

ECOLOGY Module I
Tree Sampling and Statistics
by Dr. Greg Capelli
College of William and Mary

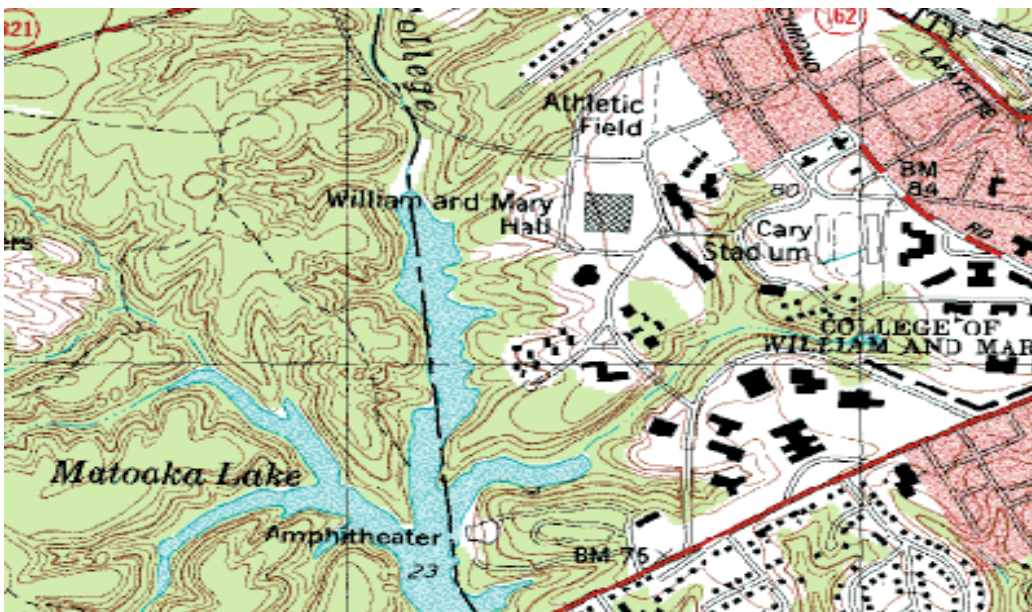
Module Goals

- 1) To gain a beginning understanding of the challenges involved in sampling populations**
- 2) To understand the basic ideas of statistical testing- assessing and interpreting data and extracting additional information from the data**
- 3) To learn about forest ecology and important ecological process (succession)**

WEEK #1

Biol204L

The College Woods is a preserve that surrounds a large part of the William and Mary Campus. The topological map shows the College Woods (green) in relation to the campus. The College has owned the land since the 1920s and it is a tremendous resource for the William and Mary Community. Several of the biology faculty conduct research and teach labs on Lake Matoaka and surrounding College Woods.



William and Mary Hall, USGS Williamsburg (VA) Topo Map

On week #2 of this course, you and your lab partners will conduct a field study in the College Woods. Your class will meet-up near William and Mary Hall and walk out to our field site with your lab instructor. Along the way, you will talk about the components of a forest ecosystem and learn about some natural history of the College Woods. Once at the field site, your Lab Instructor will provide specific instruction on tree-sampling techniques and efficient data collection. Most of the lab will be spent gathering data on the species type, number and circumference of individual overstory trees within the sample area. On the third week of lab, your lab class will return to Millington and perform several statistical exercises to estimate tree density and determine distribution patterns of various tree species within the College Woods.

In order to prepare you for your tree-sampling lab (week#2), your lab instructors will have a three laboratory exercises for you to complete this week. The exercises will help you learn to identify the dominant tree species in the College Woods. You will also learn some basics about forest stratification and recording field data for statistical analysis. You will also talk about details of the upcoming week's lab and learn about field safety. Although the tree-sampling lab is one of our longer labs, we anticipate that you should have time to return to your room and change your clothes before your next class.

WEEK #2

Biol204L

Ecology is the *scientific* study of the *interaction* of *organisms* with their *environment*. Another definition is simply the study of the distribution and abundance of organisms. These are ultimately one and the same, because distribution and abundance are determined largely by interaction with the environment. These seemingly straightforward definitions mask an enormously complex subject that is also of critical practical importance. Let's look more closely at key terms in the first definition.

