1a

Ecological Sampling

1. Introduction

Ecologists generally wish to collect quantitative information about a habitat, community, or population, for quantitative data may allow objective and illuminating presentation, summary, and interpretation of ecological phenomena. However, it usually is impossible or impractical to monitor the entire habitat or to obtain measurements of all the organisms in a given area. Biologists rarely can collect all of the data about which they wish to draw conclusions. For example, it may be desired to draw conclusions about the body weights of all mice in a particular habitat. The only way to make statements about the weights of all mice with 100% confidence would be to weigh every mouse, probably an impossible task. Instead, only some of the total number of mice are weighed, and we can then infer from this portion of the total the weights of all the mice. The entire set of data of interest (i.e., the weights of all of the mice) is called a **statistical population** and the measured portion, or subset, of the population is a **statistical sample.**

Established sampling procedures exist for obtaining information about organisms and their environment. In this section we shall deal with the general principles of sampling underlying the specific techniques of sampling habitats and biological populations given in Units 2 and 3. The theoretical bases for ecological sampling procedures may be found in such texts as Greig-Smith (1983), Krebs (1989), Pielou (1977), Poole (1974), Seber (1982), and Southwood (1978).

A statistical population is that entire set of data about which one wishes to draw conclusions. This is not to be confused with a **biological population,** which is the aggregation of individual organisms of a single species inhabiting a given area. A statistical population, then, is an entire set of measurements from a habitat, a community, a biological population, or a portion of a biological population. Though a statistical sample is a portion of a larger set of data (the statistical population), a **physical sample** is a portion, or subset, of a collection of one or more material objects, either biotic or abiotic. As an example of physical sampling, we can take a 1-liter sample of pond water (meaning we collected a 1-liter portion of the entire volume of water in the pond), or a sample of vegetation from a forest (i.e., a small portion of all the forest vegetation), or a sample of 100 mice from an entire biological population of that species. A statistical sample, on the other hand, refers to a collection of data such as measurements of the temperature or phosphate content of pond water, the biomass of vegetation, or the tail lengths of mice.

When collecting samples in an ecological study, one *must* know what natural entity is being sampled. A particular study may require a precise definition of the strata, zones, microhabitats, and/or times being sampled. Also, one may wish to study only a certain taxon or a particular collection of taxa. For example, if we obtain a collection of pond animals with a fine-mesh plankton net, we have not sampled all the pond fauna. Rather, we must be aware of the particular kinds of animals the particular sampling procedure can collect. Sweeping an insect net through the herbaceous vegetation of a forest would not yield a sample of all animals in that forest, but only a sample of those forms inhabiting a particular portion of the ecological community (i.e., the herb stratum, rather than the soil, shrub, or tree stratum), and only those not escaping capture by the net. Also, a sample of an ecological population seldom contains all the stages of the life cycle, which is important to realize when making inferences about a population or community. No single sampling device or technique can provide data on an entire habitat, community, or biological population. Thus, we must always define the ecological entity actually sampled by a given procedure.

2. Selecting Samples

After defining the ecological entity to be sampled and choosing the sampling technique (detailed in Unit 3), one can then do the actual sampling. However, assurance of a truly representative sample of the defined population, community, or habitat is usually a difficult problem in ecology. Normally, samples should be taken at random. Random sampling implies that each measurement in the population has an equal opportunity of being selected as part of the sample, and that the occurrence of one measurement in a sample in no way influences the inclusion of another. Sampling procedures are **biased** if some members of the population are more likely to be recorded than others, or if the recording of some affects the recording of others. If the sample is taken at random from a statistical population, legitimate conclusions may be drawn (with known chance of error) about that population, even though only a small portion of it has been measured.

A table of random numbers (Table 1A.1) often helps obtain random samples. In Table 1A.1, each integer from 0 to 9 has an equal and independent chance of occurring at any location in the table, each two-digit number from 00 to 99 has a random chance of occurring anywhere in the table, and so on. Each time this table is used, it should be entered at random; that is, do not always begin at the same point in the table. Once entered, numbers in the table may be read in any predecided direction—horizontally, vertically, or diagonally. If members of a population of objects (e.g., mice or trees) could be numbered, then a random sample of n objects from that population could be designated by considering n different numbers from the random number table. This is equivalent to placing each member of the population in a hat and drawing n of them by chance. However, this method generally is impractical, for numbering the individuals in the population would mean obtaining all of its members, and if this could be done there would be little need for sampling.

Random numbers may be used to select random map coordinates or numbered sampling sites. Sampling sites can be numbered easily by arbitrarily selecting a point within the habitat and marking off four compass directions (N, E, S, W) from this point to define four quadrants. A randomly selected number could represent the number of meters, or tens of meters, along one axis of a quadrant, and a second random number could do the same along the other axis for that quadrant. Thus, each pair of random numbers would establish a specific point in the quadrant at which to collect a physical sample. A quadrant could be selected at random by picking a random number from 1 to 4 and this process repeated until a sufficient number of random points had been selected.

