

Quadrat Sampling in Population Ecology

Background

Estimating the abundance of organisms.

Ecology is often referred to as the "study of distribution and abundance". This being true, we would often like to know how many of a certain organism are in a certain place, or at a certain time. Information on the abundance of an organism, or group of organisms is fundamental to most questions in ecology.

However, we can rarely do a complete census of the organisms in the area of interest because of limitations to time or research funds. Therefore, we usually have to **estimate** the abundance of organisms by **sampling** them, or counting a subset of the population of interest. For example, suppose you wanted to know how many slugs there were in the forests on Mt. Moosilauke. It would take a lifetime to count them all, but you could estimate their abundance by counting all the slugs in carefully chosen smaller areas on the mountain.

Accuracy vs. Precision.

Obviously, we would like our method for sampling the population to produce a good estimate. A "good" estimate should maximize both precision and accuracy. In everyday English we often use these terms interchangeably, but in science, they have different meanings. **Accuracy** refers to how close to the true mean (μ) our estimate is. That is, if we somehow could know the true number of slugs residing on Mt. Moosilauke we could compare our estimate to it and find out how accurate we are. Obviously, we would like our estimate to be as close to the true value as possible.

In addition, we would like to avoid any **bias** in our estimate. An estimate would be biased if it consistently over- or under-estimated the true mean. Bias may arise in many ways, but one frequent source is by the selection of sample plots that are nonrandom with respect to the abundance of the target organism. For example, if we looked for slugs at Moosilauke only in sunny, dry open fields, our estimate would probably be much lower than the true abundance. **Random sampling** avoids this source of bias. A random sample is one where every potential sample plot within the study area sample has an exactly equal chance of being chosen for sampling. Random sampling is not the same as haphazard sampling. True random sampling usually requires the use a random number table (available in some books), or a random number generator (such as is contained in some calculators, most spreadsheets, and some other software packages).

In addition to obtaining an accurate, unbiased sample, we are also concerned with the **precision** of our estimates. Precision refers to the repeatability of our estimates of the true sample mean. If we were to estimate slug abundance many times and got nearly the same estimate each time, we would say that our estimate was very precise. Note that it is possible to have

accuracy without precision and *vice versa*. Sokal and Rohlf (1981) wrote:

"a biased but sensitive scale might yield inaccurate but precise weight. By chance, an insensitive scale might result in an accurate reading, which would however be imprecise, since a repeated measurement would be unlikely to yield an equally accurate weight."

If measurements are unbiased, precision will lead to accuracy. In this exercise, we are concerned mainly with precision.

Overview

We will use the computer to simulate and sample populations of two plant species, virtual beech trees (representing *Fagus grandifolia*) and virtual hobblebush (representing *Viburnum alnifolium*). The objective is to develop an intuition for the issues involved when estimating the size of a population with **quadrat sampling**. Quadrat sampling is based on measurement of replicated sample units referred to as quadrats or plots (sometimes transects or relevés). This method is appropriate for estimating the abundance of plants and other organisms that are sufficiently sedentary that we can usually sample plots faster than individuals move between plots. This approach allows estimation of **absolute density** (number of individuals per unit area within the study site). Our challenge is to identify sampling strategies that will provide satisfactory precision with minimum sample effort. Some of the factors that affect precision are:

- 1) **Measurement error.** In the real world, it is important to count organisms carefully and lay out plots accurately for good estimates of density. This is not a concern here however, because the computer will be laying out the plots and counting the plants.
- 2) **Total area sampled.** In general, the more area sampled, the more precise the estimates will be, but at the expense of additional sampling effort.
- 3) **Dispersion of the population.** Whether the population tends to be aggregated, evenly spaced, or randomly dispersed can affect precision. Note that the dispersion pattern of the same population may be different at different spatial scales (e.g., 1 x 1 m plots vs 100 x 100 m plots).
- 4) **Size and shape of quadrats.** The size and shape of the plots can affect sampling precision. Often, the optimal plot size and shape will depend on the dispersion pattern of the population.

