Chapter 2. Introduction to Ecological Methods.

Outline of today's activities

1. Discuss ecological studies and statistics
2. Design an ecological experiment
3. Discuss Paine and Vadas 1969
4. Library visit (depending on need)

What you should get out of today's class

You should be able to articulate what ecology is, what kind of studies are used in ecology, and describe the reasons we use sampling, randomization, and statistics in ecology. You should be capable of finding peer-reviewed scientific journal articles on-line about topics that interest you.

Handouts

1. Chapter 3. Basics of Scaling in Ecology.
2. Damuth, J. 1981. Population density and body size in mammals. Nature 290:699-700.

Introduction

Ecology is the study of how organisms interact with their environment. Ecology is often separated into various sub disciplines that specify the organizational unit that is central to the sub discipline. For example, population ecology looks at how populations grow or decline, behavioral ecology focuses on what individual organisms do, ecosystem ecology examines the flows of resources through an ecosystem, and landscape ecology looks at how landscapes are patterned and formed. You might also encounter predator-prey ecology, restoration ecology, or microbial ecology! Ecology often focuses on a particular **level of organization** such as an individual, a population, or a landscape.

Types of ecological studies

There are four basic types of ecological studies: experiments, field studies, modeling studies, and meta analyses. The type of study you choose depends on the nature of your question, the system under consideration, and the amount of time and money available. By **system** we mean some set of organisms and abiotic resources, such as energy, water, and nutrients that interact to create organized patterns and behavior. Some types of studies lend themselves particularly well to certain questions, but to really understand a particular problem it is often necessary to have a little bit of all four types of studies at work. This semester we will work on labs that incorporate all of these types of studies.

Experiments are studies that manipulate a part of the system, leaving other parts of the system untouched, in order to find out how the manipulation influences the rest of the system. It is by replicated, random application of a **control** (no manipulation) and a **treatment** (some type of manipulation) that gives you confidence in your results. For example, you may want to know how a forest fire influences the **diversity** of flowering plants on the forest floor. In an experiment, one would identify the forest of interest, randomly pick some areas to burn and some areas to leave unburned, burn the treatment areas, and then go measure the diversity of flowering plants in the burned and unburned area. Because of the random assignment of treatments and controls, you could be sure that the burn really had the effect you measured. Experiments usually involve adding or removing something living to the system, introducing a non-living component such as a disturbance, or changing the physical parameters of the system such as its temperature.

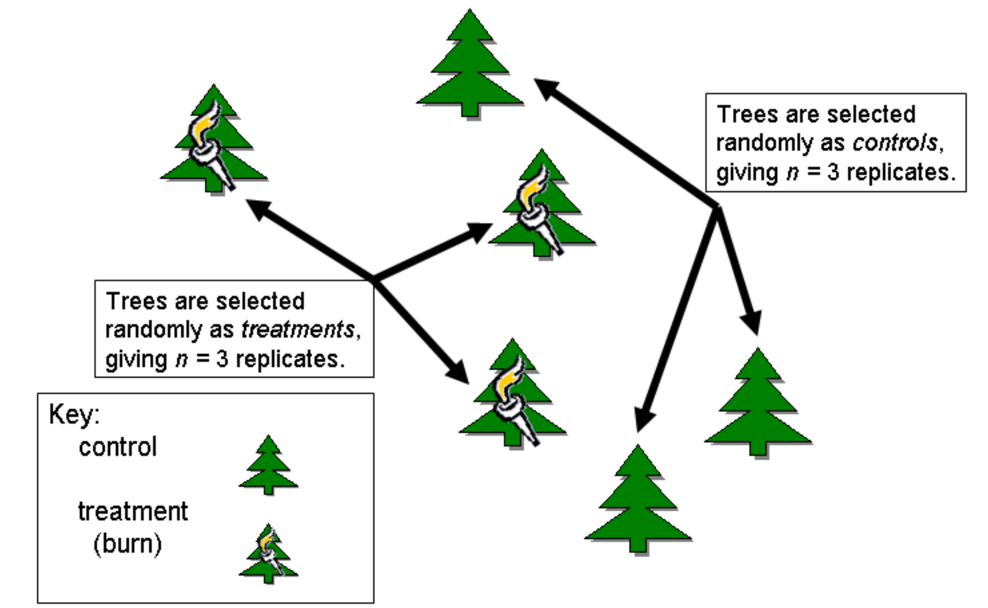

Figure 2.1. A simple experiment composed of a control and a burning treatment. Three trees were selected randomly to serve as replicate controls and another three trees serve as randomly assigned replicates of the burn treatment.

