An Introduction to Sampling: Vegetation Analysis of Prairies

To begin our study of prairies, we will undertake the critical first step in any scientific endeavor, careful description. The goal of descriptive studies is to seek patterns in nature, which are then best tested experimentally -- although *well-planned* descriptive studies can also be used to test hypotheses when experimental approaches are not feasible. As you will see, descriptive studies require a great deal of forethought and can be highly quantitative -- patterns are most precisely described mathematically rather than simply verbally. In your first investigation, you are going to describe the features of a population of your focal plant species, in one particular area of CERA. Like all good descriptive investigations, your results should raise a number of questions that could be addressed in further studies, including experimental ones.

One of the first concerns we should address if we want to measure any aspect of a community is how we deal with spatial variation. How can one know that any measurement of a physical or biological parameter is an accurate representation of a community when it varies within a relatively small area?

What is a sample?

A sample is an independent subset of the population that you are studying. When you do a biological study (experimental or descriptive), you are studying a population of organisms, be they plants, animals, or fungi. If we were to analyze the prairie vegetation at CERA, the population would be all of the plants growing in prairies at CERA.

Why sample?

We sample for one simple reason: in most biological studies, it is logistically impossible to enumerate the entire population of organisms. Imagine having to individually count all of the plants in even a small plot of prairie. Or trying to count the individuals in a population of mobile animals such as butterflies. A biologist would only be able to do a single study in his or her lifetime! Not surprisingly, sampling plays a very important role in biological studies.

The three types of sampling

There are three basic ways to sample a population: **haphazard**, **random**, and **systematic** sampling. **Haphazard** sampling is the simplest type of sampling: you just go into the area you are studying and arbitrarily decide where to take your samples. It is also the least desirable way to sample for one simple reason: your samples may be biased. For example, you may unconsciously decide to sample plants in areas where the vegetation is healthier, biasing your results. For logistical reasons, you sometimes have no choice but to haphazardly sample, but it is never desirable because of the potential for bias.

Random sampling is the best type of sampling because it is unbiased (unlike haphazard sampling). When you randomly choose where to take your samples from, you can be assured that your samples are an unbiased subset of the population. So if you remember only one thing about sampling, let it be this: **sample randomly whenever possible!**

Systematic sampling falls between haphazard and randomly sampling, both in ease of sampling and desirability. In systematic sampling, you take your samples at some regular interval (say every 2 m) across the whole population you are sampling. Researchers studying sessile organisms such as plants or intertidal marine invertebrates commonly use systematic sampling. For these organisms, running a transect or series of transects across the population and taking samples at fixed intervals is simply the easiest thing to do. While systematic sampling is exceedingly easy to do, it suffers from the same problem as haphazard sampling: possible bias. If there is periodic variation in the population you are studying, then systematic sampling can yield a biased set of samples. For example, if soil moisture increases and then decreases every 2 m, then systematically sampling every 2 m could result in a biased set of samples. Fortunately, there is little evidence of periodic **spatial** variation in natural systems (although **temporal** variation in natural systems is often periodic). Although systematic sampling is unlikely to be biased, random sampling is generally preferable.

In addition, there is a sort of hybrid between random and systematic sampling that is often used. Say that in your (non-quantitative) observation of a field, you notice that there is a slope in elevation and you're worried that organisms might be responding to that slope in some way. If you sampled purely randomly, you might not pick up certain areas of the slope, and that wouldn't characterize the overall population well. If you divided the field into three sections (say low, mid and high elevation) and then randomly sampled with in each section, you would be doing **stratified random sampling.** By general definition, this term means that you predetermined groups based on some parameter, then sampled within them randomly.

How many samples?

Within reason, the more samples you take, the better, as your populations estimates will be more precise. Having said that, you should gauge the number of samples you take by how variable whatever you are sampling is. For example, if you wanted to look at the effect of fire on populations of prairie plants and insects, you would take more insect than plant samples. Why? Because insect populations vary more in time and space than plant populations do. To get good estimates of insect populations, you would simply need more samples. Unfortunately, you often do not have good information on the natural variation in whatever you are measuring. In these cases, you will have to do the best you can.

Today's exercise

Over the period today and next Wednesday, you and your partner are going to develop a quantitative description of a population of your focal prairie plant. Your data should address both of the following questions of interest:

1. **How abundant is your species in the area (what is its *density*)?** Depending on the density of your species you can use one of the following ways of measuring density:

    a. For very abundant species, define a random sampling point, and count the number of stems within 1-meter radius of the point.
    b. For less abundant species, determine a random point on the edge of the population and lay down a 30-meter tape through the population. Using a meter stick count the

number of plants within ½ meter of the tape as you walk along it. This is called a **belt transect**.

Measure density as many sampling points or transects as you can today. Make sure you write down the density at each point or for each transect, since that will be important for understanding the population dispersion – i.e. whether individuals are distributed in space in a clumped, even, or random fashion.

2. **How do characteristics of individuals vary?** In plants there are several features that might be good measures of relative vigor including:
    a. Height (in meters) of the tallest part of the plant.
    b. The stem width, measured with calipers at the base of the stem near the ground.
    c. The number of flowers, inflorescences and/or fruits.
    d. The length or width of a leaf. (You might have to think about how you are going to sample leaves on the plant to get this measure in a non-biased way).
    e. The distance to the nearest neighbor of the same species – this might be hard for very rare species, but not so hard for abundant ones. (Obtaining this data will also allow you to measure dispersion in a slightly different way than with density data.)

Focus on this task on the second day of sampling by picking random sampling points and measuring each aspect on the individual of your species that is closest to the sampling point. Get as many points as you can in the time we have. The more points, the more enjoyable your data analysis will be!