When we think of how the *environment* affects organisms, physical/chemical variables usually come to mind first. Light, rainfall, or soil nutrients obviously might affect how and where a plant grows. Warmer temperatures or lower oxygen might limit the areas of a lake where fish can live. Beyond these *abiotic* variables, however, *biotic* factors are also extremely important. Other individuals or species are part of an organism's environment, and may compete with the organism for resources such as food or shelter, prey on it as a food source, or serve as a food source.

The word *interaction* implies an important consideration. At first, especially for abiotic factors, we may think primarily in terms of the environment as acting *on* and affecting organisms. But it is a two-way street: organisms by their very presence inevitably alter their environments, often significantly, for themselves and other species. A large amount of algae and plant growth may raise the level of oxygen in a lake substantially during the day, but metabolism by those

same organisms (and the lake's animals) at night may reduce oxygen to critically low levels, and lower the pH as well. Pine trees reduce the level of light on the forest floor, making it impossible

for their own offspring to live there. Often the effects of a biological change extend in complex ways through a whole community of organisms. For example, a change in the abundance of a predatory fish several steps up in the food chain may affect the abundance of its prey, which in turn affects the prey's food, and so on in what is known as a "trophic cascade"—ultimately altering the abundance of algae and even chemical nutrients in the water.

We've been speaking here in general of *organisms*, and have mentioned "species" and "community" as well. Let's now define some useful terms that refer to groups of organisms. Within ecology, as in much of biology, the concept of *species* is central. A general working definition of species is a collection of organisms potentially capable of interbreeding. Thus the collection of birds we call Blue Jays can interbreed with each other; those in New York could produce offspring with those in Virginia if given the chance. Blue Jays don't interbreed with Cardinals, so we call them different species. Although the concept of species matches well with much of what we see in nature, it is still to some degree an organizational construct that we humans impose on nature. This is simply to say that not all of what we see fits neatly into the usual definition. Some organisms usually or always reproduce asexually, and therefore "interbreeding" doesn't apply in the first place. And hybridization between two "different" species is very common, although often the reproduction is not as successful as with a member of the "same" species.

In any given area, the species present occur in smaller functional units known as *populations*. A population is simply a group of individuals belonging to the same species in a given area. There is a population of largemouth bass in Lake Matoaka separate from other populations of this species. With the bass, typically found in lakes, populations are physically isolated from other populations. With other species, distribution may be continuous over large areas, but we may still refer to different populations, as with the New York and Virginia Blue Jays. Broadly, when ecologists speak of "species" they are often actually referring to "populations."

Many important questions in ecology are focused primarily at the species level: what determines the number of species in a given type of habitat, how do species differ in how they use that habitat, how do species compete with each other for resources, how do we protect endangered species, etc. Specifically at the population level, basic questions regarding how and why populations change in size are central to much of ecology.

As a very broad generalization, individuals in the same species/population are typically ecologically similar: feeding on the same kinds of things, using the same kind of nests or shelters, responding to environmental variables in the same way, etc. In ecological terms, they occupy the same *niche*. However, in some cases there may be important ecological differences within the members of the same species. For example, small, young fish typically feed on different items than large adults. And sometimes males and females of the same species may be ecologically very different.

Moving up from the species/population level, ecologists often speak of a *community*, which can be defined as all the species in a particular area or habitat. We could speak of the Lake Matoaka community, or a grassland community, or a rainforest community. In this, its broadest sense, “community” would refer to all the plants, animals, bacteria, etc., in an area. However, ecologists often use the term in a more narrow ad hoc way, speaking of the amphibian community in a swamp or the microbial community in lake sediments, etc.

Biome is a term used in terrestrial ecology, referring to major regional communities. Biomes are usually defined by the major vegetation occurring in the region, often with names reflecting this: deciduous forest, grassland, coniferous forest. No parallel system exists for aquatic communities, which are usually referred to by terms based on major physical/chemical characteristics: fresh or salt water communities, lake, stream, or estuary communities, etc.

An *ecosystem* is simply a community—the organisms—in combination with the physical/chemical factors and processes. Simply put, it’s the biotic plus abiotic components. There is no size restriction: we could speak of a rotting log, Lake Matoaka, or the whole earth as an ecosystem.

A great many basic and applied questions involve the community or ecosystem level: why certain species are disproportionately important in their effect on overall community structure, how a toxic chemical is passed through a food chain, how much food the oceans can provide, the effect of global warming on the structure and function of coral reefs, etc.

Ecologists typically work at the population level or higher. However, it is important to recognize that an organism’s ecology is based on or closely connected to other aspects of its biology along a continuum that includes its behavior, physiology, genetics, and evolutionary history. So too with the questions that ecologists ask.