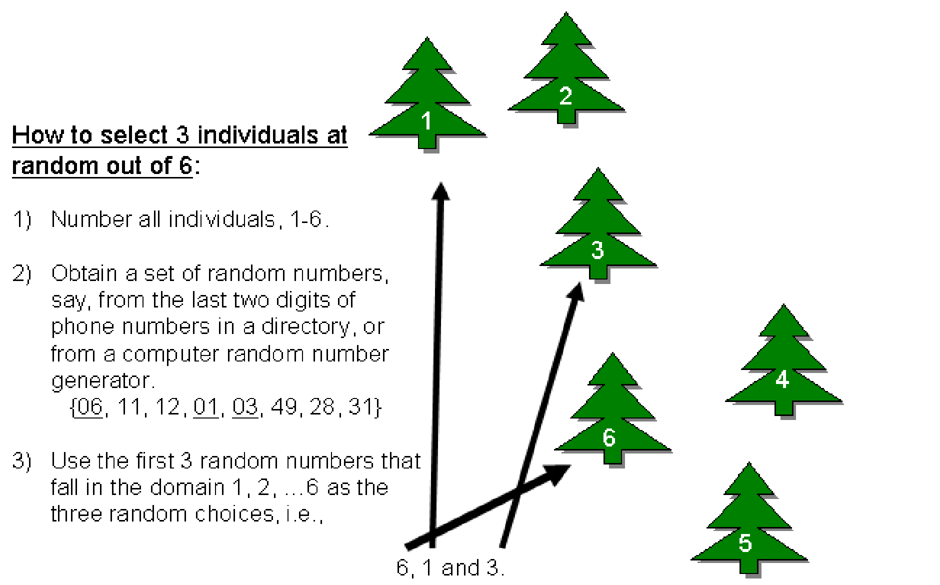

Figure 2.2. Method for choosing 3 out of 6 trees. The initial numbering of trees in step 1 is arbitrary. The procedure ensures that each tree has an equal chance of being selected.

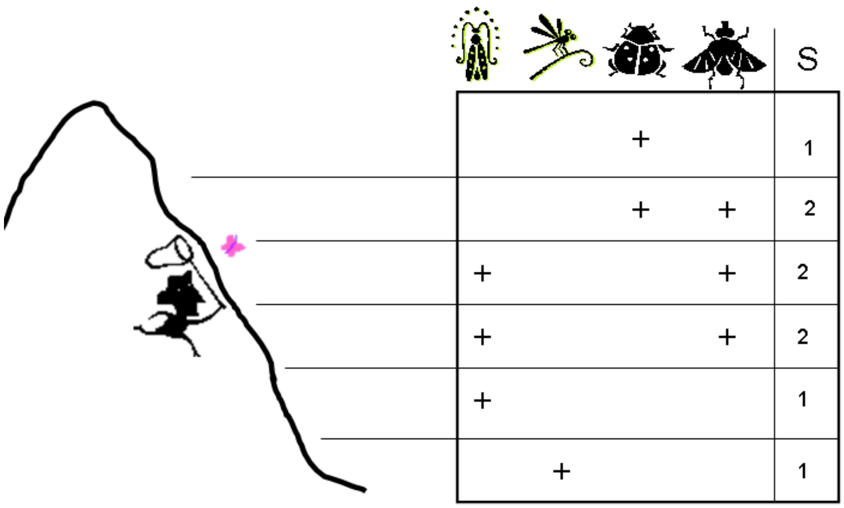

Figure 2.3. A field study to explore elevation preferences of several insect species and to characterize the pattern of species richness, S, with elevation. Notice that each species tends to favor low, mid, or high elevations.

