Lab 3. SAMPLING AND COMMUNITY ANALYSIS

A. Introduction

Ecological systems such as populations or communities are usually quite large and complex. Yet in order to compare them to one another or assess changes in them, ecological systems must be described and quantified. In order to describe populations and communities, several ecological variables can be measured, such as density, frequency, coverage and biomass. These measurements are then used to characterize aspects of populations and communities, such as population distributions, species diversity, and productivity.

Ideally, one would count all individuals in a population or community to determine density, frequency, or other characteristics of particular species. This sampling method (i.e., census) is often impractical or impossible due to time or economic constraints. For instance, it would not be practical to count every *Lemna* plant growing in Tuttle Creek Reservoir. Instead, by sampling several 1-m$^2$ areas and counting the number of *Lemna* growing in these samples, the total number of *Lemna* in the whole reservoir can be estimated by multiplying the number of *Lemna* per sample by the total area of the reservoir.

It is important to choose a sampling technique that will provide the most accurate representation of the population or community studied. Choosing an appropriate sampling technique is a critical task, and should be conducted with prior knowledge of the population or modeled after other sampling experiments involving the same organisms or habitat types. Once the sampling technique is chosen, it is important to follow prescribed procedures carefully, taking shortcuts will invalidate statistical comparisons and any inference based on the data. If the sampling technique chosen leads to overestimation or underestimation of characteristics of a population or community, then it is considered to be biased (or inaccurate). Biased estimates cannot be used to make inferences from the sample about the entire population or community, because they are not representative of the whole population.

For this lab, we will sample a desert plant community in southern California. Since the class does not have time (or funds) to go to California, maps of the sites will be used. The community to be sampled covers a 4800 m$^2$ area and includes two zones of vegetation: bajada (upland community) and wash (lowland community). Circles indicate the exact location and size of 17 species of desert shrubs and cacti. The objectives of this lab are to 1) determine what type of sampling technique and design to use to measure density and frequency of each plant species in the two vegetation zones, 2) compare the community structure and species composition of the two zones, and 3) evaluate the effectiveness of the sampling techniques utilized and determine sampling adequacy.

B. Sampling Techniques and Estimation

To obtain a representative sample of a population, the sample must be unbiased and it must be adequate in size. An unbiased sample is obtained by randomly sampling from a population. A random sample is one in which every member of the population of interest has an equal and independent probability of inclusion. Contrasted with random sampling are systematic sampling
Sampling and Community Analysis

and any other type of sampling technique. Systematic sampling involves taking samples in some sort of systematic (or uniform) manner. For example, sample plots may be located every five meters along a 50 meter transect. Although systematic sampling is simpler than random sampling, there may be bias in sampling. Therefore, if this technique is employed, it is important that care is taken to avoid bias. Unless it can be demonstrated that no bias is present, random rather than systematic sampling (or other techniques) should be used.

With random sampling, the goal is to choose sample units from a population with an equal and independent likelihood. Although haphazard placement of samples may appear to avoid conscious bias, this is not an adequate technique. When samples are located in this way, biases usually emerge because samples often turn out to be too evenly spaced or are not placed in areas with dense vegetation or other characteristics, such as presence of rocks or large trees. The basic approach of random sampling is to use chance to determine which units will be selected. For example, we may want to locate several plots in a field in order to characterize the plant community. To ensure randomness of the sample, we could establish a system of coordinates for the field, letters for north-south and numbers for east-west. We could then put the letters and numbers in separate hats and draw them out in pairs. If we selected a D and a 5 then the plot would be located in the field where the line D crossed the line 5. This technique, as well as rolling dice and flipping coins are all appropriate for obtaining random samples. A more simplified technique, however, involves using random number tables or random numbers generated by a computer.

A random number table consists of a long series of digits that are random. These tables can be found in any elementary statistics text. To use it, pick a starting point on the table and omit the first four to five digits. Then select numbers in a predetermined order, such as left to right. If for example you have a total number of 40 plots that you need to select, you would only select two-digit numbers from 01 to 40 and omit any others that fall outside that range.

A number of sampling techniques and designs are available for sampling plant populations and communities. A few of these are listed below:

1. **Sampling Techniques**
   a) Census - all individuals counted
   b) Sample - plot or **quadrat** (square, round, rectangular), arranged randomly or in line or belt transects
   c) Sub-sample - use a portion of a sample
   d) Transects - line-intercept
   e) Point-quarter or **quadrant** - points located randomly or along a transect

2. **Sampling Designs**
   a) Simple random sampling without replacement or with replacement - most commonly used
   b) Stratified sampling - population is partitioned into regions or strata (e.g., different soil types)
   c) Systematic or cluster sampling - consists of both primary units and secondary units (e.g., stream reach with quadrats)
d) Multistage sampling (two-stage, three-stage, ...k-stage) - to obtain a sample of plants for a particular species, first take a sample of plots and then a sample of plants from each selected plot.

