^7(18)$$

$$\sum_{i=3}^{6}(-1)^{i+1}(3i) = 1(9) + (-1)(12) + 1(15) + (-1)(18)$$

$$\sum_{i=3}^{6}(-1)^{i+1}(3i) = 9 + (-12) + 15 + (-18) = -6$$

Q7 (Super Solver Problem): How could you write "the sum of all multiples of 3 from 6 to 27" in sigma notation?

There are actually many possible solutions, but one of the ways you could write this is as follows:

$$\sum_{i=2}^{9} 3i$$

Other possibilities include

$$\sum_{i=0}^{7} 3i + 6$$

and

$$\sum_{i=1}^{8} 3i + 3$$

Did you find any other possible solutions?

Exercise 3b: Measures of spread

Exercise 3b.1: Calculating measures of spread

Q8 (Super Solver Problem): If you completed Exercise 3a.2, how would you write the formulas for population variance and sample variance using sigma (summation) notation?

$$Population\ variance\ (\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$Sample\ variance\ (s^2) = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Exercise 3b.2: Sample variance vs population variance: What a difference "1" makes

NOTE TO TEACHERS: For this exercise, your "population" container should hold a total of 100 beads, with 75 beads of one color representing a value of 1, and 25 beads of another color representing a value of 5. The resulting population mean μ=2, and the resulting population variance σ^2=3 (population standard deviation $\sigma \approx 1.73$). However, you may wish to double check the numbers of each color of beads and recalculate these population parameters if necessary to give students the correct values.

Q9: What do you notice about the average values of sample variance (s^2) using n versus (n-1) in the denominator of the equation when compared to the true population variance (σ^2)?

When calculating sample variance using n in the denominator, you should notice that, on average, the resulting value is smaller than the true population variance. In other words, the use of n in the denominator tends to underestimate the true amount of variance in the whole population. When using (n-1) in the denominator, the calculated sample variance might be lower or higher than the true population variance, but provides a much closer estimate of the true population variance than when using n in the denominator of the formula. This is why the formula for finding sample variance (which is used as an estimate of the true population variance) uses (n-1) in the denominator.

Q10: In effect, what are you doing when you are calculating the mean of all students' sample means?

When calculating the mean of all sample means, you are essentially pooling the data from each sample, and calculating the sample mean of a much larger sample size.

Q11: Compare your calculated value of the mean of all students' sample means, as well as each of the individual sample means obtained, to the true population mean. Which value is closest to the true population mean? Considering your answer to Q10, what does this tell you?

You should notice that some of the individual sample means (of size n=3) are lower, and some are higher, than the true population mean. However, when all of these samples are pooled to make a much larger sample size, the resulting sample mean is much closer to the true population mean. This should illustrate the fact that the larger the sample size, the better the sample mean is as an estimate of the true population mean. This also holds true for sample statistics that estimate variation in a population (variance and standard deviation). This is why scientists typically try to use the largest sample sizes that time and funding allow, as larger sample sizes give them greater confidence that their estimates of characteristics of entire populations are close to their actual values.

Exercise 3c: Visualizing measures of central tendency and spread

Q12: Figure 4 below shows three data sets that are normally distributed. How do their distributions differ? Do they have the same mean? Do they have the same variance? Which distribution has the greatest mean? Which distribution has the smallest variance?

These distributions differ in both their means and variances. The green distribution has the greatest mean. The blue distribution has the smallest variance.

Q13: How can you interpret the area of region C (in Figure 5) in terms of probabilities?

Region C represents the probability that an individual in the population has a leaf width between 5cm and 8cm.

Q14: Now suppose that we select a leaf at random from the population. Use Figure 5 to rank the following events from most to least probable.

d) **The leaf's width is between 0 cm and 3 cm.**
e) **The leaf's width is between 3 cm and 8 cm.**
f) **The leaf's width is between 3 cm and 5 cm.**

From most to least probable, the events are b,c,a.

Answers for Exercise 3d: Comparing leaves from two samples.

Q15: Examine Figure 8. Which species appears to have narrower leaves? Which species appears to have a broader leaf width distribution?

Species 2 appears to have narrower leaves. Species 1 appears to have a broader leaf width distribution.

Q16: Use your bar graph to answer the following questions about species A and species B. Do you think that one of these trees has larger leaves on average? Why or why not? Do you think that one of these trees has a broader distribution of leaf widths? Why or why not? Now use your statistics to answer the same questions again. Did your answers change?

Student answers to this question will vary, based on the samples of Species A and Species B provided in this unit.

Q17: Assuming that a large number of leaves were collected from all of the trees at the respective localities, which of the following best describes the population to which MTL1 belongs?

- **e) All of the maple leaves in North America**
- **f) All of the maple leaves from suburban yards**
- **g) All of the maple leaves found at the locality L1**
- **h) All of the maple leaves that receive approximately 48 inches of rainfall a year**

The correct answer is c) all of the maple leaves found at the locality L1.

Q18: What are the means of FBL1, FBL2, MTL1, and MTL2?

The mean of FBL1 is 3. The mean of FBL2 is 4. The mean of MTL1 is 8. The mean of MTL2 is 9.

Q19: What is the difference between the means of FBL1 &FBL2? What is difference between the means of MTL1 &MTL2?

FBL2 - FBL1=1; MTL2-MTL1=1

Q20: Compare the x-axis of Figure 9 to the x-axis of Figure 10. Which difference appears to be more significant, the difference between the means of FBL1 and FBL2 along the x-axis of Figure 9, or the difference between the means of MTL1 and MTL2 along the x-axis of Figure 10? Explain your choice.

The difference between the means of MTL1 and MTL2 ***appears*** to be more significant because the distance between the mean of FBL1 and the mean of FBL2 is longer than the distance between the mean of MTL1 and the mean of MTL2.

Q21: Thinking about the influence of variance on the significance of the difference between sample means, what is another method, using a ruler, you could use to evaluate the significance of the differences in the comparisons in Figures 9 &10? (HINT: Think about the differences between sample means relative to the sample variances!)

Using a ruler, you could measure the distances between the sample means in each figure. However, keeping in mind the importance of sample variances (reflected by the total span of the x-axis in each figure), you would also want to measure the length of the x-axis in each figure, and calculate the proportion of the x-axis that is taken up by the distance between means in each figure. The greater this value, the greater the significance of the difference between the means compared in that particular figure.

Q22: Using the method you came up with in the previous question, which comparison of sample means (FBL1 vs FBL2 or MTL1 vs MTL2) appears to be more significant? Why?