Field studies occur outdoors in the natural environment, where investigators observe what is happening in nature. Field studies help us see what an ecological system is really like. For example, field studies tell us simple things like what species are present, where organisms occur, and what organisms eat. They also tell us more complicated things like the dispersion of habitat patches across the landscape or the dynamics of populations through time. A field study also can mimic an experiment. Consider the question posed in the paragraph above: What is the influence of forest fires on the diversity of flowering plants? In the real world you may not have the luxury of choosing forests to burn. Instead, you could sample many patches of forest that already have burned and many patches of forest that have not burned, and compare the diversity of flowering plants in each. It is possible that you will get the same answer as in the experiment above, but you could not with complete assurance say the fires were the cause of what you observed, because there could be some other variable, such as forest age or slope aspect, that caused the fires to burn where they did, and such factors could also be influencing the flowering plant diversity. Nonetheless, a field study like this would tell you what is really occurring on the ground, and therefore would provide useful information about flowering plants.

Modeling is a way of representing a complicated system in a simple mathematical way. All models involve two types of numbers: (1) **state variables** and (2) **parameters**. A state variable describes the condition, or state, of a system. For example, your state may be described by your location, represented by a state variable x, e.g., $x = 32$ km from Albuquerque. If you go for a walk, your location will change at some rate, dx/dt, which, being your speed, is a useful parameter that describes your motion. A model enables you to explore many important properties of the system based on a manageable set of state variables, parameters, and relationships. Typically, experiments or field studies are used to test whether the model correctly predicts how the system behaves. For example, perhaps you think that among the many possible factors involved, the number of fish in a lake determines how many offspring osprey (a predatory bird that eats only fish) raise in a year. You could make a model where the number of fish is multiplied by a parameter (in this case a number that represents how many fish are needed to make a young osprey) to give you the number of young produced. You could then go out and count the number of fish and the number of young osprey produced, plug the number of fish counted into the model and see if it correctly predicted the number of young produced. If so, you can say that you identified a useful dynamic of the system. If not, you may have to find more variables to include in the model so that it works better, such as how many osprey were competing for the fish, or how far away the osprey had to fly to get the fish.

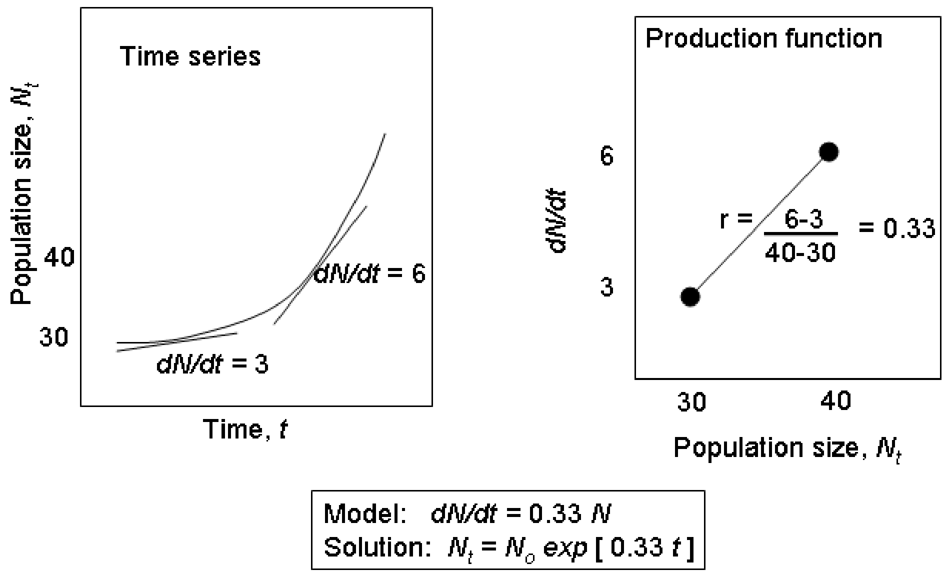

Figure 2.4. A model of population size, N. To display the time series (left panel, see Figure 2.8), we plot the number of individuals through time. The curve shows how population size increases upward. The slopes of tangents at selected times give the rate of change, dN/dt, at two particular population sizes, namely 30 and 40. The production function (right panel) examines how the rates of change increase linearly with population size. The key parameter of the model is r, the slope of the production function.