Scientific Method in Ecological Studies

There is one remaining term of great importance in our first definition: *scientific*. You’ve all learned about the “scientific method,” involving observation followed by hypothesis generation, then hypothesis testing through experiments involving controls and a variable of interest. The first important point is that not all “science” necessarily involves the “scientific method.” In fact, there are whole areas of science in which it is difficult or impossible to answer questions experimentally: geology and astrophysics, for example. In these areas very careful observation and inference account for much of what is known. So too with ecology, in which careful descriptive work—the kinds of species in an area, the food habits of a Robin, the timing of reproduction in trout, changes in the density of algae over time—is sometimes called “natural history.” Descriptive work provided the foundation of ecology, and remains essential to it.

But there is no denying the power of experimentation in the advancement of science. Unfortunately, however, bringing the “scientific method” to ecological questions has been a slow and challenging process. There are several related reasons for this, involving the complexity of the systems and the time and geographic scales often involved in ecological questions. A chemist can probably safely assume that a nitrate molecule behaves like a nitrate molecule whether it is high in the atmosphere or inside a beaker in a laboratory. Some ecological questions may be in the same vein: the amount of nutrients needed by a plant in a greenhouse may accurately reflect what it needs in the wild. But a bass or a squirrel is more complex than a single molecule or a plant, and may not feed in a lab aquarium or cage as it would in a lake or forest. There would be even less reason to think that several fish species placed together in a lab would interact normally. Essentially, it’s often difficult to duplicate natural conditions in a lab, and then to carefully manipulate the variable of interest while controlling everything else.

It is also difficult to take the experimental approach out to nature. To illustrate, let’s imagine two contrasting experiments, both of which are *conceptually* simple. Let’s say you know that the hormone thyroxine is involved in animal growth, and that it contains iodine. You hypothesize that laboratory mice deprived of iodine in their food will grow more slowly than those on a normal diet. You set up a simple experiment in which similar sized young mice are weighed, then kept under identical conditions except that some have a normal diet (the control group) and the others are fed the same food but with iodine eliminated. At the end of three weeks, you reweigh them all, and judge whether your hypothesis was supported. Simple enough. Now assume you’ve noticed squirrels in the forest burying acorns in the fall. You suspect that the squirrels don’t always find them all again, and that this burying process is important to seedling germination and ultimately to maintaining oak trees in the forest. You’d like to test that hypothesis. Conceptually the experiment is simple: find a bunch of similar tracts of forest, set up a system to keep all the squirrels out of some, and come back in a hundred years or so to see what has happened to oak populations in the two forest groups. But obviously the logistics of this pose serious problems. Nonetheless, despite the difficulties, ecologists have sometimes been able to use certain natural systems in experimental ways, providing extremely valuable information about basic ecological processes. Still, many of the most pressing ecological issues, such as the impact of global warming, are very difficult to address experimentally.

Because of the difficulties of doing actual experiments, ecologists have often turned to mathematical modeling of ecosystems and processes. Bits and pieces of descriptive information, or data from “do-able” experiments, are put together to describe a larger process of interest. For example, in a given community, there may be data on how light, temperature, moisture, and nutrients affect plant growth, how much plant material is consumed by herbivores, how fast the herbivores grow, how many of the herbivores are consumed by higher-level predators, etc. These

could be linked in a mathematical model that theoretically describes how a change in one variable would affect the others. Then an “experiment” can be done on a computer involving, say, how global warming might affect the number of higher level predators, given that warming = higher

temperatures and more cloud cover (due to more evaporation) = less sunlight for the given system. Of course, the possible changes in temperature and cloud cover are themselves derived from models. Much ecological insight, both theoretical and applied, has been gained by modeling. Obviously, however, this approach is only as good as the assumptions and data used to build the model, and for many ecological questions these remain major limitations.

There is an important upshot to all this. In other sciences, including most other areas of biology, there are typically “laws,” “principles,” “rules,” (or whatever they may be called), both large and small, that describe pretty accurately how things work, pretty much all the time: acceleration of an object due to gravity, the number of electrons found in a 2p orbital, the volume of a gas as a function of temperature and pressure, expected genotypes and phenotypes in F_1 and F_2 generations, etc. But this is less the case with ecology. There are of course fundamental principles that apply universally to ecological processes. For example, the Second Law of Thermodynamics describes what happens to energy as it transforms from one kind to another, and applies equally well to distant stars or food chains. But this is a law of the universe, not specific to ecology. When it comes to genuinely ecological-level questions, the answer is often more in the vein of “sometimes this, sometimes that.” How will additional food affect this population? Ans: it will allow it to increase—unless, at the higher density, disease gets a foothold and spreads fast, so the population actually crashes. How does the amount of plant growth in a community relate to the total number of species in a community? Ans: Sometimes with more plant growth there are more species, but sometimes fewer.

