Techniques and Practice of Ecological Sampling

One of the first things a field ecologist will want to know about an animal or plant species is: How **dense** is the population [units of density are number of individuals {or colonies etc.} per unit area {or volume}]. Another important question is: How are the organisms **dispersed** [The pattern of distribution in space] within the habitat? In most cases it is impossible to count every individual or plot their location on a map [This would be a **census**] because of the time, effort or money involved. So it would be useful if there were some way that we could get an accurate representation of some spatial characteristics of the population without having to map every organism.

By **sampling** the population we can do this, BUT the sampling must be done properly if we want our representation to be valid. To insure an adequate representation, some guidelines must be followed. This laboratory exercise is designed to introduce you to the methods, and rules of ecological sampling in the spatial domain, and to allow you to use some common tools to practice these techniques. You will use these methods in most of the rest of your labs this semester.

Choosing sample sites: Random vs. Haphazard:

Experimenter bias in sampling is a common hazard and must be continuously guarded against (for example: "It's late and this rock is too hard to look under so it wouldn't hurt to sample under this smaller rock" right?) To obtain an unbiased estimate of the population, sampling should be done at **random** –or more specifically *the sampling should be conducted in such a way that the probability of each individual being selected in the sample is the same*. There are several ways of insuring this criterion is met – or at least approximated. **Random numbers** are series of numbers such that the chance of selecting, for example, any digit (0 – 9) is equal at any point in the sampling procedure. If the random numbers can be assigned to organisms or to locations in the habitat, they can be used to select the sample from the population. One way to generate a series of random numbers is to write the numerals 0 through 9 on slips of paper, mix them in a hat, draw the slips out, write the number down, then replace the slip in the hat, remix, and draw again, etc. etc. etc. A faster and less cumbersome method is to use a **random number table**. You worked with ways of getting random numbers in the previous lab. You can use the numbers in the table to select sampling positions (e.g. paces along a trail, GIS coordinates, termite holes in a wall that you have numbered etc.). Most calculators and spreadsheet applications also have random number generating functions,

Often a person will think that they can make up random numbers and/or sample "randomly" without recourse to the use of some **randomizing method**. This rarely works [you will get a chance to test this statement in this lab] and such samples are termed **haphazard**.

Commonly Used methods for Sampling in the Spatial Domain:

There are three general types of sampling methods used to select individuals from a population situated in space: **quadrats**, **transect lines** and **plotless techniques**.

1) A quadrat is a frame (usually a square or a circle) of **known area** used to isolate a subset of the population. This subset will comprise one sample. Quadrats come in various sizes (and shapes) with the size selected determined by features of the organisms in the population to be sampled. This description might sound circular, but a postage stamp size quadrat might work well for mites on a leaf but will be hard to use on elephants. The use of a quadrat is very simple: It is placed randomly in the **sampling area** (the habitat of the species of interest) and all the individuals within the quadrat are counted and/or measured.

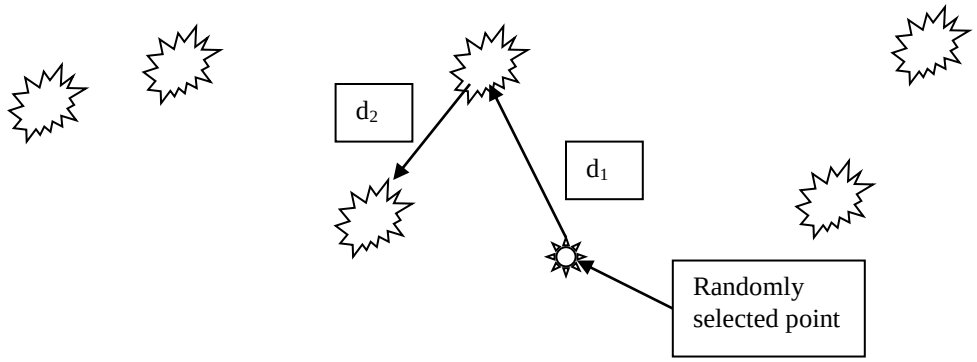
Exactly what the sampling area is, how its limits are determined, and how the quadrats are placed, are all very important points in using this method. Quadrats are most useful when the area is fairly uniform and movement within the area is easy.

2) The transect method is most useful when the area to be sampled is zoned in some way, or has some sort of gradients running through it. Think of the intertidal habitat on a small scale or the vegetation running up the side of a mountain on a larger scale. If you want to study the zonation itself, the transect lines will usually be run *normal* to the margins of the zones, while if the between zone differences are to be minimized, the transect lines are laid *within* zones parallel to the boundaries. You might say this introduces bias – and indeed it does. However, the researcher is aware of the bias, it is intentionally introduced into the sampling scheme to maximize the useful information, and, of course, the results can be interpreted only when this bias is taken into account.