Meta analyses are studies that summarize data from other studies. They take values measured in a variety of places and times or of a variety of species and put them into a new data set, then answer their questions using the new data set. Usually, meta analyses involve compilations of field measurements to test a mathematical model. For example, you might be interested in finding out how biodiversity varies with latitude. To find out, you could compile biodiversity measurements from studies conducted everywhere from the tropics to the poles and evaluate the biodiversity relative to how far the study site was from the equator. You would then know something that you could not have known from each individual field study – biodiversity gets higher as you move from the poles to the equator!

Statistics and sampling

There is extensive variation in nature. When we measure a tree, count frogs, or quantify nitrogen in the soil, we are likely to obtain different values each time and place we do it. For example, there are millions of cottonwood trees in existence. Few of these cottonwoods will have the same height, diameter, or number of leaves. Therefore, if you measured some group of cottonwoods, you would have a range of values, from the smallest trees to the largest.

You can use a variety of **statistics** to give you an idea of what the measured cottonwoods are like in general. For example, the **mean**, or average, tells you the central tendency of a range of measurements. You could also calculate the **standard deviation** (the mean distance of values around the mean – a measure of dispersion), or the minimum and maximum values. Hence, statistics tell us something about a set of measurements in a concise way. Otherwise you would be stuck with a long list of numbers that would be very hard to get your mind around.

If you could measure the height of every cottonwood in existence, and you calculated the average of all those heights, you would have the correct value for the average height of cottonwoods. However, given that you have limited time and resources (this is always true in ecology), you have decided that of the millions of cottonwoods to choose from, you are only going to measure 50 of them. The average height of these 50 trees will give you an estimate of the average (not the true average, for which you would need to measure the height of all of the trees). Your estimate of the average may be very close to the real average, or it may be not so close. The primary determinant of how accurate the average is how the trees are selected for measuring. The average of one set of 50 trees is likely to be different from the average of other possible sets of 50 trees. So which 50 cottonwood trees should you measure? How do you choose trees so that the estimate of the average is as close as possible to the real average? The two main considerations in choosing your trees (or sample of measurements) are bias and inference.

Bias is the systematic movement of an estimate away from its real value. For example, if you had a tendency to pick shorter trees because it was easier to see their tops for measuring, you would calculate an average that was biased low. If you tended to pick taller trees because for some unexpected reason it was easier to get to them, then your average would be biased high. Therefore, you need to have some method for selecting trees that ensures that you have not unintentionally (we will assume that no one intentionally wants to introduce bias into their data set) created a biased estimate of the average. The tool we use for this is randomization, where with the aid of dice, a random number table, or a computer, your selection of individuals to measure is determined at random and therefore is not susceptible to unconscious choice by the scientist.

Inference is making a claim about a whole population based on the sample. The way you sample determines what kind of inference you can make. You start by choosing a **sampling frame** that is specific to the inference you want to make. If you want to infer the average height of all cottonwoods in existence, then you must randomly sample from throughout the entire universe of cottonwood trees. Or perhaps you just want to infer the average height of cottonwood trees is in New Mexico, in which case you would randomly sample from all of the cottonwood trees in New Mexico. (The New Mexico average would likely be different from the universal average – can you think of some reasons why?) Similarly, if you wanted to know the average height of cottonwood trees along the Rio Grande, you would sample from all of the cottonwood trees along the Rio Grande. You could continue to reduce your sampling frame to smaller and smaller units, and the level of your inference would be reduced as well. The bottom line is your inference only extends to your sampling frame, and attempting to infer to a larger, unsampled area usually leads to an error, and is therefore usually not a good idea.