This is not at all to say that ecology is a jumble of random observations lacking in any unifying principles. Quite the contrary. There are a great many well-supported concepts. Often, however, many operate at the same time in a given situation, and because it is difficult to know the relative “strength” of each, it is hard to predict the outcome, or the effect of a change in one of them. Even when ecologists *can* predict with reasonable certainty what is likely to happen and why, the lack of absolute certainty is often exploited regarding important applied issues. Just as the tobacco industry resisted regulation for decades because there was no “proof” that smoking is bad, so too those opposed to various environmental regulations often make the case that proof is lacking regarding the harmful effects of whatever is in question. The most notable example in recent years involves President Bush, who continues to insist that there is insufficient proof of global warming or of what the effect of any warming will be.

Estimating Tree Density in the Northwest Sector of the College Woods.

To get the most accurate determination of the composition of the college woods, you could go out and count all of the trees in a given acre. Although it would prove accurate, it would also be

very expensive and time consuming. A more practical way to do it is to estimate the composition and forest density.

You will use two different survey methods in your lab to sample trees at the field site.

1. Quadrat Sampling

For terrestrial organisms that don't move around—meaning, most often, plants—fairly simple and straightforward methods can be used for estimating density. A common one involves quadrats, which are simply appropriately-sized defined areas located at random in the study area. All the organisms of interest that occur within the quadrat are counted, and these data points in combination with the area of the quadrat are used to calculate an estimate of the overall density of the organism in the area. Smaller quadrats may be delineated by use of rings or squares carried to the sampling area; larger ones can be delineated as needed on site by use of a tape measure.

The size of the quadrat is important, and should be matched approximately to the expected density of organisms. Obviously you don't want a quadrat so large that it's cumbersome to count all the organisms within it—that defeats the purpose of using it in the first place. And you don't want one that is so small as to rarely “capture” individual organisms. For both practical purposes, and for some statistical considerations that we won't discuss here, quadrats should generally encompass “a few” of the organisms, with perhaps some occasional “zeroes.” For our tree work, we will use quadrats 10 m X 10 m.

It is ***crucially important*** to getting good data that the quadrats be placed ***at random*** in the sampling area. “At random” does ***not*** mean casually looking around, placing one “here”, then maybe saying to yourself “how about over there,” etc. It is all too easy, consciously or unconsciously, to be biased in where quadrats are placed. For example, if you haven't been getting a lot of organisms in your recent samples, and you notice an area where they seem to be more abundant, you might have a tendency to think you should place a quadrat there.

So, before sampling, you must decide on a method that will place the quadrats at random. How to do this? One useful item that often comes in handy when you need to do something randomly is a table of random numbers, typically included in the appendix of statistics books, and readily available on the internet. It is simply, as the name suggests, a bunch of numbers (0-9) arranged at random in columns and rows. If you impose on this table *any* regular, consistent pattern for selecting numbers, the numbers generated will be random. Simply going in sequence straight across a row or down a column will generate random numbers. Or you could start at the ninth row and pick every third number. Whatever—just so it's consistent. The table, of course, needs to be adapted in some specific way to the work at hand. Your lab instructor will demonstrate how best to use the table.

For our tree sampling, we will use transects, which are simply straight paths through the sampling area. At various points along the transects, we will place quadrats. The distances at which the quadrats are placed will be determined by selecting numbers from the table. Measuring and laying-out the quadrats is quite simple. From your random point along your transect, place a flag. Next, measure off 10 meters to the right and perpendicular to the transect vector. You will

place a flag which is your second corner of the transect. From your flag, you will measure 10 meters running parallel to the transect vector and place a flag to become your third corner of your transect box. From the third corner of the transect box, measure off 10 meters perpendicular to your transect vector and place a flag. The flags define the four corners of your quadrat.

The sampling will be done on those trees inside your quadrat. Using the datasheets, have two team members recording the species identity, location and dbh of all of the trees greater than 15cm. Other team members can record the identity and location of any saplings and the composition of the shrub and debris layers.