Using the method from the previous question, you should see that the difference between the means of FBL1 and FBL2 takes up a greater proportion of the x-axis in Figure 9 than does the distance between the means of MTL1 and MTL2 relative to the x-axis in Figure 10, which means that the difference between the means of FBL1 and FBL2 is more significant than the difference between the means of MTL1 and MTL2, even though the magnitude of these differences is the same. In other words, for a given difference between means, the lower the variance, the greater the significance of the comparison.

Q23: What is the most likely value of the difference between sample means? Does this make sense? Why?

The sample difference is most likely to be close to zero. Since the populations from which the samples were drawn have the same mean, it makes sense that the samples are most likely to have similar means as well, therefore the difference in the sample means is likely to be close to zero.

Q24: Estimate the probability that the difference between the sample means is greater than 2.5 cm. Is this event likely?

The probability that the difference in the sample means is greater than 2.5 is less than 0.01. This event is very unlikely.

Q25: Suppose that you sample two normally distributed populations with the same variance as the populations that we used to make Figure 11. After computing the sample means you find that the difference between them is 3 cm. Do you think that these populations have the same mean? Why or why not?

These populations probably have different means because if they have the same mean then the probability of drawing samples with such different means is very small, less than 0.01.

Q26 (Super Solver Problem): Show that the formula for the t-statistic with equal sample sizes is equivalent to the formula for the t-statistic with unequal sample sizes if $n_A = n_B$.

If $n_A = n_B$ then we can substitute n_A for n_B in the equations for the pooled standard deviation and the t-statistic and simplify as follows:

$$s_{AB} = \sqrt{\frac{s^2{}_A(n_A - 1) + s^2{}_B(n_A - 1)}{n_A + n_A - 2}} =$$

$$\sqrt{\frac{(s^2{}_A + s^2{}_B)(n_A - 1)}{2(n_A - 1)}} = \sqrt{\frac{(s^2{}_A + s^2{}_B)}{2}}$$

$$t\ statistic = \frac{\bar{x}_A - \bar{x}_B}{s_{AB}\sqrt{\frac{1}{n_A} + \frac{1}{n_A}}} =$$

$$\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{(s^2{}_A + s^2{}_B)}{2}}\sqrt{\frac{2}{n_A}}} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s^2{}_A + s^2{}_B}{n_A}}}$$

Q27: How does the variance of the student's t distribution change with the degrees of freedom?

The variance of the student's t distribution decreases as the degrees of freedom increases.

Q28: If a t-statistic has a student's t-distribution with 58 degrees of freedom, how does the probability its absolute value is greater than 3 compare to the probability that its absolute value is less than 3?

The probability its magnitude is greater than 3 is less than the probability that its magnitude is less than 3.

Q29: As the sample size increases, what happens to the degrees of freedom?

As the sample size increases the degrees of freedom increases.

Q30: How does the probability that the absolute value of the t-statistic is bigger than 3 change as the sample size gets bigger? Does this make sense? Why or why not?

The probability that the magnitude of the t-statistic is bigger than 3 decreases as the sample size gets bigger. This makes sense because as the sample size grows larger the sample means approach the population means. Since we assume that the population means are the same this forces the difference of the sample means, and therefore the t-statistic, to be closer to zero.

Q31: Why do we reject the null hypothesis if the absolute value of the t-statistic is greater than or equal to x_c?

We reject the null hypothesis if the t-statistic is greater than or equal to x_c because it is highly improbable that the t-statistic is this large if the null hypothesis is true.

Answers for Exercise 6: Finding Species

<u>**Key to salamander species**</u> (species names followed by image numbers)

- **Spotted Salamander (*Ambystoma maculatum*)**: 1, 12, & 13
- **Marbled Salamander (*Ambystoma opacum*)**: 3, 5, & 14
- **Mole Salamander (*Ambystoma talpoideum*)**: 4, 6, & 11
- **Smallmouth Salamander (*Ambystoma texanum*)**: 7, 10, & 16
- **Eastern Tiger Salamander (*Ambystoma t. tigrinum*)**: 2, 8, 9, & 15

This set should have been fairly easy, but a few of the species may have tricked you a little due to some variation. For example, both the Spotted Salamander and the Eastern Tiger Salamander have spots. Though the coloration of the spots on the Spotted Salamander may be similar in coloration to the spots on the Eastern Tiger Salamander, the Spotted Salamander's spots are nearly perfectly round, and not as numerous as the less-perfectly round, more abundant spots (which often creep down onto the sides and merge with the yellow belly) in the Eastern Tiger Salamander. The Mole Salamander and the Smallmouth Salamander have similar coloration, but the Mole Salamander's head is much larger in proportion to the rest of its body. Also note the strong keel (ridge) on the dorsal (upper) surface of the

Mole Salamander's tail! The Marbled Salamander also displays some variation, with the light coloration on its back ranging from dark grey to almost white. However, its overall pattern and large body size make it practically unmistakeable for any other salamander in the U.S.

Salamander photo credits: Emily Moriarty Lemmon (5 & 13); Matt Niemiller (1-4, 6-10, 12, 14-16); John Williams (11)

<u>Key to Lady Beetle Species</u> (species names followed by image numbers)

- **Nine-spotted Lady Beetle (*Coccinella novemnotata*)**: 2, 10, 16, & 20
- **Multicolored Asian Lady Beetle (*Harmonia axyridis*)**: 3, 4, 5, 6, 9, 12, 13, 14, 17, 18, & 24
- **Thirteen-spotted Lady Beetle (*Hippodamia tredecimpunctata*)**: 8, 15, 19, & 21
- **V-marked Lady Beetle (*Neoharmonia venusta venusta*:** 1, 11, & 22
- **Ashy Gray Lady Beetle (*Olla v-nigrum*)**: 7 & 23 (the ones in the exercise are the black and orange form of this species, which also comes in a gray, spotted form like the one pictured below:

More than likely, your decisions on how to group these lady beetles into species relied solely on the obvious variation in color, pattern, and possibly shape, which are the most commonly used means of identification of beetles within this family (Coccinellidae, which is represented by 481 species in America north of Mexico alone). However, other characteristics such as spines on the legs and feet, as well as grooves on the underside of the thorax, which you were not able to see, are also used to identify genera and species in this family. Almost half of the lady beetles in the images belong to the species *Harmonia axyridis*, an invasive species from Asia, which is highly variable, displaying an astounding amount of variation in color and pattern, though it can often be distinguished by a black "M" shape on the pronotum (the area between the head and the wings), though this shape may be broken up, or obscured by a larger dark blotch. See the image below for an example of the variation observable within *H. axyridis*:

Harmonia axyridis was introduced to the United States in California in 1916, again in California in 1964 and 1965, in Washington from 1978-1982, and throughout Canada and the Eastern U.S. from 1978-1981 as a means of biological control of aphids, which are their primary food. However, this beetle spread throughout the U.S. and Canada rapidly, becoming very common throughout North America by 1994. This species has actually become so numerous as to be a nuisance, often invading homes, where they seek shelter from the cold in the winter. This species has also become problematic in that in many regions, it is outcompeting many native lady beetle species, such as the other ones pictured in this exercise, causing populations of those native species to dwindle rapidly in numbers. In fact, since the introduction of *H. axyridis*, *Coccinella novemnotata* (the Nine-spotted Lady Beetle, which also comes in an unspotted, or nearly spotless form, as seen in images 2, 16, and 20), has almost disappeared entirely from the eastern U.S.

Lady beetle photo credits: Bob Barber (23); Christy Beal (1); Christoph Benisch (19); Borealfalcon (21); Margarethe Brummermann (20); Devin Carroll (2, 16); Carol Davis (8); Denis A. Doucet (15); Keith Edkins (3-6, 9, 12, 13, 17, 18, 24); Matt Edmonds (7); Paul Kleiman (11, 22); Tom Murray (14); W. Louis Tedders (10)

<u>Key to spider species</u> (species names followed by image numbers)

- ***Agelenopsis aleenae*:** 3, 5, & 15
- ***Agelenopsis aperta*:** 1, 10, & 16
- ***Agelenopsis emertoni*:** 6, 9, 12
- ***Agelenopsis pennsylvanica*:** 4, 8, 13
- ***Agelenopsis potteri*:** 14
- ***Agelenopsis spatula*:** 2, 7, 11

You may have noticed that most of the spiders in these images were very, very similar. However, taxonomists who specialize in arachnology (the study of spiders) classify members of the genus *Agelenopsis* into different species based on the shapes and characteristics of male and female reproductive organs, which are not readily visible in these images. Also, many of these species occur in geographic ranges (and habitats within ranges) that do not overlap, which many warrant as worthy of species status. These are basically the only means of distinguishing among some species in this genus.

Traditionally, many spider biologists viewed the shapes of male and female spider genitalia as a sort of “lock and key” model of species, in that the shape of a male’s pedipalps determined and limited with which females he would be able to mate. However, though this does appear to be the case in several species of *Agelenopsis* with overlapping ranges (males of a “species” only mate with females of the same species in those areas), genetic data from a few species suggests that some hybridization within some “species groups” composed of several formally described species may occur (Ayoub, Riechert, & Small 2005). How do *Agelenopsis* spiders tell each other apart? It seems to be based on olfactory cues, with females releasing a species-specific pheromone that attracts males of the same species and elicits courtship from them.

Spider photo credits: All photos by J.R. Jones, specimens identified by Dr. Susan Riechert.

Additional Notes on this Exercise

Below are definitions of three major species concepts (aside from the biological species concept, discussed earlier):

- **Typological species concept** – A species is a certain “type” of organism that looks different than another.
- **Ecological species concept** – A species is a set of organisms exploiting (or adapted to) a single niche (Ridley 1993).
- **Evolutionary species concept** – A species is a single lineage of populations which has its own unique evolutionary history and tendencies (Wiley, 1981).

In determining which individuals constituted separate species in this exercise, you undoubtedly relied on the typological species concept, as you had no prior information on which of the species represented are capable of interbreeding, nor on the populations from which they came. Also, keep in mind that although the species identifications listed in the keys for this exercise are agreed on by most specialists who work with those particular organisms, further changes in taxonomy are always possible, as more data about those organisms are gathered!

One reason (of many) that biologists care about species concepts is that how we define a species often has a big impact on when and how conservation efforts are implemented! For example, the U.S. Endangered Species Act defines a "species" as "any *subspecies* of fish or wildlife or plants, and *any distinct population segment* of any species of vertebrate fish or wildlife which interbreeds when mature." This means that specific populations of vertebrates in a particular area, such as grizzly bears, salmon, wolves, etc. can be classified as endangered, even if they are not threatened in other parts of their range, even if they are not described as separate species/subspecies. However, significant populations of other organisms (invertebrates, plants, etc.) can only be protected in parts of their range where they may be threatened only if they are recognized and ***formally described*** as a distinct species or subspecies! Thus, you can see why species concepts are important to lots of biologists, and why they are a hot topic!

Answer to Exercise 7: Mystery Shell

The mystery shell is actually the operculum of a snail. The operculum is a shell door, which many (but not all) snails have. With this door, the snail is able to close its shell to survive periods of drought and to gain protection against predators. Snails that live along the seashore are often exposed to air when the tide goes out, and there are many animals that feed on snails, especially birds. In those situations, an operculum is a valuable adaptation!

Answers for Exercise 8: Mechanisms Underlying Variation

Q1. Can you rank the leaves in Figure 1 from youngest to oldest? Which species of tree did these leaves come from? Check your answer by looking at the Tree Guide in the answer booklet.

The order from youngest to oldest (most mature) leaf is C, A, B.

Q2. Why does the top of the tree typically receive more shade than the bottom of the tree?

The top of the tree typically receives more sun than the bottom of the tree because throughout most of the day the bottom of the tree is shaded by the top of the tree.

Q3. Which type of leaves would you expect to have a greater surface area: shade leaves or sun leaves? Use the information from the preceding paragraph to create a hypothesis as to why your answer is correct.

Sun leaves tend to have a smaller surface area than leaves that develop further down the tree in the shade. Sun leaves also tend to be thicker than shade leaves, and have multiple layers of energy producing cells. Because sun leaves are exposed to intense light that is capable of penetrating multiple cell layers, sun leaves are able to increase food production and minimize water loss by growing thicker instead of wider. Shade leaves, meanwhile, do not have to contend with water loss due to exposure to strong wind and light. Instead, they struggle to obtain enough sunlight. In order to increase their exposure to the sun, shade leaves grow larger surface areas.

Of course trees can't consciously alter the way individual leaves grow, so how does this variation occur? Leaf growth is regulated by multiple genes, some of which are activated or inhibited by light or a lack of light.

Q4. Which of the leaves in Figure 2 is the sun leaf?