The other definition of statistics goes much beyond concise representations of data such as the mean and the standard deviation. It is a vast and sometimes challenging field, but the important point is that statistics are a mathematical process that allows us to assess the likelihood that our conclusions are correct. Going back to the entire universe of cottonwood trees, we decide to compare the average height of cottonwoods that are 1 to 10 years of age with the height of cottonwoods that are more than 10 years old. Upon measuring every cottonwood in existence, we would undoubtedly find that the older cottonwoods have a larger mean height than the younger cottonwoods. Given that we have obtained the true means by measuring every tree in existence, we can compare the means and be confident in the difference we observe. But since we are only measuring 50 trees, let's say 25 in each age category, we only have estimates of the average height. Remember, if we took a different set of 50 trees, we would obtain two slightly different estimates of mean height. Because there is a possibility that, just by chance, we could obtain the same estimate of the mean for the two categories of trees (or even a higher one for the younger trees), we need to know how much confidence we should ascribe to our results. This is where statistical tests, probability distributions, and confidence intervals come in, and there are whole courses on these topics. But as an example, we will run through a Student's *t*-test in lab today.

Figures

Let us cover the names and briefly the construction of some commonly used graphing techniques in ecology. We cover these to get you started interpreting graphs in the papers you will be reading and discussing this semester in lab and lecture. You should get familiar with **time-series graphs**, **histograms**, **scatter plots**, and **bar charts**.

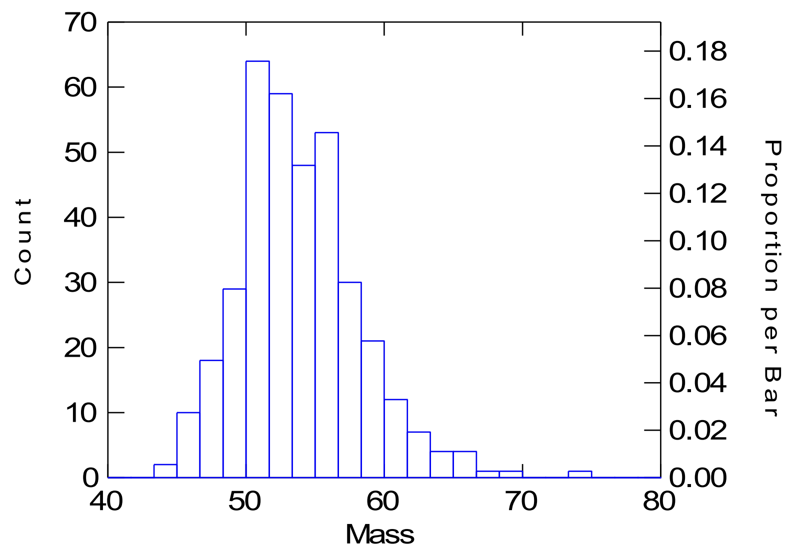

Figure 2.5. *A* ***histogram*** *depicts how often each value in a data set occurs. The y-axis shows you the "count" of the value in the data set, such as one 1, eight 2's, twenty 3's, and eight 4's. Data are often from a continuous variable (any value is possible along a continuum), but histogram programs usually pool data within a range (e.g., 0-1, 1-2, 2-3, etc.). Hence, a histogram shows you easily where most of the values are and what the spread of the values looks like. This histogram shows the distribution of values for the mass of flammulated owls. Note that most owls are 51-57 g or so, and a few owls weigh much more than the others, but not as many dip much below the majority range. This distribution is what is known as "right-skewed", as opposed to a "normal" distribution which has little skew in either direction.*