We will explore the role of some of these factors in influencing estimates of absolute density.

The Lab

1. Double click on the Ecobeaker icon to open it.
2. Load the "Sampling" situation file by choosing 'Open' in the File menu.
You should see four windows open on the screen: (1) 'Species Grid' showing brown and green dots. This is an aerial view of our forest, 100 m on each side. The brown dots represent beech trees and the green dots represent hobbleshbush; (2) 'Sampling Parameters' is where we specify the type and number of quadrats we would like to sample. They are automatically placed randomly; (3) 'Total Population' shows the number of beech and hobbleshbush present on the grid; (4) 'Control Panel' is used to run the simulation and tell the computer when to sample. In this exercise we won't be using 'STOP', 'GO', or 'RESET'. If you accidentally hit 'GO' or 'RESET', thus changing the species abundances, you will have to reload the situation file.
3. First, run a sample to familiarize yourself with the procedure. Specify plot size and number in the 'Sampling Parameters' window. Use 5 x 5 m plots, and sample $n=20$ of them. After you enter the parameters in the window, press 'Change'. Now go to the 'Control Panel' window and press sample. One by one, a plot will appear on the screen in a randomly selected location, and you will be told how many beech and how many hobbleshbush were in that plot.
4. For this exercise, you will need to calculate the mean, standard deviation, standard error, and 95% confidence intervals for your samples (see Appendix). So before we begin, you will need to specify how you will receive the data. You could copy the results by hand as the sampling is performed, but it is easier to save the data into a file that can be opened by Excel. To do so:

Choose 'Sampling...' from the Setup menu.

Press the 'Set Save File' button to specify where you would like to save the file. You should create a separate file for each sampling run. So each time you change the sampling parameters, change the file name BEFORE you press 'Sample'. When you want to do the calculations on this data, import the data file into Excel as "Tab-delimited text".

To speed up the sampling, turn off the dialog boxes put up during the sampling runs. Choose 'Sampling...' from the Setup menu again. Press the 'Advanced Stuff' button. Set verbosity to 'No Feedback'.

5. Begin the data collection by sampling beech with a small plot size, 5 x 5 m. Assume that our funding for this study is limited, and we can only sample a total of 500 m², so set $n = 20$. As before, set the sampling parameters in the appropriate window, and press 'Change'. Now, sample the population (press 'Sample') and calculate the mean density of the plots (see Appendix 1). To facilitate comparisons with subsequent sampling using different plot sizes, convert your raw data to individuals / m². At this plot size (5X5m= 25m², you should divide each sample count by 25; do this in Excel with an equation in the adjoining your raw data (see Appendix 2 for a recommended structure for your Excel worksheet). **Record the mean on the answer sheet in the Results table, on the appropriate line. Be sure that they are in the correct units.** This is your estimate of the true mean density. If you sampled the population again with the same plot size, how close do you think the next estimate would be? We can estimate the precision of the sample based on only one run in order to find out how variable our estimates would be if we sampled the population many times. Calculate the standard deviation (SD), standard error (SE), and a 95% confidence interval (CI) (see Appendix). **Record your calculations in the Results table.**
6. You have just received a research grant, and can therefore increase our total area sampled 4-fold to 2000 m². Consider what will happen to our estimate of (1) the mean, (2) the SD, and (3) the CI. Then change the number of plots sampled (n) in the sampling parameters window to 80. Record data for BOTH beech and hobbleshbush. Sample, and calculate the mean, SD, SE, and CI for beech, as before. **Enter your results in the Results table, in the appropriate space. Again, be sure that they are in the correct units.**
7. Now consider how dispersion would affect our estimates of the mean, and our precision. First, look at the pattern of beech in our forest. Would you describe the dispersion as aggregated, evenly spaced, or random? Answer at **line 1 on the answer sheet**. A simple calculation can give us a quantitative estimate of the degree of aggregation in a population. This statistic (δ) is simply the variance divided by the mean (see Appendix). A value of 1 indicates random dispersion, values less than 1 indicate even-spacing, and values greater than 1 indicate aggregation. (In this case, the calculations should be based on the raw data units of individuals / sample plot; see Appendix 2). Calculate this statistic for beech (5 x 5 plots, $n = 80$) and evaluate how it