Once the line has been randomly laid, points along the line are selected (or quadrats could be used too). The points might be randomly or **uniformly** placed, depending on the question being asked. In this way the distribution of a species across (or within) the environmental gradient can be determined.

A useful discussion of spatial sampling that goes into more detail than we can here can be found at : http://media.wiley.com/product_data/excerpt/03/04700444/0470044403.pdf

3) The plotless technique we will use is the nearest-neighbor method. It is particularly useful in areas where carrying large quadrats would be difficult to manage such as a dense forest. It is also most suited to habitats that are relatively *uniform*. Plotless methods do not use a defined area (like a quadrat), nor are they arranged linearly as in transect methods, rather, they use **distances** between a point and an individual, or between two individuals.

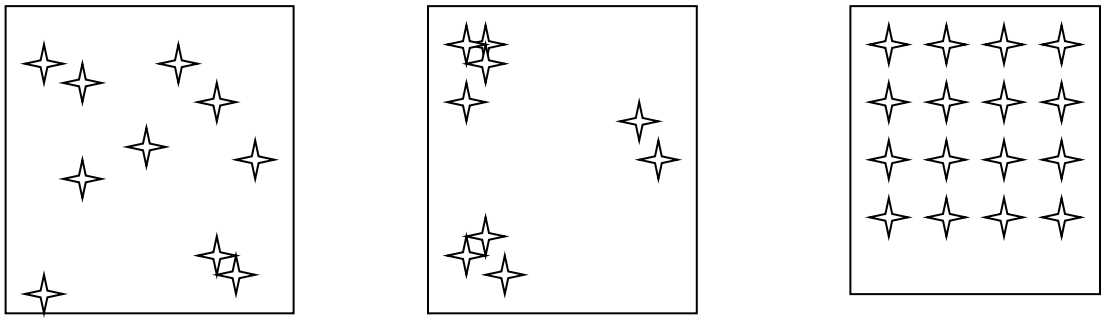


d_1 = the distance from the randomly selected point to the closest individual of interest.
d_2 = the distance to the nearest neighbor

In this technique each sample consists of a *pair* of measurements, d_1 and d_2. To obtain the next sample, a new random point is chosen and another pair of measurements is made. While nearest neighbor methods are especially useful in certain field situations like dense forests etc. you will see that unlike quadrat methods, they are not that useful for determining density. They are however quite useful for determining the dispersion pattern of the organisms.

Density and dispersion patterns:

As noted earlier, the number of individuals per unit area is termed the density and the units are things like; trees $acre^{-1}$, moose km^{-2}, zooplankton m^{-3}, parasites $fish^{-1}$, etc. Dispersion is the **pattern** of the distribution of organisms in space. There are three basic dispersion patterns (and their combinations - depending a lot on the **scale**): **random**, **regular** and **clumped** (or **contagious**).



Random Contagious Regular.

A random dispersion pattern means that there is an equal probability of an individual occurring at any point in the habitat and that the presence of an individual does not influence the probability of occurrence of another individual. Contagious dispersion patterns are those where the presence of an individual increases the probability of finding another one near by. Regular dispersion, indicated by more even spacing that would be predicted by a random dispersion may suggest territoriality or some limiting resource. What pattern do you think is most common in nature? Why?

Sample Size:

Hopefully, by now the question of **sample size** has already occurred to you. How many samples (of any kind) will you need to take before you are confident (how confident?) that your estimate of density or dispersal reflects the true situation? Clearly, the larger the sample the better, but things like time, manpower and money also enter the picture. How can you determine the appropriate sample size? There are many methods some simple and some complex. One easy method is graphical and should be done while you are in the process of sampling. The method consists of plotting a **running mean**. The X-axis is the number of samples (1, 2, 3, etc.) and the Y-axis is the **mean** number of individuals per sample (a **cumulative** value averaged over the

continuously increasing number of samples you have taken.) As the number of samples increase (as you move to the right along the X-axis) the running mean should begin to stabilize. If you are plotting quadrat data, once the optimum number of samples has been determined (and this is NOT a constant it can change with species, terrain, time of day and many other factors), it is straightforward to determine the density since density is simply the mean number of individuals per sample area. The shape of the running mean curve will be strongly influenced by quadrat size. This is one of the things you will investigate in this exercise.