You will sample at least three random quadrats along your transect.
(*POINTER*) *DO the next method before moving o a new location.*)

2. Point-Quarter Sampling

As you will see, delineating the fairly large quadrats needed for tree sampling is somewhat cumbersome and time-consuming. We will also use another method which is faster and easier, though not at all immediately intuitive as to why it should work. It's called the "point-quarter" method (or "point quarter center method" or point-centered quarter method"). Select, as before, random points. At each point, divide the immediate area into four equal *quadrants* (that's quadrants, not quadrats). Within *each* quadrant, find the closest tree and measure the distance to it. Continue on to other random points and do the same.

When finished with the field work, simply find the mean value for all measurements combined. This value is an estimate of the mean distance between trees, and when *squared* is an estimate of the *average area* that each tree "occupies." (Note: the explanation for this involves moderately complex math with calculus, so we'll just take it on faith). Then, if you know the average area occupied by each tree, it is a simply a matter of dividing this into whatever area unit you want to use, to get the number of trees per that unit.

Determining Abundance: Just count them all—but what if you can't?

Given the second definition of Ecology, as the study of distribution and abundance, an essential part of many ecological studies involves determining the number of individuals in the population(s) of interest. As you might expect, this is typically not easy to do, for all kinds of potential reasons (mobility of the organism, difficulty of working in the given habitat, and many other considerations). And obviously methods will vary: determining the number of fish in a lake requires very different methods than those for trees in a forest.

But before any discussion of specific methods, there are some *absolutely fundamental*—and quite simple—ideas that you need to understand. Consider, say, a 100 hectare area of woods. You wish to determine the average number of trees per hectare. How can you do this? There is *only one way*: **count them all** and, in this case, divide by 100. If you do *not* count them all, and someone asks you what the average number of trees per hectare is, there is only one honest answer you can give: “***I don’t know.***” Similarly, if you want to know the average number of crayfish per m² throughout a large lake, you must **count them all**. Otherwise, you simply *can’t know*.

Of course, “counting them all” is usually not a real possibility. And of course ecologists don’t simply throw up their hands and walk away. Instead, as you would expect, some kind of sampling is done in a limited way, such that only some of the individuals in the population are actually counted, within limited areas. This information is then used to calculate an average number of individuals/area. But where does this leave us? If asked at this point “so what is the number of trees per hectare in this forest,” what would you say? The only answer is “I still ***don’t know***, but I do have an ***estimate.***” What’s the *next* question? We really don’t care about the *estimate* per se, except as a *guide* to the *real* average (or mean) number of trees/hectare. So the obvious next question is “how good is the estimate,” i.e., how close is it to the *actual* mean number of trees/hectare. And what is the answer to *this* question? There is only one: “***I don’t know.***” The only possible way of knowing for sure how good the *estimate* is would be to count *all* the trees to get the *real* number, and compare the estimate to that. Of course, if you counted all the trees, you wouldn’t need the estimate in the first place!

So we seem to be stuck. But there *is* a way to attach *some* meaning to our estimate, through the use of *statistical* analysis of the data. Statistical analysis can be done for all sorts of reasons in all sorts of ways, but at its most fundamental level it involves describing the relationship between our *estimate* of something and the *real* value in terms of *probability*. After doing an appropriate statistical analysis of our tree data, we might again be asked “so what is the mean number of trees/hectare.” This time the answer would be: ***I still don’t know—and I never will know***—but I do have an ***estimate***, and the ***true mean*** is ***probably*** (but not definitely) within this range of the ***estimate***.

To get a beginning understanding of what some statistical tests involve, we can start with some very simple, intuitive ideas. Let’s say you’re interested in determining the mean number of

petals on the daisies growing throughout a large field. You collect five plants at random and count the petals on each. Here are two hypothetical sets of results:

	<u>#1</u>	<u>#2</u>	<u>#3</u>	<u>#4</u>	<u>#5</u>	
Number of Petals:	32	79	20	67	52	x = 50/plant <u> </u> = ?
Number of Petals :	51	48	52	50	49	x = 50/plant <u> </u> = ?