Either quadrats, quadrats arranged along transects, line-intercept, or point-quarter could be used to sample the desert plant community, and we could choose several of the designs listed above to obtain an unbiased estimation of frequency and density of each species. It is important to keep the sampling design as simple as possible; otherwise, estimation of density and frequency may be difficult. Therefore, we will use the simplest sampling design (simple random sampling without replacement) and use two different sampling techniques, quadrats and line-intercept transects.

B. Procedure

1. Sampling techniques
   a) Quadrat sampling
   Divide the study area into bajada and wash communities. We will use 10 quadrats to sample each community. Use a random number table (or a calculator's random number generator) to generate 10 two-digit numbers, which will be the x and y coordinates for the bajada community and will correspond to the scale on the edge of the map. Place the lower left-hand corner of the quadrat at these coordinates and count and determine the coverage of each individual that has greater than 50% of its area within the quadrat. Repeat the process for the wash community.

   b) Line-intercept sampling
   Again, divide the study area into bajada and wash communities. We will use 3 transects to sample each community. Use a random number table (or generator) to generate 3 two-digit numbers, which will correspond to the x coordinates on the map scale for the bajada community (Fig. 3.1). Draw the transects parallel to the y-axis and mark off 5-m intervals on the line. Each interval will be treated as a separate unit of the transect. Begin counting at one end of the transect, and count all plants that are intercepted by the transect and determine the intercept length for each plant (i.e., the portion of a line intercepted by a plant) (Fig. 4.1). This data will be used to estimate species frequency and coverage.

![Figure 3.1](image-url)

Figure 3.1. A) Example of random placement of transects in the bajada and wash communities. B) The line-intercept method used to estimate coverage of each plant species.
2. Calculations

We will use the following calculations to estimate abundance, density, coverage (i.e., dominance), frequency, and importance values for each species encountered in the bajada and wash communities.

**a) Quadrats:**

Abundance \((A_i)\) = number of individuals for species \(i\)

Coverage \((C_i)\) = cover for species \(i\)

\[
\text{Density} \ (D_i) = \frac{A_i}{\text{Area}}
\]

\[
\text{Dominance} \ (d_i) = \frac{C_i}{\text{Area}}
\]

\[
\text{Frequency} \ (f_i) = \frac{\# \text{ of quadrats with species } i}{\text{total } \# \text{ of quadrats sampled}}
\]

Importance value \((IV_i)\) = \(RD_i + Rd_i + Rf_i\)

**b) Line-intercept transects**

Abundance \((A_i)\) = number of individuals for species \(i\)

Length \((L_i)\) = length of transect \(i\)

\[
\text{Linear density index} \ (ID_i) = \frac{A_i}{\sum L_i}
\]

\[
\text{Linear coverage index} \ (IC_i) = \frac{l_i}{\sum L_i}
\]

\[
\text{Relative coverage} \ (RC_i) = \frac{IC_i}{\sum IC_i}
\]

\[
\text{Frequency} \ (f_i) = \frac{\# \text{ of line - intercept intervals with species } i}{\text{total } \# \text{ of line - intercept intervals sampled}}
\]

Importance value \((IV_i)\) = \(RD_i + RC_i + Rf_i\)

Shannon diversity index = \(H' = -\sum p_i [\ln(p_i)]\), where \(p_i = \text{decimal fraction of individuals belonging to the } i^{th} \text{ species}\)
C. Measures of Community Structure

Species diversity is an expression of community structure. This measure incorporates both the number of species (i.e., richness) and the distribution of individuals among the species (i.e., evenness). A community has high species diversity if there are many species (high richness) and abundance of these species is nearly equal (high evenness). Conversely, a community with low diversity has either very few species (low richness) or only a few species that are very abundant (low evenness). High species diversity indicates a complex community with a variety of species and a complexity of species interactions. Recently, there has been much debate (e.g., Grime 1997, Huston 1997) concerning the relationship between species diversity and community stability or resilience, which is the ability of a community to respond to disturbance. Some researchers have shown that with increasing diversity, stability of a community increases (Tilman and Downing 1994, Tilman 1996); however, other researchers have concluded that there is no simple relationship between diversity and stability (e.g., Huston 1997). Because diversity may be linked to stability and functioning of ecosystems, the concept of diversity (or biodiversity) has important implication for conservation of ecosystems.

There are a wide variety of formulas used to calculate diversity and evenness depending on a number of different factors and the system being sampled (see Magurran 1988, Brower et al. 1998, Krebs 1999). For this class, we will calculate richness (i.e., number of species in a community), diversity (Shannon diversity index) and evenness (i.e., distribution of individuals among species) for the bajada and wash zones. These measures will allow us to compare the community structure of the two zones.