Determining the dispersion pattern:

With quadrat methods the detection of dispersion patterns is based on the following series of possible values for the **variance: mean ratio:**

$s^2/\bar{X} = 1$ implies random $s^2/\bar{X} < 1$ implies regular $s^2/\bar{X} > 1$ implies contagious.

Think about what the two variables mean and why these inferences can be drawn from this ratio.

To test whether the difference is large enough to be considered significant you can use the fact that $s^2/\bar{X}$ is distributed like χ^2 with n-1 degrees of freedom (Southwood 1966). Remember this is a **two-tailed** test. Why? A more general test is discussed in Greig-Smith (1964) where a χ^2 test based on expected values derived from the **Poisson distribution** is used. (You will learn more about this statistical pattern in lab 7). You should be familiar with the second method, but the first one is all right for "quick and dirty" tests. Distributions of this statistic can be found in statistics books or on line ex.
http://brd4.ort.org.il/~sdror/quality/Poisson.htm

For quadrat data, the mean and variance are calculated taking the total number of individuals counted in each sample as the individual samples. The theory behind the use of the variance:mean ratio can be understood as follows: In a perfectly regular dispersion, the variation about the mean will be very small since all individuals are equally spaced so you would expect most samples will have about the same number of individuals. Thus the low variance will result in a variance mean ratio less than one. From statistical theory we know that in a purely random distribution the variance is equal to the mean. But the more relevant question to you as an ecologist might be how different from one does the ratio have to be before you are convinced that the dispersion is NOT random? You can use the χ^2 distribution as a significance test for the variance:mean ratio:

Consider the expected value (E) to be the mean ($\bar{X}$) of all the observations, and the observed values to be the individual observations. Thus:

$$\chi^2 = \Sigma (O - E)^2/E \quad = \Sigma (X_i - \bar{X})^2/\bar{X}$$

with n-1 degrees of freedom, where n is the number of samples. You can then simply use a χ^2 table to determine the level of significance.

Using nearest neighbor techniques. With this class of methods there are no quadrats to count individuals rather it is the spacing itself that is the variable measured. The dispersion pattern and its corresponding statistical significance test is outlined here:

$$A= \quad \Sigma(d_1^2) / \Sigma (d_2^2)$$

So that
A < 1 implies regular dispersion
A = 1 implies random dispersion
A > 1 implies contagious dispersion

Why should this be? Try to think it through.

For random dispersions the parameter x = A/ (1+A) has a value of 0.5 Think why this should be the case.

You can use the attached graph to determine whether or not the dispersion pattern deviated significantly from random at the probability level desired

A more general method to estimate dispersion patterns from nearest neighbor data is from Clark and Evans (1954). Using their method, in a population of individuals with a density D, the distance from each individual to its nearest neighbor is measured. The mean distance is:

$$\bar{r}_a = \Sigma d/n$$

If the population is distributed randomly, the expected value of the average distance would be r_e where :

$$\bar{r}_e = 1/(2D)^{1/2}$$

The ratio R (= $\bar{r}_a$ / $\bar{r}_e$) is a measure of the departure from randomness. If R=1 the pattern is random, if aggregation is maximum, R=0. Can you figure out why this is so? If spacing is maximum R = 2.1491. A significance test for this value is given by $c = (r_a-r_e)/s_D$, where s_D is the standard error of the mean distance to the nearest neighbor in a randomly dispersed population with density D. In this case

$$s_D = 0.26126/ (n*D)^{1/2}$$

The c values of 1.96 and 2.58 represent the 5% and 1% levels of significance respectively (for a two tailed test). A refinement of this method is given by Simberloff (1979)

References:

Clark, P. J. and F. C Evans,. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecol 35: 445-453.

Greig-Smith, P. 1964. *Quantitative Plant Ecology*. Plenum Press NY

Greig Smith 1979. Pattern in Vegetation J. Ecol. 67:755-779.

Kinzie III, R. and R. H. Snider 1978. A simulation study of coral reef survey methods. IN *Coral Reef Research Methods*. D. R. Stoddart & R. E Johannes eds. UNESCO Paris

Pielou, E. C. 1969 *An Introduction to Mathematical Ecology*. Wiley Inter Science NY

Poole, R. W. 1974 *An Introduction to Quantitative Ecology* McGraw-Hill NY

Simberloff, D. 1979. Nearest neighbor assessments of spatial configuration of circles rather than points Ecology 60:679-685

Southwood, T. E. E. 1966 *Ecological Methods*. Methuen London