Leaf A

Q5. How much larger is leaf B's surface area than leaf A's surface area? Come up with a way to estimate the surface area of the leaves and then use your surface area estimates to perform the following calculations. Calculate how many times larger leaf B's surface area is than leaf A's surface area. Calculate the percent difference between the surface areas of leaf A and leaf B.

We can calculate how much larger leaf B's surface area is as follows:

a) Measure the width of leaf A at its widest point and the length of leaf A at its longest point. Multiply these two measures to obtain an estimate of leaf area: $A_A = l \times w$.

b) Do the same for leaf B.

c) Take the ratio of the two areas to find how many times greater the amount of sunlight leaf B can capture is than the amount of sunlight leaf A can capture:

d) $\frac{A_B}{A_A} = x$

e) Leaf B is capable of capturing x times more sunlight than leaf A.

f) We can also calculate the percent difference in the amount of sunlight that the two leaves can capture as follows

g) $\frac{A_B - A_A}{A_A} \times 100 = x$

h) Leaf B is capable of capturing x% more sunlight than leaf A.

Q6. Can you think of other examples of polymorphisms?

Other examples of polymorphism include butterfly and flower colors within the same species. Some spider species have different body patterns such as the Hawaiian Happy Face spider (*Theridion grallator*), that has several different "faces" on its abdomen. Some fish species, such as the three-spined stickleback (*Gasterosteus aculeatus*) are polymorphic with regard to body size and shape, with one morph being adapted to benthic habitats (on or near the bottom of aquatic habitats), and the other being adapted to more limnetic (open-water) habitats. Many animals such as some fish and lizards have different male morphs which use different mating strategies. Large male morphs defend females from other males, but small male morphs look more like females, and are not chased out of larger males' territories, and can sneak up on females to mate. Many salamanders within the genus *Plethodon* also have two forms: a dark, unstriped "leadback" form, and a reddish, striped "redback" form. These are just a few examples, as polymorphisms can be found in all sorts of other organisms! See the following images for pictures of the organisms discussed here.

Left: the butterfly *Heliconius erato*; **Right:** color polymorphism in the morning glory *Ipomoea purpurea*

Left: larger benthic (bottom-dwelling) and smaller limnetic (water-column dwelling) forms of the three-spined stickleback (*Gasterosteus aculeatus*). Both individuals pictured are gravid (pregnant) females; **Right:** color and pattern polymorphism in the Hawaiian Happy Face Spider (*Theridion grallator*)

Left: leadback (top) and redback (bottom) morphs of the Southern Red-backed Salamander (*Plethodon serratus*)

Q7. Does your ranking according to apparent size agree with your size measurements? Why might your visual rankings differ than rankings based on your measurements?

Students' answers to this question may vary, depending on whether their visual size rankings agree with their ranks based on actual measurements. If these rankings do not agree, this could be because students' visual rankings may have been based on a particular trait (such as shell length, width, etc.), while the trait they actually measured may have resulted in different rankings. This is because different size measures of various parts of the shell may or may not be correlated with the trait they used in their visual rankings.

Q8. Since the equation for the circumference of a circle is $\boldsymbol{C = \pi d}$, and the diameter of a circle is equal to two times its radius ($\boldsymbol{d = 2r}$), how would you solve for the radius to help you calculate the volume?

You could first substitute 2r for d in the first equation to obtain $C = \pi(2r)$. If you then divide both sides of the equation by 2π, you obtain the following: $r = \frac{C}{2\pi}$. Alternatively, if you had obtained the diameter by measuring a traced outline of the shell, you could simply divide the diameter by 2, since the radius of a circle is equal to half its diameter.

Q9. How is a snail's shell volume related to its body size?

The body size of a snail should be positively correlated with shell volume. In other words, the larger the volume of the shell, the larger the body of the snail that inhabited it.

Q10. What is the factor by which the value of a nickel is less than the value of a dime?

Remember, if y is greater or less than x by a factor of c, then $y = cx$. In this case, we can substitute the values of a nickel for *y*, and the value of a dime for *x*, to solve for *c*:

$$\boldsymbol{y = cx}$$

$$\boldsymbol{5 = c(10)}$$

$$\boldsymbol{c = \frac{5}{10} = 0.5}$$

Q11. What is the factor by which the value of a nickel is less than the value of a quarter?

$$\boldsymbol{c = \frac{5}{25} = 0.2}$$

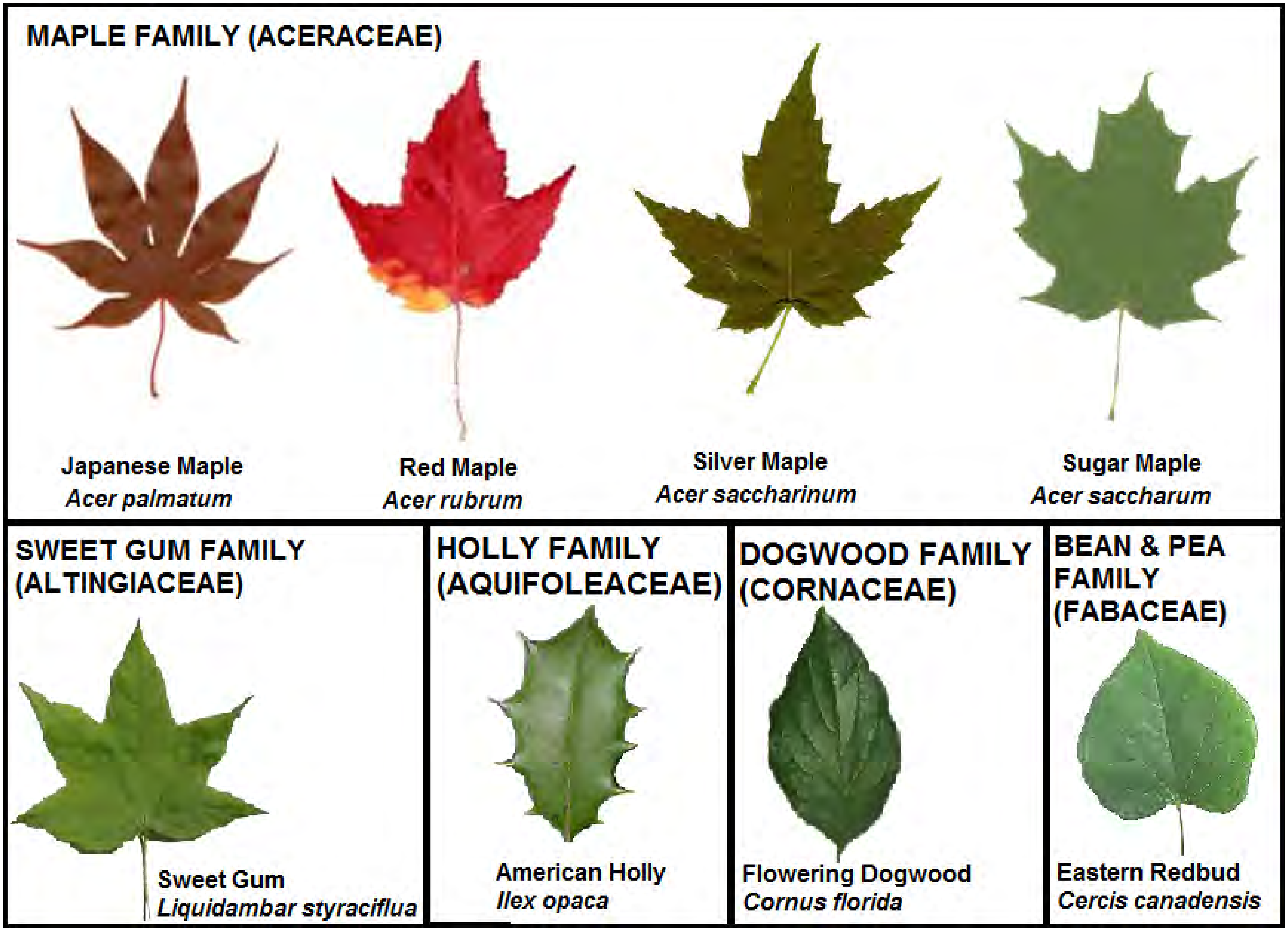
Q12. What is the factor by which the value of a quarter is greater than the value of a nickel?