D. Coefficient of Similarity

The Sorenson’s coefficient of similarity (C_c; Cox 2002, p. 202) will be calculated to determine how similar (or dissimilar) the bajada and wash communities (i.e., species composition) are to each other. A similarity coefficient that equals 100 indicates the two species assemblages are identical, whereas a value of 0 means that the communities are completely dissimilar. Although there a number of different ways to calculate community similarity (see Magurran 1988 and Bower et al. 1998 for details), for this class we will calculate C_c using the following equation:

\[ C_c = \left[ \frac{2W}{(A + B)} \right] \times 100 \]

where
- \( W \) = sum of lower of the two \( IV_i \) for shared species
- \( A \) = sum of \( IV_i \) for community a
- \( B \) = sum of \( IV_i \) for community b
- \( IV_i \) = importance value for species \( i \) (see calculations below)

Table 3.1 shows an example of two communities with a similarity index of 56.2 (\( C_c = \frac{2(11.5)}{(26.0 + 14.9)} \times 100 \)). Therefore, overlap between the two communities is approximately 56%.
Table 3.1. Example of calculation of coefficient of similarity ($C_c$) for communities A and B.

<table>
<thead>
<tr>
<th>Species $i$</th>
<th>Importance Values ($IV_i$)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>min $IV_i$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.8</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>4.2</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>2.2</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>6.0</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>A = 26.0</td>
<td>B = 14.9</td>
<td>W = 11.5</td>
<td></td>
</tr>
</tbody>
</table>

F. Evaluation of Sampling Techniques and Sampling Adequacy

As indicated earlier, replication is necessary to prevent bias in estimation of population or community characteristics. A single measurement is not enough to draw conclusions about a certain characteristic. Through replication we can estimate the mean of a population and determine how much error exists in this estimate.

How many samples are enough? In general, the larger the sample the greater the adequacy; however, taking a sample larger than necessary is a waste of time and money. Thus, there are no set answers to this question, and the number will vary from study to study. If the study area in question is homogeneous or species are distributed randomly, then most likely only a small number of samples will be needed. In contrast, if an area is heterogeneous or species are distributed in clumps or non-randomly, then a larger number of samples will be needed.

One way of estimating if enough measurements have been taken is illustrated by the species-area curve (Table 3.2, Fig. 3.2). This is a plot of the accumulated number of species against the size of the sample (e.g., number of quadrats or cumulative area sampled). This curve illustrates whether the sample size is large enough to include most of the species in the community we are studying. With the first sample, all species encountered are new. With the second, we will probably encounter some of the same species in the first quadrat but will also add new species. As we proceed in our sampling, we will continue to add rarer species so that each new sample adds fewer species to the accumulated total. The number of samples is considered sufficient when the curve reaches an asymptote where no new species are added with increased sampling effort.

A performance curve can also be used to determine sampling adequacy. This curve plots the mean value of a measure (e.g., density, biomass, weight, etc.) against the sample size. The first two samples will likely not reflect the population mean, but as more samples are added, these values will approach this mean. Therefore, the curve will appear jagged at the beginning (because of greater variability between samples) but will level off as more samples are added. The point in which the curve flattens out indicates that the sample is adequately estimating the population mean of a particular measure.
Table 3.2. Data for generating the species-area curve of Fig. 3.2. Each sample represents a 10 m² quadrat (Modified from Brower et al. 1998).

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Cumulative area sampled (m²)</th>
<th># of species</th>
<th># of new species</th>
<th>Cumulative # of new species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>5</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>6</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>4</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>5</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>5</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>70</td>
<td>5</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>4</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>4</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>5</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>

Related References


Sampling and Community Analysis Questions (20 pts. total)

Due February 25, 2002

1) (6 pts) Graph the relative density, relative dominance, and relative frequency for each species using the plot (quadrat) and transect data.

2) (4 pts) Compare the relative density, relative dominance, and relative frequency of the three most abundant species in each plant community calculated using the plot data to those calculated using the transect data. How different (or similar) were the estimates of these variables using the two sampling techniques? Did one technique tend to overestimate or underestimate any of these variables?

3) (6 pts) Calculate richness and diversity for the bajada and wash communities using the plot data. Summarize in a table. Compare and contrast the two plant communities using these measures of community structure.

4) (4 pts) Calculate the coefficient of similarity between the bajada and wash plant communities using the plot data. Summarize in a table like Table 3.1 in Lab 3. How would you interpret the coefficient of similarity for the two plant communities?

Include copies of all your worksheets and calculations.

Extra Credit (3 pts)

(2 pts) a) Plot a species area curve for each community type for the plot and transect data. Did you have an adequate number of samples using each sampling technique (i.e., did the curves level off)?

(1 pt) b) Which sampling technique would you use if your goal was to sample most of the species in the two plant communities using the smallest sample size?