3. Sampling Replication

A single measurement generally is insufficient to draw conclusions about an ecological characteristic. This is because a single datum is not adequate to judge how reliably that characteristic had been estimated. Repeated measurements may vary greatly; hence a single value could have an uncomfortably high probability of being far from the typical or average value. Therefore, a series of repeated, or **replicated,** measurements should be taken. From this collection of replicates (i.e., the statistical sample), we can estimate the mean of the statistical population and determine how much error exists in making this estimate (see Sections 1B.2.1 and 1B.2.4).

How many replicate data are needed to obtain a good estimate of some aspect of a statistical population (i.e., a reliable estimate of a characteristic of an ecological population, community, or habitat)? There is no set answer, but a number of procedures can aid in determining whether enough measurements have been collected. Two common methods—the species-sample curve and the performance curve—are discussed here. A procedure using statistical considerations in discussed in Section 1B.2.5.

In a **species-sample curve,** the cumulative number of species is plotted against the cumulative number of physical samples, where each sample might be a plot, transect interval, point-quarter point, insect-net effort, seine haul, etc. (see Unit 3). If the cumulative number of species is plotted against the cumulative size of the area sampled, this is called a **species-area curve.**

Figure 1A.1 is a presentation of the data in Table 1A.2. Here each datum is a species enumeration for a 20-m^2 area. One finds three species in the first sample. Because the second sample has four species, but two are species found in sample 1 and two are species newly found in sample 2, there are 3 + 2, or 5, species found in a total of 40 m^2 of sampling. The number of samples is considered sufficient after the curve levels off (see Figure 1A.1). However, if the curve reaches a plateau after only a very few samples, then the area in each sample is too large. The species-sample curve is an aid in evaluating both the number of replicates and the size of the physical sample. Physical samples that are too small may require a very large number of replicates. On the other hand, if the physical samples are too large, then fewer samples may be taken than necessary to allow for a satisfactory estimate of statistical error. The species-area curve is also useful for comparing the diversity of different communities and may be used in conjunction with Sections 5A and 5B.

A **performance curve** examines the mean value of a set of measurements for an ecological variable. For example, the mean density or biomass for a given species (or for all species) may be plotted as a function of the cumulative number of samples or the cumulative area sampled (Figure 1A.2). It is analogous to a species-area curve, but it plots a cumulative mean of some variable instead of the cumulative number of species. For a small number of ecological samples, such a mean fluctuates widely from sample to sample, but as the number of replicates increases, the fluctuation of the mean decreases (see Figure 1A.2). The number of replicates may be considered sufficiently large when such fluctuations are so slight that the cumulative mean has become insensitive to variations in the data. For example, the data of Table 1A.3 represent ten measurements of biomass as determined from ten physical samples, and it appears, from Figure 1A.2, that seven or eight samples would be sufficient to estimate mean biomass.

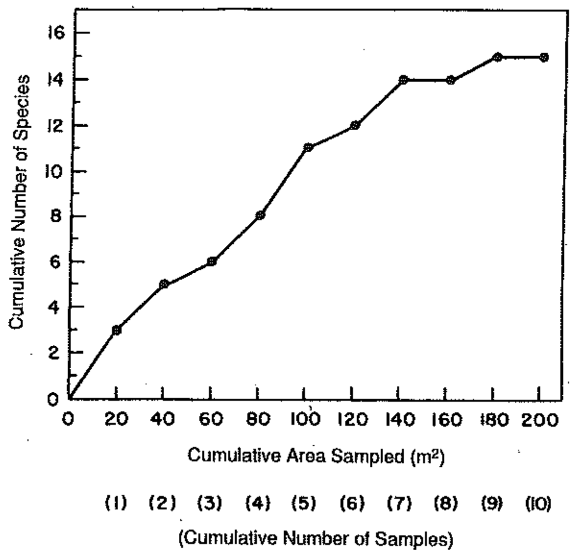

Figure 1A.1 A species-area curve for the data in Table 1A.2, plotting cumulative number of species against area sampled. If the cumulative number of species is plotted against the cumulative number of ecological samples (indicated in parentheses), this would be a species-sample curve.

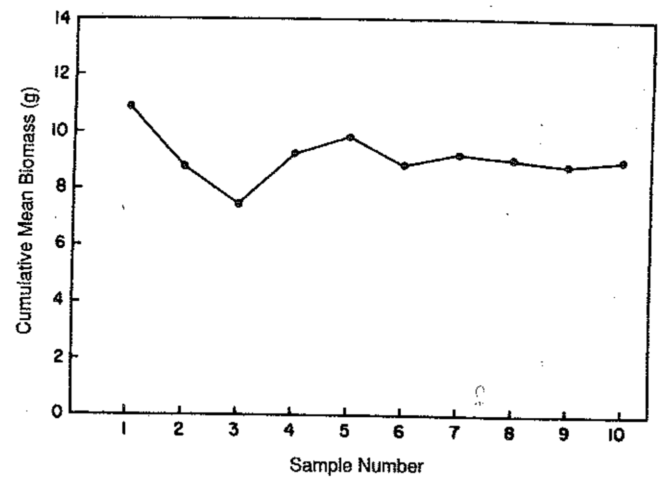

Figure 1A.2 A performance curve for the data in Table 1A.2, plotting cumulative mean biomass against cumulative number of samples.