$$\boldsymbol{c = 5}$$

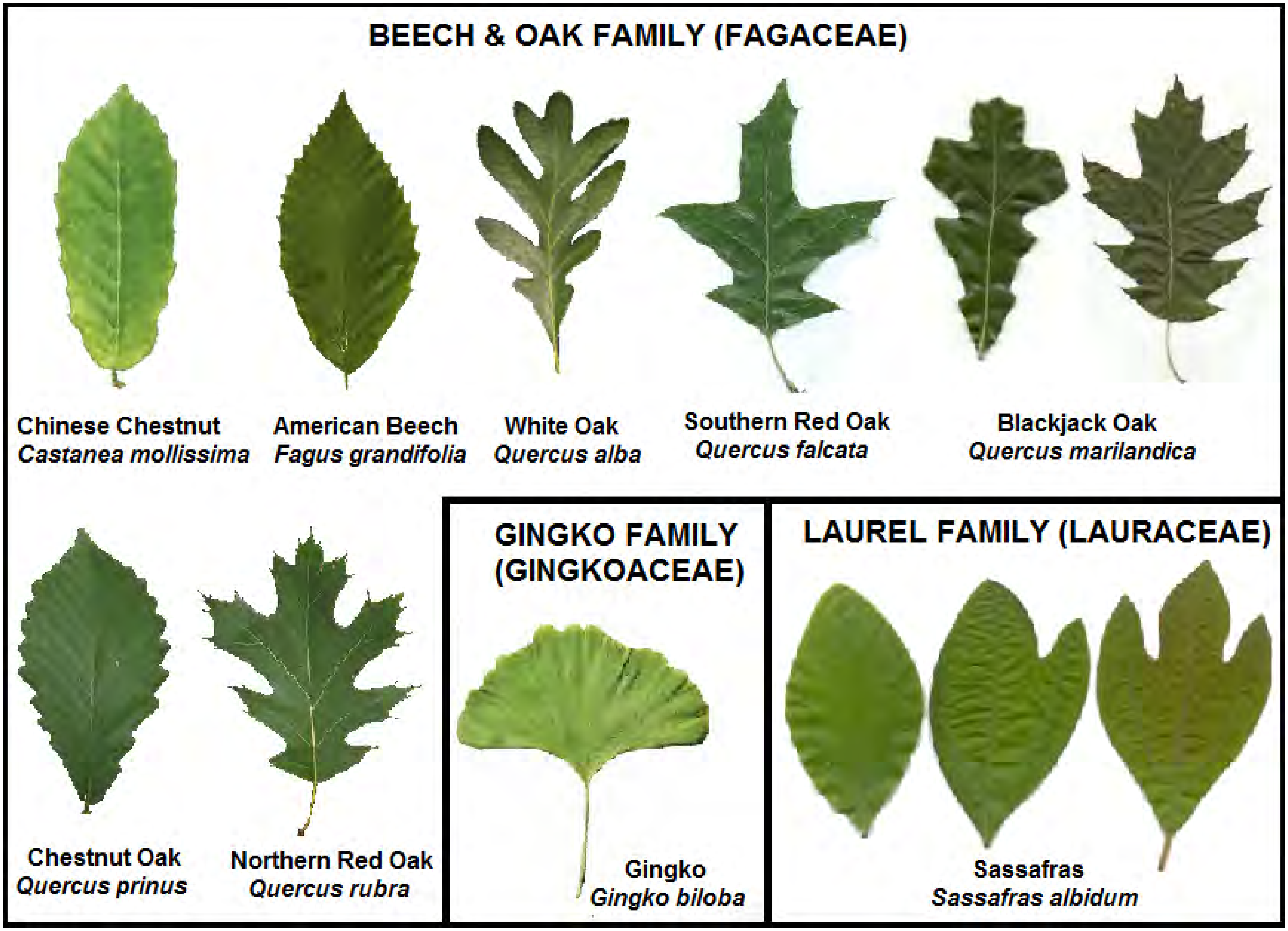
Q13. Complete the following sentence: If *x* is greater than *y* by a factor of *c*, then *y* is less than *x* by a factor of ___.

Using the answers to **Q9** and **Q10** above, we can see that the answer is $\frac{1}{c}$.