In both cases the *estimate* of the mean (symbolized by \bar{x}) is the same: 50 petals/plant. We know it's unlikely that a random sample of five plants will have a mean that matches the true mean *exactly*. But just how close is this *sample mean* to the *true mean* (symbolized by μ)? Well of course we don't know. But we notice that values in the first set seem to be all over the place, whereas those in the second have little variation. With which set would you be more confident that the true mean is actually close to 50? Now, let's say you somehow did go out and count the petals on *every* plant, and found that the true mean was 65, which is fairly far from the estimate of 50. That would be no surprise for the first set of data. And a true mean of 38 wouldn't surprise either. On the other hand, it *would be* surprising to us if the true mean for the second set of data were this far from our estimate—we intuitively feel that this is *unlikely*. Put another way, we don't have much confidence with the first set of data that the true mean really is close to 50, and if we had to guess where it lies we would probably give a fairly large range around 50. With the second set of data, however, we are much more confident that the true mean is fairly close to 50, and our best guess about where the true mean actually lies would involve a smaller range.

Here's the essential point. A first step in assessing the accuracy of an estimate is to look at the *variation* in the data that produced the estimate. One formal way of doing this is to calculate the *variance*:

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

where: X = an individual sample value
 \bar{x} = the mean
 n = the number of samples

In words, you find the difference between each sample value and the mean, then square that difference, then add up all such squared differences, then divide by one less than the number of samples. We won't worry here about the rationale for this formula, but you can see that when the samples vary a lot from the mean, you'll tend to get a large number in the numerator (and in the formula overall) as these differences are squared and then added. If the variation is low, there will be small differences from the mean, and the resultant value will be much lower. In many statistical applications what's actually used is called the *standard deviation*, which is simply the square root of the variance. Standard deviation is symbolized by s , and therefore the variance can be symbolized by s^2 , as above. We don't want to worry here about math details or formulas, but the variance is both very basic and easy to remember, so you should know it from memory.

In the first daisy example above, s^2 would be calculated as:

$$\begin{aligned} \frac{(32-50)^2 + (79-50)^2 + (20-50)^2 + (67-50)^2 + (52-50)^2}{5 - 1} &= \frac{18^2 + 19^2 + 30^2 + 17^2 + 2^2}{4} \\ &= \frac{324 + 361 + 900 + 289 + 4}{4} \\ &= \frac{1878}{4} = \mathbf{470} \end{aligned}$$

For the second set of data, the variance can be easily calculated mentally as

$$1 + 4 + 4 + 0 + 1 = 10, \quad 10 \div 4 = \mathbf{2.5}$$

So we see here that one data set has a large variance, the other one a small variance. But what constitutes “large” and “small”? There’s no formal specific answer to that, but we can illustrate a problem that the question raises. Consider two simple sets of data:

$$1, 2, 3, 4, 5 \quad x = \mathbf{3} \quad s^2 = 4 + 1 + 0 + 1 + 4 = 10, \quad 10 \div 4 = \mathbf{2.5}$$

$$95, 98, 100, 102, 105 \quad x = \mathbf{100} \quad s^2 = 25 + 4 + 0 + 4 + 25 = 58, \quad 58 \div 4 = \mathbf{14.5}$$

The second data set has a higher variance. Does that mean those data really are more “variable?” Intuitively that doesn’t seem right. If we consider the first data set, we see that some values are, relatively, much higher or lower than the mean (1 is only .33 of the mean of 3, 5 is 1.67 times larger than the mean). With the second set, the values are proportionately much closer to the mean, and we intuitively think of that as less variable. Therefore a “stand-alone” variance is meaningless; there must be some reference to the mean for interpretation. One quick way to do this is to simply calculate the variance/mean ratio (s^2/x). For the first data set this would be $2.5/3 = \mathbf{.83}$. For the second set it’s $14.5/100 = \mathbf{.145}$. So we see that variance/mean corresponds better to what we intuitively think of as higher or lower variation.

Another factor that is intuitively understandable as important in assessing an estimate is the *number of samples*, or data points. With our daisies, 5 samples is pretty small. Even with the second set, where all values are close to 50, the small sample size reduces our confidence about where the true mean lies. But if we had 20 samples, all similarly close to 50, we would feel much more confident that the true mean is quite close to 50.

We will use a simple statistical test or two to assess some of our data. Some of you no doubt have done statistical calculations, probably involving either a chi-square or t-test. However, it is

not important at this time for you to understand the mathematical theory behind statistics tests, or to know formulas and which test to use when. But: statistical testing shows up everywhere—in the scientific literature to be sure, but also every day in newspapers and other familiar sources regarding such things as election polls or an increased health risk associated with a certain drug. Although it is not important that you understand the details of the statistics, it *is* important that you understand some basic terminology typically used to describe the “significance” of the results.