compares with your visual estimate. Now examine the hobblebush population. How would you describe its dispersion pattern (**3 on answer sheet**)? Calculate the mean, variance, and dispersion statistic (δ) for hobblebush (**Enter the mean in the results table. Enter δ on line 4**). How would you expect the precision for sampling hobblebush to compare with that of beech? Calculate SD, SE and CI for hobblebush (**Enter them in the results table**) Compare the SD, SE, and CI for beech and hobblebush. Which one was more precise? Why? (**line 5**).

8. Examine the effect of changing plot size and shape on our estimates of the mean density and on precision. We'll concern ourselves only with **hobblebush** in this part of the exercise. First, test what happens when we change plot size keeping everything else constant. Change the sampling parameters so that you are using a 20 x 20 m plot, keeping total area constant at 2000 m² (n=5). Now sample the population. Be sure to convert your data to the proper units to facilitate comparison with the earlier sampling. Calculate the mean, SD, SE, and CI (**Enter in the results table**). Did the precision increase or decrease compared to the 5 x 5m plots (n=80)? Briefly explain why (**line 6**)
Now change the plot shape, keeping n and total area sampled constant. Change the plot size to be 4 x 100. Leave n at 5. Sample. Convert to individuals / m². Calculate the mean, SD, SE, and CI (**Enter in the Results table**). How do these compare with the values from the 20 x 20 plots? If they are different, explain why (**line 7**) **OPTIONAL**: Sample beech for plot sizes of 20 x 20 and 4 x 100. Calculate mean, etc. and compare for other samples of beech. How did your estimates of the mean and precision change with plot size and shape? How does this compare with hobblebush? Why? (**line 7 - optional**)
9. In this exercise we have an advantage because we can know the true density of our populations. Calculate the true density (μ) of each species using the values in the 'Total population' window. Note that the total area of our forest is 10,000 m². **Enter these values in the Results table**. Which sampling scheme, of all you have tried so far, came closest to estimating the true means? (**line 8**) Was the true mean included in all of the CI's? If not, explain why. (**line 9**)
10. Based on your pilot sampling of these virtual plant populations, suggest an optimal plot size and shape for hobblebush.

Appendix 1: a primer on the statistical description of populations

Problem: How to estimate the population mean from a sample (e.g., How can we estimate the mean density of beech in a forest where we sampled $n = 20$ quadrats that were each 5×5 m?)

There are two important components to estimating the mean. First is the estimate itself, and second is the variability associated with that estimate. You are probably familiar with the first component; it is simply the average.

1. To find the average density of a sample of plots, add the number of individuals found in each plot and divide by the number of plots. This is now the average number of individuals per plot and can be expressed as an **absolute density** (e.g., 4.2 beech / 25 m^2). This sample mean (\bar{x}) represents an estimate of the true population mean (μ).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Where: x_i = each sample observation (individuals / plot)

n = sample size (number of plots)

In Excel, the mean can be calculated as:

=AVERAGE(number1, number2...) or =AVERAGE(firstcell reference:lastcell reference)

2. The **standard deviation** is one measure of variability among the population of plots in the study area. The sample standard deviation (SD) represents an estimate of the true population standard deviation (σ).

$SD \approx$ the average amount by which a sample differs from the sample mean. (In a normal distribution, 68% of the population lies within ± 1 SD of the mean, and 96% lies within ± 2 SD's of the mean). Thus the SD is a measure of the variability in the population.