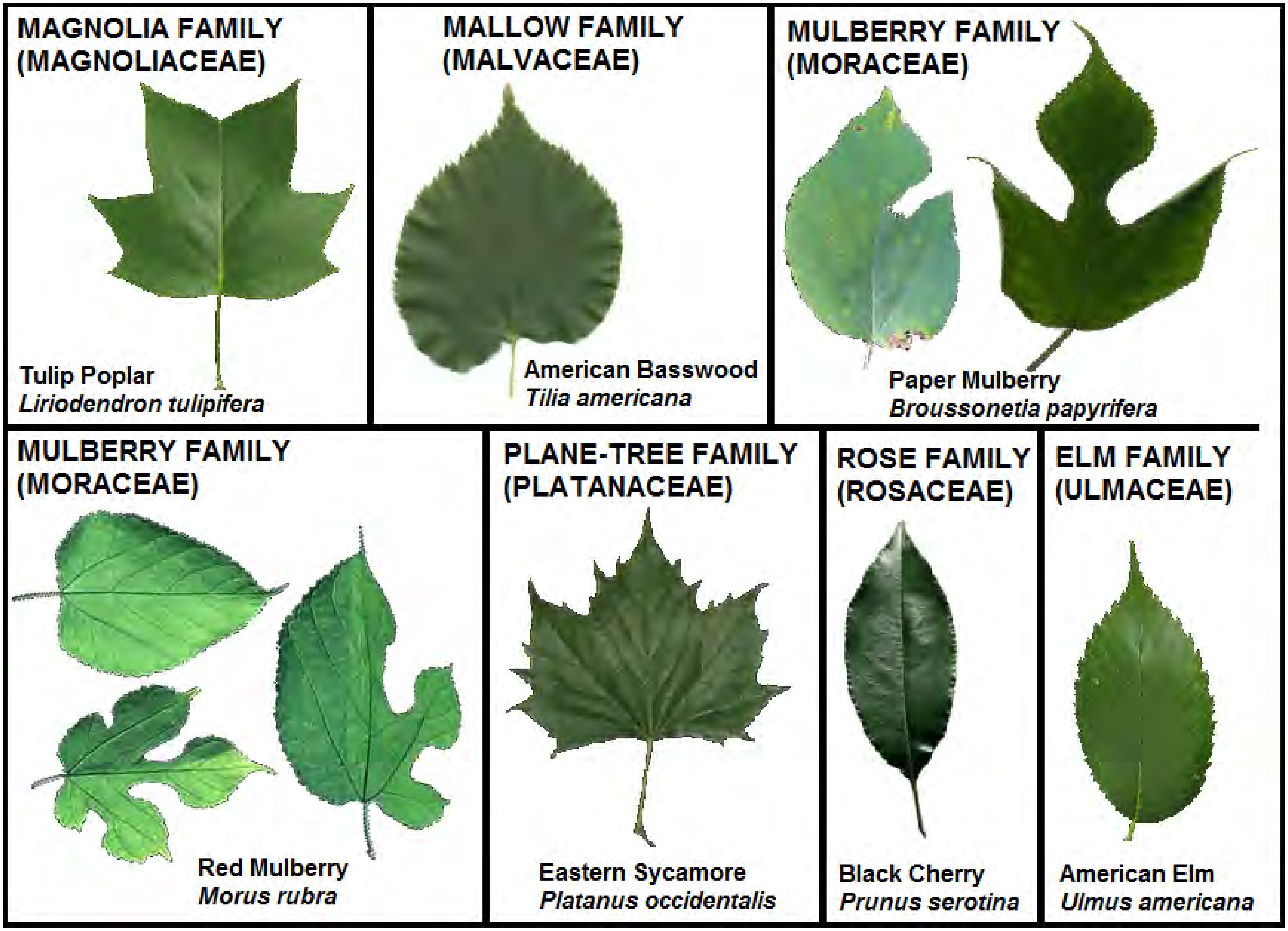
Guide to Leaves in *Biology in a Box* Unit 8: Everything Varies



Guide to Leaves in *Biology in a Box* Unit 8: Everything Varies



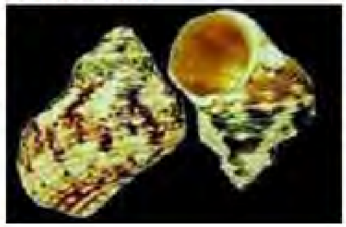
Guide to Leaves in *Biology in a Box* Unit 8: Everything Varies



Guide to Shells in *Biology in a Box* Unit 8: Everything Varies (Box C)

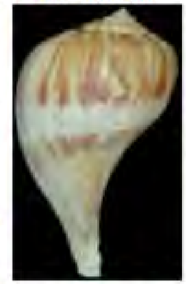
Moon snail (family Naticidae)

Turbo snail



Calico scallop

Pear whelk



Spindle (*Fusinus* sp.)