We’ll focus on one of the most common kinds of statistical testing, involving the difference in mean values from two different groups. Perhaps you are interested in whether there is a lower incidence of auto accidents among teenagers in towns that ban cell phone use while driving for this age group, vs. towns with no such law. Or whether there are more pine trees on one side of a ravine vs. the other, or whether woodpeckers on one island have longer beaks than those on another. In each case you would use appropriate methods to take a random sample of each group, and calculate the mean for each.

We are of course now confronted again with exactly the same problems discussed before. We want to know whether there is a difference between the *true mean* of one population and that of the other. But we have only *estimates* of those means. Given that these estimates are from random samples, we know it is unlikely that they are going to be identical, *even if* the true means of their respective populations are essentially identical. So the job is to assess what the difference in sample means means ☺. One possibility is that the true means are identical or very close to each other, and the difference we see in our two data sets can be explained as arising from random differences in our samples. The other possibility is that the sample differences reflect a difference in the true means. A statistical test helps us distinguish between these two possibilities. In this case, involving two means, something known as a *t test* can be used.

Here’s the formula involved in use of a *t test*:

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{\text{var}_T}{n_T} + \frac{\text{var}_C}{n_C}}}$$

where *t* and *c* refer to two different groups (“treatment” and “control”), *x* refers to their means, *var* to their variances, and *n* to the number of samples from each. The “t value” is then used to determine a probability that helps us understand what our means mean (described below). You **do not** need to memorize this formula. But **note**: consistent with what we’ve said so far, it incorporates the key considerations we’ve described: the means of two different samples, their variance, and the number of samples in each. You **do** need to understand well the following.

You can think of a statistical test, or the statistician using the test, as beginning with a basic assumption, called the *null hypothesis*. In words, this can be stated as “there is *no reason* to believe there is a *difference* between the two groups; the difference in sample means can be explained *as random variation* among a limited number of samples, and *does not imply a difference* in the true means.” A statistical test provides evidence in support of or against the null hypothesis, using the following general ideas that are consistent with everything already discussed. If there *is* a real difference in the true means and it is large, it is less likely that the sample means are going to be close to each other. Or conversely, the larger the difference in sample means, the more likely it is that the true means differ. Of course, variation must be considered: two sample means that differ, with each having only low variation, reflect a greater likelihood that the true means really are different than would be the case if each sample had lots of variation. Similarly with sample size: larger samples make it more likely that smaller differences in sample means reflect real differences.

A statistical test then takes the data and cranks it through a formula, with an outcome that is ultimately couched in terms like this: “the populations differed significantly, $p = .03$ ” or “the populations did not differ significantly, $p = .17$.” You *must* understand this terminology. Here’s the idea. Assuming the null hypothesis of no difference in the true means *is* correct, it is still not likely that the sample means will be identical. But small differences suggest no reason to reject the null hypothesis. The larger the difference, however, the more likely it is that the null hypothesis *isn’t* correct, and that the sample difference reflects a *true* difference. In this case the null hypothesis should be rejected in favor of the idea of a real difference. The probability given with a statistical test gives us guidance in this regard. Specifically, it is the ***probability of getting a difference as large as that observed in the samples if the null hypothesis is true***. As this probability becomes lower, it is increasingly likely that the null hypothesis *isn’t* correct, and therefore it is increasingly reasonable to reject it in favor of the idea that the difference is real. Of course—one more time—we can’t know for sure; we have only probabilities to guide us.

So at what probability level is it reasonable to reject the null hypothesis? There is no simple or definitive answer to this. However, much of the time the kinds of questions being asked are couched in such a way that is important to know fairly certainly if a difference really exists. Therefore the critical probability level is set rather low, meaning that there needs to be a high likelihood that the null hypothesis isn’t correct before “accepting” the idea of a difference. The most common probability “setting” is at $p = .05$. A statistical test that gives us this value is saying that if the null hypothesis is correct, 95% of the time random samples from the two populations would not differ as much as we observed, or conversely that only 5% of the time would a difference as large as ours occur. Values lower than .05, of course, indicate even greater likelihood that the null hypothesis isn’t correct.

But ***understand clearly***: there is *nothing “magic”* about .05. You’re still guessing about your data, and 1 in 20 times (on average) you will be *wrong*, i.e., you will accept the idea of a real

difference when there isn't one—because we expect differences as large as ours to occur that often when the null hypothesis *is* correct. In effect, at $p = .05$, you're just saying that you are “pretty sure” there is a difference. Of course, you could “require” a higher level of certainty (i.e., a lower probability value, such as $.01$) before rejecting the null hypothesis.