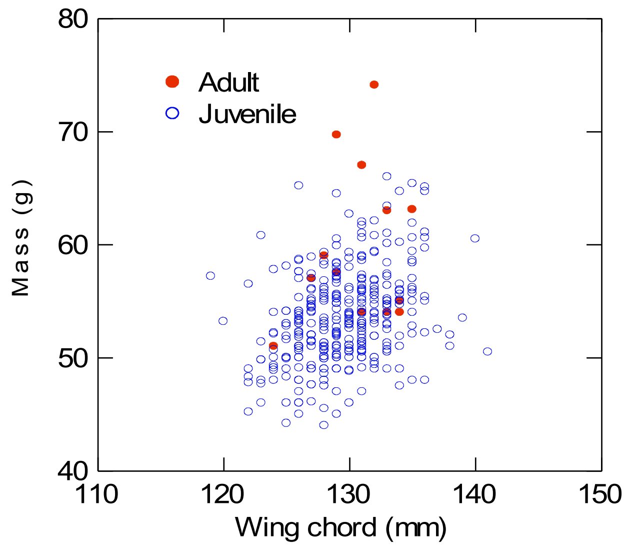

***Figure 2.6**. A **scatter plot** depicts the relationship between continuous variables. For a particular measured individual, you may have several variables, such as height or mass. One of those variables can be plotted on the x-axis and another on the y-axis, and even a third on a z-axis. Hence, a scatter plot shows how one variable changes with a change in another variable, and it gives you a sense of whether that relationship appears linear or curved. This scatter plot shows how the mass of flammulated owls increases with the length of the owl's wing – basically showing that larger owls (as measured by wing length) usually weigh more than smaller owls. You can even categorize the individuals so you can see how the relationship operates for different groups, such as age classes or species or any other class of interest. In this graph, adults (red, solid circles) are heavier than juveniles (open, blue circles).*

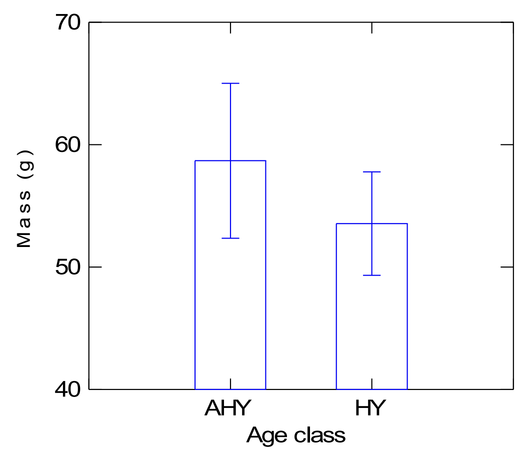

***Figure 2.7**. A **bar chart** depicts the means of different groups. The groups can be categories such as age or species, or it can be means through time or space. The standard deviation or standard error (the standard deviation divided by the sample size) is often portrayed with lines to show the spread of the data around the mean. Notice how these standard deviation marks indicate that there is much overlap in the masses between the AHY (adult) and HY (juvenile) age groups, but the means are different. You can see the same pattern in a different way in the scatter plot example.*

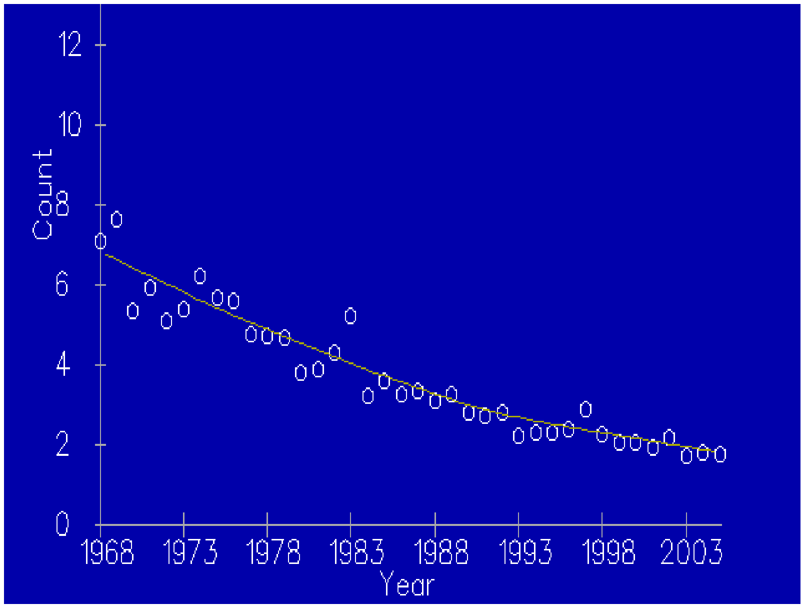

***Figure 2.8**. A **time-series** graph of the number of olive-sided flycatchers observed per breeding bird survey route in California. Time is on the x-axis and the count is on the y-axis. Although an estimated trend line is shown, time is not really an explanatory variable like number of offspring, food availability, or environmental variables. Rather, a time series shows what happens through time without ascribing the passage of time as the cause. From the USGS Patuxent Wildlife Research Center - http://www.mbr-pwrc.usgs.gov/*