Rose cockle

Oyster

Flat scallop

Orange scallop

Nucleus scallop

Green limpet

Star limpet

Florida whelk

Fig whelk

Jingle clam

Turkey wing clam

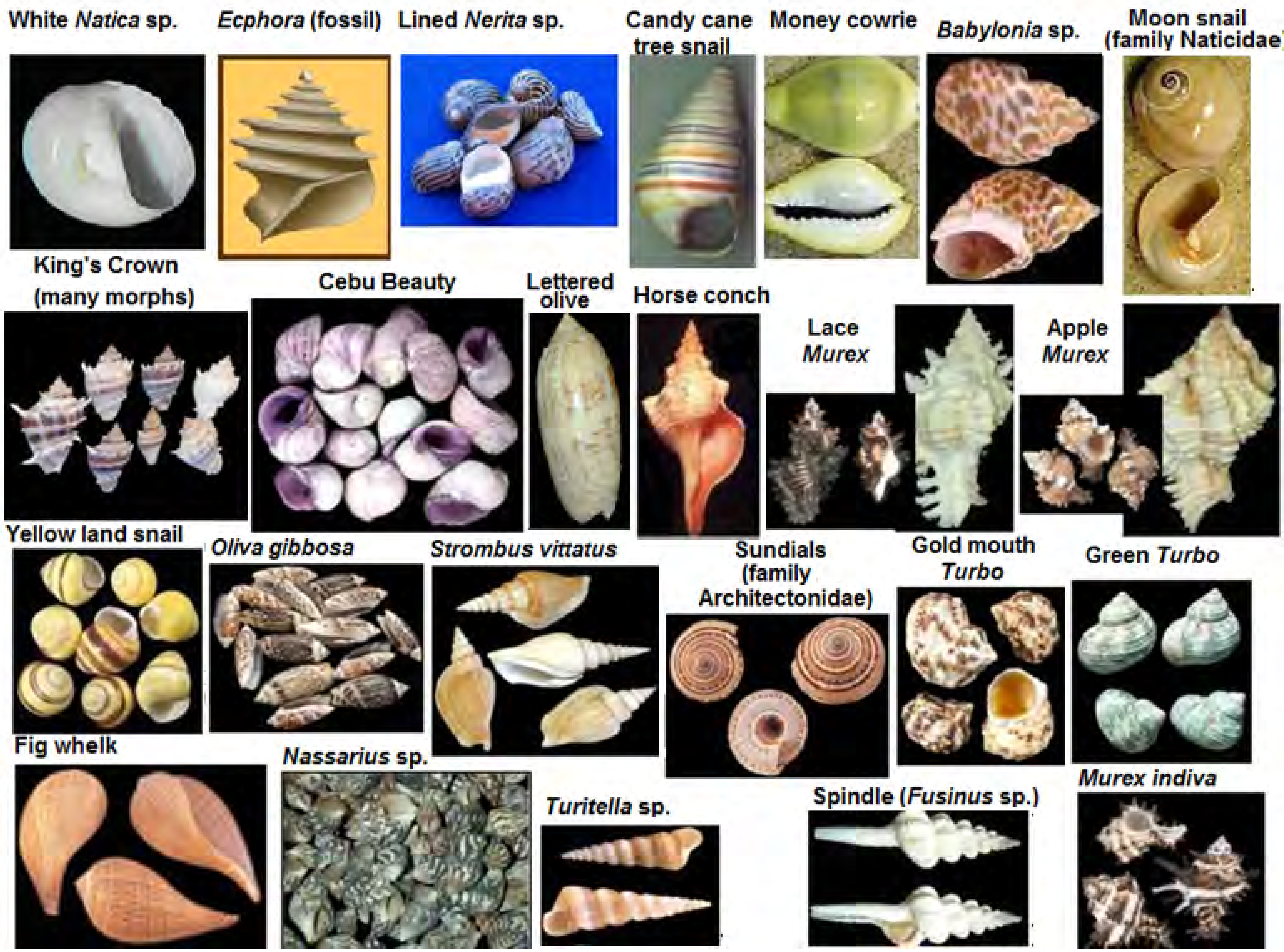
Auger shells

Turitella sp.

Murex indiva

Nassarius sp.

Guide to Shells in *Biology in a Box* Unit 8: Everything Varies (Box D)



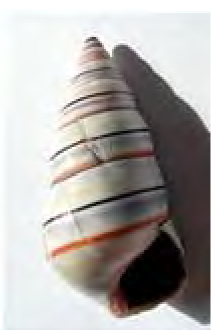
Guide to Snail Shells from Exercise 8

Snails in the genus *Nassarius* are saltwater snails that typically live in shallow waters. They are scavengers that feed on films of algae and detritus (decaying organic material). They are often seen in marine aquariums, where they help keep the glass clean.

Striped nerite snails are in the family Neritidae. These snails are found all over the world, but are most common in tropical areas. They can live in marine and freshwater habitats, and are most often found in intertidal areas, where they are often exposed to the sun. The majority of species in this family are herbivores that graze on algae, but some also have carnivorous diets, and eat larvae of flies on the roots of mangrove trees.

Species in the genus *Strombus* (the true conchs) live on muddy and sandy bottoms in marine habitats, and range from shallow waters to depths of over 160 feet. They are mostly herbivores, feeding on algae and marine plants. An unusual "leaping" motion is common in this genus, in which the conch digs the pointed end of their operculum (shell door) into the sand or mud, then thrusts its body forward with its muscular foot.

The lettered olive (*Oliva sayana*) is a species native to North America. This species typically lives near shore, on flat sandy areas, where they capture bivalves and small crustaceans for their carnivorous diet. This species is the state shell of South Carolina.



Haitian Tree Snails (*Liguus virgineus*, sometimes also called candy cane snails) live in trees, where they feed on fungi, algae, and lichens that grow on their host trees' bark. During dry seasons, they glue themselves to the bark of a tree and seal their shells with the slimy mucus produced by their bodies, and aestivate (pronounced ESS-tiv-ate), which is a term used to describe a period of summer dormancy (similar to hibernation in winter). When the rainy season comes again, the snails again become active.

Sundial snails (in the family Architectonicidae) typically live in warm, shallow waters, on sandy bottoms in marine habitats. They are found along the coasts of North and South America, as well as in the Pacific. Cnidarians, such as sea anemones and corals, are the preferred food of these predatory snails, which can be pests in marine aquaria.

Moon snails (in the family Naticidae) are a diverse group of approximately 300 species, which can be found in marine waters worldwide, even in the Arctic! They are highly predatory, and actively pursue bivalves, as well as other snails, including their own kind!

Land snails and slugs (formerly classified into the order Pulmonata, which is no longer considered a valid group) breathe with a single lung. Several species of land snails, such as the one in your box, are polymorphic, occurring in many different forms, or morphs. Some species vary widely in color, ranging from almost white to yellow, pink, green, and brown. They may also display great variation in the number and intensity of bands on their shells. Different morphs are more common in different types of habitats. The yellow form is most commonly found in grasslands, where they are more well-camouflaged from their bird predators. Most land snails and slugs are herbivores that eat grass, leaves, algae, and fruits.

SUGGESTED READING

<u>Grades K-3</u>

Find the Seashell - Liza Alexander
Backyard Explorer Kit with Leaf and Tree Guide - Rona Beame & Lionel Kalish (Illustrator)
The Little Yellow Leaf - Carin Berger
Seashells by the Seashore - Marianne Berkes & Robert Noreika (Illustrator)
Leaves Fall Down: Learning About Autumn Leaves - Lisa Marie Bullard & Nadine Takvorian
I Wonder Why Trees Have Leaves and Other Questions About Plants - Andrew Charman
A Simple Brown Leaf - L.J. Davis
Leaf Man - Lois Ehlert
Tell Me, Tree: All About Trees for Kids - Gail Gibbons
The Seashell Song - Susie Jenkin-Pearce & Claire Fletcher (Illustrator)
My Shell Book - Ellen Kirk
We're Going on a Leaf Hunt - Steve Metzger & Miki Sakamoto (Illustrator)
Fletcher and the Falling Leaves - Julia Rawlinson & Tiphanie Beeke (Illustrator)
Autumn Leaves - Ken Robbins
Temperate Decidous Forests: Lands of Falling Leaves - Laura Purdie Salas & Jeff Yesh
Swirl by Swirl: Spirals in Nature - Joyce Sidman & Beth Krommes (Illustrator)
Seashells (True Books: Earth Science) - Ann O. Squire
Look What I Did with a Leaf! (Naturecraft) - Morteza E. Sohi
Leaves - David Ezra Stein
The Leaf That Wouldn't Leave - Trish Trinco & Bryan Langdo (Illustrator)
I Can Name 50 Trees Today!: All About Trees (Cat in the Hat) - Bonnie Worth, Aristides Ruiz (Illustrator), & Joe Mathieu (Illustrator)
Fall Leaves Change Color - Kathleen Weidner Zoehfeld
What Lives in a Shell? - Kathleen Weidner Zoehfeld & Helen K. Davie (Illustrator)