Table 1A.2 *Data for Generating the Species-Area Curve of Figure 1A.1. Each ecological sample is from a 20-m² area.*

Sample Number	Cumulative Area Sampled (m²)	Number of Species	Number of New Species	Cumulative Number of New Species
1	20	3	3	3
2	40	4	2	5
3	60	5	1	6
4	80	3	2	8
5	100	4	3	11
6	120	4	1	12
7	140	4	2	14
8	160	3	0	14
9	180	5	1	15
10	200	4	0	15

Table 1A.3 *Biomass Data for Generating the Performance Curve Plotted in Figure 1A.2.*

Sample Number	Biomass (g)	Cumulative Mean Biomass (per sample) (g)
1	10.9	10.9
2	6.7	8.8
3	4.9	7.5
4	14.7	9.3
5	12.3	9.9
6	3.9	8.9
7	11.7	9.3
8	7.7	9.1
9	7.3	8.9
10	10.9	9.1

4. Subsampling

Occasionally, ecological samples are taken in the field and only portions of them, or **subsamples,** are later examined in the laboratory. The principle of subsampling is like that of sampling: The subsample must be randomly taken from the sample. This may require (as in a chemical analysis) shaking, mixing, or blending the sample before taking the subsample. In this way, subsample characteristics reflect the characteristics of the entire sample.

5. Experimental Design

Closely associated with the concept of sampling is that of **experimental design,** the planning of field or laboratory studies. Experimental design details with the questions to be asked in a study, the selection of variables to be studied, and the choice of a sampling program. The design is constructed, prior to the data collection, with specific procedures of sampling and data analysis in mind (see Section 1B and Units 2 and 3). There are many complex designs by which data may be collected and analyzed. A few of the simplest and most common will be discussed here and in Section 1B.

The most commonly used experimental design in ecological work is the two-sample comparison. Here, one selects two situations in which all conditions but one are equal (or nearly equal). For example, one may measure the population density of caddisfly larvae in a stream to conclude whether there is a difference between the densities in two different velocities of current. One then selects two sites with similar habitat characteristics (dissolved oxygen, stream substrate, depth, etc.) but with different current velocities. On examining the collected data, you may conclude that the population density of caddisfly larvae is different at the two current conditions. However, you cannot automatically conclude a direct cause-and-effect relationship and assert that the difference in population size was due to the current *per se* (e.g., a faster current may result in more food availability or better protection from predators.)

6. Selected References

Andrewartha, H. G. 1971. Introduction to the study of animal populations. 2nd ed. University of Chicago Press, Chicago.

Barbour, M. G., J. H. Burk, and W. D. Pitts. 1980. Terrestrial plant ecology. Benjamin/Cummings Publishing Co., Menlo Park, Calif.

Bormann, F. H. 1953. The statistical efficiency of sample plot size and shape in forest ecology. Ecology 34:474–487.

Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species-area relationship. Amer. Natur. 113:791–833.

Dixon, W. J., and F. J. Massey, Jr. 1983. Introduction to statistical analysis. 3rd ed. McGraw-Hill Book Co., New York.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John C. Wiley & Sons, New York.

Greig-Smith, P. 1983. Quantitative plant ecology. 3rd ed. University of California Press, Berkeley.

International Business Machines Corporation. 1968. System/360 scientific subroutine package (360A-CM-03X). Version III. Programmer's manual. White Plains, N.Y.

Kershaw, K. A., and J. H. H. Looney. 1985. Quantitative and dynamic plant ecology. 3rd ed. Edward Arnold, London.

Krebs, C. J. 1989. Ecological methodology. HarperCollins*Publishers,* New York.

Mead, R., R. N. Curnow, and A. M. Hasted. 1993. Statistical methods in agricultural and experimental biology. 2nd ed. Chapman and Hall, London.

Pielou, E. C. 1977. Mathematical ecology. John C. Wiley & Sons, New York.

Poole, R. W. 1974. An introduction to quantitative ecology. McGraw-Hill Book Co., New York.

Prince, S. D. 1986. Data analysis, pp. 345–375. In P. D. Moore and S. B. Chapman (eds.), Methods in plant ecology. 2nd ed. Blackwell Scientific Publications, Oxford.

Rohlf, F. J., and R. R. Sokal. 1995. Statistical tables. 3rd ed. W. H. Freeman and Co., San Francisco.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. 2nd ed. Macmillan Publishing Co., New York.

Snedecor, G. W., and W. G. Cochran. 1989. Statistical methods. 8th ed. Iowa State University Press, Ames.

Sokal, R. R., and F. J. Rohlf. 1985. Biometry: The principles and practice of statistics in biological research. 3rd ed. W. H. Freeman and Co., San Francisco.

Southwood, T. R. E. 1988. Ecological methods. Chapman and Hall, New York.

Steel, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics: A biometrical approach. 2nd ed. McGraw-Hill Book Co., New York.

Zar, J. H. 1996. Biostatistical analysis. 3rd ed. Prentice-Hall, Upper Saddle River, N.J.