An Illustration of the use of statistics in ecology (Student's *t*-test)

When we take samples from a population, we can derive estimates of the mean for the population that was sampled. Remember, we will not have determined the true mean, but an estimate. The estimate could be very close to the real mean, or fairly far away, just by chance (or because of bias, but we will ignore that for now). So when we compare sample means to each other for the purpose of determining whether the true means are different, we need a way to say what the chances are that our estimated means really differ, given the probabilities associated with the sampling. The *t*-test provides a way to do this.

We conduct a *t*-test using the following basic steps. First, count up the samples to determine the sample size. Second, calculate the mean. Third, calculate the variance, which is the standard deviation squared. Keep the samples separate and calculate the variance for each group. Fourth, use a formula to determine the *t*-statistic. Fifth, determine the degrees of freedom. And finally, compare the *t*-statistic with a table of *t*-values to evaluate the significance of the test.

The significance of a test, symbolized by the letter *p*, is a probability value that says how likely it is that the difference in means that you estimated was just a result of sampling error. So, if the significance of your test is 0.02, then there is only a 2% chance that, just by random sampling error, your estimated means would appear different when the true means really are not different. That is, there is a 2% chance that concluding that the sample means are different is wrong. That is a pretty low chance of being wrong. In fact, a rule-of-thumb threshold, or critical significance level, that many people use is 5%, although there is debate about how useful it is to set hard significance cut-offs such as 5%, 10%, or any other value. The point is that the smaller the chance of being wrong, the sounder your conclusions. By the time your significance level gets to 10%, you should recognize that there is probably a good enough chance that you are wrong to want to hedge a bit on your conclusions.

Step 1. Determine the sample size, known as *n*. This is just a simple tally of all the measurements you have made. If we measured 25 trees, then our sample size is 25.

Step 2. Calculate the mean, with the following formula, where the *X*'s are the measurements, and $\bar{X}$ signifies the mean:

$$Mean = \bar{X} = \frac{\sum X}{n}$$

Step 3. Calculate the variance, s^2, with the following formula:

$$Variance = s^2 = \frac{\sum (X - \bar{X})^2}{(n-1)}$$

Step 4. Calculate the *t*-statistic with the following formula:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

Step 5. Determine the degrees of freedom:

$$d.f. = n_1 + n_2 - 2$$

Step 6. Compare the *t*-statistic to a table of *t*-values. Run down the column on the left-hand side to find the correct degrees of freedom, then run along that row until you find the column with the highest *t*-value that your *t*-statistic exceeds. Then look at the top of the column to find the *p*-value.

Table 2.1. Use this worksheet to guide your calculations.

Sun leaves (*n* =)			**Shade leaves (*n* =)**		
Measurements	**(X - $\bar{X}$)**	**(X - $\bar{X}$)^2**	**Measurements**	**(X - $\bar{X}$)**	**(X - $\bar{X}$)^2**
10.5			14.1		
9.8			15.5		
12.3			16.6		
11.8			15.6		
11.7			13.8		
11.2			10.3		
11.3			12.7		
12.2			15.8		
8.8			17.1		
16.2			13.5		
$\bar{X}$:		**Σ**:	$\bar{X}$:		**Σ**:
Variance =			Variance =		

Table 2.2. Table of t-values.

	Probability, *p*		
d.f.	**0.100**	**0.050**	**0.010**
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878

Homework # 2 – Questions about Damuth 1981 (10 points).

Reading: Damuth, J. 1981. Population density and body size in mammals. Nature 290:699-700.

Read the article and answer the following questions. Just write your answers out by hand on this piece of paper. You may have to read ahead through the next chapter for help.

1. Which of the four types of ecological studies does Damuth's 1981 article represent? Justify your answer?

2. What was the “scaling exponent” of the density/body size relationship? What is the ecological significance of the scaling exponent being less than 1 and being negative?

3. Explain, mathematically, the result that energy used by a local population is independent of the body size of the individuals in the population. Look up the rules for managing exponents if need be.