<u>Grades 4-7</u>

Shell (Eyewitness Books) - Alex Arthur
Exploring Statistics in the Elementary Grades: Book 1, Grades K-6 - Carolyn Bereska, L. Carey Bolster, Cyrilla A. Bolster, Richard Schaeffer, & Rachel Gage (Illustrator)
Variation & Classification (Life Science in Depth) - Ann Fullick
Seashells in My Pocket: AMC's Family Guide to Exploring the Coast from Maine to Florida - Judith Hansen & Donna Sabaka (Illustrator)
Secret Lives of Seashell Dwellers - Sara Swan Miller
The Secret Spiral - Gillian Neimark
A Wasp is Not a Bee - Marilyn Singer & Patrick O'Brien (Illustrator)
The Tree of Life - Peter Sis
Tree of Life: The Incredible Biodiversity of Life on Earth - Rochelle Strauss & Margot Thompson
Animal Kingdom: A Guide to Vertebrate Classification and Biodiversity - Kathryn Whyman

Grades 7+

Compendium of Seashells: A Full-Color Guide to More than 4,200 of the World's Marine Shells - R. Tucker Abbott & S. Peter Dance

Shells (Smithsonian Handbooks) - S. Peter Dance

Evolution (DK Eyewitness Guide) - Linda Gamlin

Biodiversity - Dorothy Hinshaw Patent & William Munoz (Illustrator)

National Audubon Society Field Guide to North American Seashells - National Audubon Society

The Biodiversity Crisis: Losing What Counts (American Museum of Natural History Books) - Michael J. Novacek (Editor)

The Art of Shelling: A Complete Guide to Finding Shells and Other Beach Collectibles at Shelling Locations from Florida to Maine - Chuck Robinson

Seashells: Jewels from the Ocean - Budd Titlow

A Natural History of Shells - Geerat J. Vermeij

Data, Graphing, and Statistics Smarts! - Rebecca Wingard-Nelson

All Ages

Seashells of the World (Golden Guide) - R. Tucker Abbott, Herbert Spencer Zim (Editor), & George F. Sandstrom (Illustrator)

The World's Most Beautiful Seashells - Pete Carmichael, Leonard Hill, & Tim Ohr (Editor)

Peterson First Guide to Shells of North America - Jackie Leatherbury Douglass & John Douglass (Illustrator)

Seashells - Josie Iselin & Sandy Carlson

National Audubon Society Pocket Guide to Familiar Seashells - National Audubon Society

Scientific Journal Articles (included on Teacher CD)

Ayoub, N.A., S.E. Riechert, & R.L. Small. 2005. Speciation history of the North American funnel web spiders, *Agelenopsis* (Araneae: Agelenidae): Phylogenetic inferences at the population-species interface. *Molecular Phylogenetics and Evolution* 36:42-57.

Bruschi, P., P. Grossoni, & F. Bussotti. 2003. Within- and among-tree variation in leaf morphology of *Quercus petraea* (Matt.) Liebl. natural populations. *Trees* 17:164-172.

Cain, A.J. & P.M. Sheppard. 1954. Natural Selection in *Cepaea*. *Genetics* 39(1):89-116.

Cotton, P.A., S.D. Rundle, & K.E. Smith. 2004. Trait compensation in marine gastropods: shell shape, avoidance behavior, and susceptibility to predation. *Ecology* 85(6):1581-1584.

Hugall, A.F. & D. Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. *Nature* 485:631-634.

Ito, M. 2009. Variation in leaf morphology of *Quercus crispula* and *Quercus dentata* assemblages among contact zones: a method for detection of probable hybridization. *Journal of Forest Resources* 14:240-244.

Oxford, G.S. & R.G. Gillespie. 2001. Portraits of Evolution: Studies of Coloration in Hawaiian Spiders. *BioScience* 51(7):521-528.

Tang, X. & S. Huang. 2010. Fluctuating selection by water level on gynoecium colour polymorphism in an aquatic plant. *Annals of Botany* 106:843-848.

LINKS

Archerd Shell Collection – A web-based tour of the Archerd Shell Collection at Washington State University Tri-Cities Natural History Museum. Provides great info on the main classes of mollusk shells, shell development, torsion in gastropods, and more!

http://shells.tricity.wsu.edu/ArcherdShellCollection/ShellCollection.html

The Arthropod Story – Website from the University of California Berkeley's Museum of Paleontology; contains information on the history and diversity of arthropods, as well as symmetry.

http://evolution.berkeley.edu/evolibrary/article/_0_0/arthropodstory

Biology4Kids.com – A great site with information on invertebrate and vertebrate animals, as well as organisms from the other kingdoms of life. The subpages most appropriate for this unit are those for the invertebrates and vertebrates:

http://www.biology4kids.com/files/invert_main.html

Class Gastropoda (Snails and Slugs) - Biodiversity of Great Smoky Mountains National Park - a great website about the gastropod diversity found in our own home state, with lots of information on land snail importance, life histories, anatomy, and identification!

http://www.dlia.org/atbi/species/Animalia/Mollusca/Gastropoda/index.shtml

Real Trees 4 Kids – A great site sponsored by the National Christmas Tree Association with TONS of information on trees for kids, broken down into several pages based on grade level, as well as resources for teachers.

http://www.realtrees4kids.org

The Space Place: Sorting Out Trees in the Forest - Another awesome site from NASA describing how different tree species can even be identified from space with current technology!

http://spaceplace.nasa.gov/en/kids/eo1_1.shtml

eNature: FieldGuides - Excellent online field guide that allows you to enter your ZIP code to access field guides to plants and animals in your area!

http://www.enature.com/fieldguides

The Wonders of the Seas: Mollusks - A brief introduction on mollusk diversity from the Oceanic Research Group.

http://www.oceanicresearch.org/education/wonders/mollusk.html

Learner.org Life Science Session 5 - A great teacher resource and lesson ideas on variation, adaptation, and natural selection.

http://www.learner.org/courses/essential/life/session5/

BBC - Schools Science Clips - Variation - A fun interactive exercise on natural variation for kids ages 6-7.

http://www.bbc.co.uk/schools/scienceclips/ages/6_7/variation.shtml