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

where: x_i = each sample observation

\bar{x} = sample mean

n = sample size

In Excel, the standard deviation can be calculated as:

=STDEV(number1, number2...) or =STDEV(firstcell reference:lastcell reference)

3. The **standard error** (SE) is a measure of the repeatability of the population estimate (\bar{x}).

$SE \approx$ the average amount by which a sample mean (\bar{x}) differs from the true population mean (μ). (Given a normal distribution and random, independent samples, $\approx 68\%$ of the possible sample means lie within ± 1 SE of the true, and $\approx 96\%$ lies within ± 2 SE's of the mean). Thus the SE is a measure of the variability that could be expected of repeated samples from a population. The SE is an estimate of the SD around the \bar{x} 's that would be obtained from repeated samples of size n from the study population. It estimates the precision of a sampling scheme.

$$SE = \frac{SD}{\sqrt{n}}$$

where: SD = sample standard deviation

n = sample size

In Excel, calculate SE by referencing the cell holding the SD and divide by the square root of the sample size: e.g., '= C6/(n^0.5)', where C6 is the cell holding the SD , and n = sample size.

4. The precision of an estimate of the sample mean can be expressed in terms of a **confidence interval** (CI). A confidence interval represents the range of values that can be assumed to contain the true mean with a certain (specified) probability. For example, suppose we wanted to find out the range of values that would contain the true mean with a probability of 0.95. In other words, if we were to take 100 random samples of our population, 95 of the calculated CI's based on those samples would contain the true population mean. This would be called a 95% confidence interval. Other values, (e.g. 99%, 99.9%) may be used, but 95% is most common. The higher the percentage, the wider the confidence interval, thus it becomes a trade-off between a range narrow enough to be meaningful, and precise enough to be useful. The confidence interval is expressed as a lower and upper limit (L_1 and L_2 , respectively), and is calculated as follows:

$$L_1 = \bar{x} - t_{\alpha[n-1]}SE \quad \text{and} \quad L_2 = \bar{x} + t_{\alpha[n-1]}SE$$

where L_1 = lower confidence limit

L_2 = upper confidence limit

\bar{x} = sample mean

t = the t statistic, from statistical table or Excel

α = the probability associated with the critical t - value. $\alpha = 1 - P$ where P is the desired confidence of the interval (e.g. for a 95% CI, $\alpha = 0.05$)

n = the sample size

SE = standard error

To calculate confidence limits in Excel, be sure to take advantage of previously calculated values for the mean and the standard deviation. Note that Excel uses the SD in its calculation, and NOT the SE. To calculate a CI in Excel, the lower limit is '= (mean) - TINV(alpha, n-1) * SE'. The upper limit is identical, except the TINV() * SE term is added to the mean, not subtracted. To calculate these limits, substitute cell numbers or calculated values for the variables in italics.

3. A simple statistic for describing the **spatial dispersion** of a population is simply the variance ($= SD^2$) divided by the mean. It is denoted here as δ , delta.

$$\delta = \frac{SD^2}{\bar{x}}$$

Values of δ greater than 1 indicate aggregation, while values less than 1 indicate a uniform dispersion. A value of 1 indicates random dispersion.

Appendix 2: Example spreadsheet

Sampling 20 quadrats, each 5 by 5 at time step 365

X_pos	Y_pos	Individuals / sample plot		Individuals / m2	
		Hobblebush	Beech	Hobblebush	Beech
52	78	0	1	0	0.04
62	39	0	1	0	0.04
87	94	0	0	0	0
93	8	0	0	0	0
48	74	0	0	0	0
34	53	4	0	0.16	0
74	11	0	0	0	0
13	30	0	2	0	0.08
74	81	0	0	0	0
59	79	0	2	0	0.08
48	74	0	0	0	0
34	53	4	0	0.16	0
74	11	0	0	0	0
13	30	0	1	0	0.04
72	78	0	0	0	0
91	13	0	1	0	0.04
6	19	1	0	0.04	0
26	31	0	1	0	0.04
93	52	0	0	0	0
94	77	0	1	0	0.04
Mean		0.450	0.500	0.0180	0.0200
SD		1.234	0.688	0.0494	0.0275
N				20	20
SE				0.0110	0.0062
CI(lower)				0.0101	0.0112
CI(upper)				0.0396	0.0321
δ		3.39	0.95		