There is a strong tendency, certainly among students whenever they learn about “.05,” but sometimes even among scientists, to misinterpret values *above* $.05$. The misinterpretation is that higher values mean there is *no* difference in the true means. It is true that if $p = .15$ we would usually “not accept” the idea of a difference in the true means, because we want to be more certain. But what does $.15$ actually mean? It means we have a sample difference that would be expected to occur, if the null hypothesis were really true, only 15% of the time, and that 85% of the time the difference wouldn't be this great. To put it another way, if you *did* reject the null hypothesis in favor of a difference, you'd be *right* 85% of the time and wrong only 20%. So what's your best bet?

It is also *very important* to understand the use of the word “*significant*” in a statistical context. It is unfortunate that this word was chosen, because it is easily misunderstood. In ordinary English, “significant” means something like “large enough to make an important difference”: a significant advantage, significantly higher costs, significantly better gas mileage. In a statistical context it simply means “*probably real*, given the selected p level.” Whether that difference is “significant” in the other sense is a different question. In particular, “probably real” isn't the same as “large and important.” Sampling of deer may indicate a male/female sex ratio of $.8$ in one area and $.65$ in another, and this difference may be *statistically* significant with $p = .03$. But this has no significance to population growth, because one male can mate with many females, and so having fewer males in one area doesn't make any difference. On the other hand, if it was a monogamous species (like Canada geese), fewer males might be “significant” in leading to lower population growth. Other common examples of this kind of thing occur in reporting about health issues. A recent study (July 2005) from Virginia Commonwealth University in Richmond found that using low-dose birth control pills “significantly” increased the risk of heart attack in younger women—in fact, actually *doubled* the rate. Sounds bad. But as is often the case, the report described only the *change* in the rate, without saying what the rate actually *is*. Heart attack in young women who don't have other serious risk factors for it (obesity, heavy smoking, etc) is extremely rare, maybe 1 in 20,000. Doubling that means 1 in 10,000. Is that of practical “significance” to a woman wishing to use this form of birth control? Probably not. And finally, keep in mind that the larger the sample sizes, the smaller the difference that can be shown to be “probably real.” For example, a study involving over 50,000 humans found some “significant” differences among various ethnic groups in cranial (skull) capacity (i.e., the volume available for the brain). Some individuals seize on this kind of information to promote racial stereotyping. But the actual mean values varied within only about 1% of each other, and are of no practical significance.

Units

You're all familiar with common English units of length such as inches, feet, yards, and miles. And all of you learned about the metric system somewhere along the line, and maybe used it in your high school science classes. Although lots of you are probably rusty on it, you may have some idea of length units such as centimeter, meter, and kilometer, and how they relate to English units. And small-scale area units are simple and familiar: a square foot, a square meter. However, it's likely that few of you can describe or explain *large-scale area* units, even in the English system. You've heard of an "acre" and you may have a vague visual idea of its size, maybe because you know your home is on a 1 acre, or $\frac{1}{2}$ acre (or whatever) lot. But you'd probably be hard-pressed to describe an acre mathematically. And the great majority of you probably haven't even heard of the large-area metric unit. So let's talk about this.

An acre is the amount of land that can be plowed by a yoke of oxen (= 2 oxen) in a day. It is equivalent to an area 1 furlong long (furlong = the distance the yoke of oxen can go without resting) by 4 rods wide (rod = the length of the long stick used to control the oxen). Well OK, that's the origin anyway.☺ More relevantly, an acre is $43,560 \text{ ft}^2$, or about 209 ft/side if square-shaped. It follows that, as any old-time farmer can tell you, you can step off an acre as 70 paces per side (for an average-sized male, 1 step = ~ 3 ft).

In the metric system the large-scale area unit is the *hectare*, abbreviated "ha." As always with the metric system, there is a simple relationship between it and smaller units: it is $100 \text{ m} \times 100 \text{ m} = 10,000 \text{ m}^2$. A hectare is quite a bit larger than an acre, by a factor of about 2.5.

With our tree sampling, we will have a mean number of trees per 100 m^2 quadrat. Therefore the number of trees per hectare can be estimated as $(10,000 \div 100)(\text{mean}) = 100(\text{mean})$.

The diagram below shows the proportional relationship between an acre and a hectare.

