SAMPLE SIZE ADEQUACY:

A LFA Procedure for Estimating the Adequacy of Sampling

Aim:

The specific aim of this document is to provide landscape restoration practitioners with a procedure for evaluating whether the number of query zones evaluated along a landscape function analysis (LFA) transect is adequate for accurately estimating LFA indicators. An overview of LFA procedures for measuring indicators of landscape functionality is provided in chapter 13 in our book, “Restoring Disturbed Landscapes” (Island Press, 2011).

Although the adequacy of sampling procedure briefly described below is specific to LFA, it is generally applicable to most scientific studies. For those interested in a thorough discussion of sampling adequacy, which is only one component of what statisticians call Power Analysis, we refer you to chapter 7 in a book by Quinn and Keough on “Experimental Design and Data Analysis for Biologists” (Cambridge University Press, 2002).

Preamble:

The number of samples, N, observed, for example, in quadrats or query zones along LFA transects, that are needed to accurately estimate a parameter (e.g., mean LFA nutrient cycling index) for a population (e.g., for the potential for a specific vegetation patch type to cycle nutrients in a rehabilitated landscape) depends on the spatial variability of the population (e.g., variation in index data among patches within the type). If this variability is high, a larger sample size will be needed to achieve an acceptable level of accuracy for the parameter of interest. In other words, if the potential for nutrient cycling varies greatly between vegetation patches across the rehabilitated landscape, then soil surface nutrient cycling indicators will need to be observed in a large number of query zones along LFA transects to accurately estimate the mean value for the nutrient cycling index.

Background:

The key words or terms in the above Preamble are “number of samples” (i.e., N), “to accurately estimate a parameter” (mean), and “variability of the population” (variance). These terms are usefully related to each other. Let’s look at these relationships in two ways: graphically and statistically.

<u>Graphical relationship between mean and variance</u>

First, let’s graph an example of how mean and variance change with increasing number of samples, N. For this example, we use data for observations on LFA indicators in 12 query zones selected from a sample of vegetation patches identified as “grassy sward” along LFA transects positioned in a woodland located in Western Australia, which is in good condition because it’s ground layer is intact (Figure 1). An example of the same type of woodland in poor condition has lost its ground layer vegetation (Figure 2).

Figure 1. A woodland in ‘good condition’ because it has an intact ground-layer of grasses and shrubs.

Figure 2. A nearby woodland in ‘poor condition’ because disturbance has greatly reduced the ground-layer of grasses and shrubs.

In the good condition woodland, LFA indicators were observed in 12 grassy sward query zones along LFA transects; these observations were used to derive the following 12 nutrient cycling index values:

59, 71, 73, 63, 65, 68, 71, 63, 69, 69, 68, 66

Now, we plot the ‘running’ means after every 2 observations for this set of 12 observations (i.e., means for 2, 4, 6, 8, 10 and 12 observations; Figure 3). Note how the mean goes up slightly but then becomes relatively constant. The mean value after 12 observations is 67.1. This mean LFA nutrient cycling index value is quite typical for grassy sward patches in woodland sites in good condition.

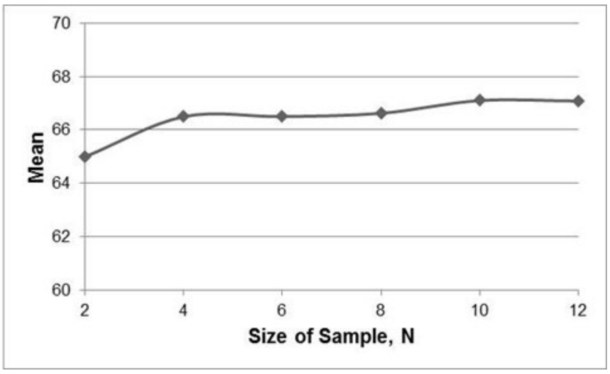

Figure 3. Running means for the LFA nutrient cycling index in grassy sward patches located in good condition woodland.

Now let’s plot the ‘running’ variances after every 2 observations for this set of 12 observations (i.e., variances for 2, 4, 6, 8, 10 and 12 observations; Figure 4).

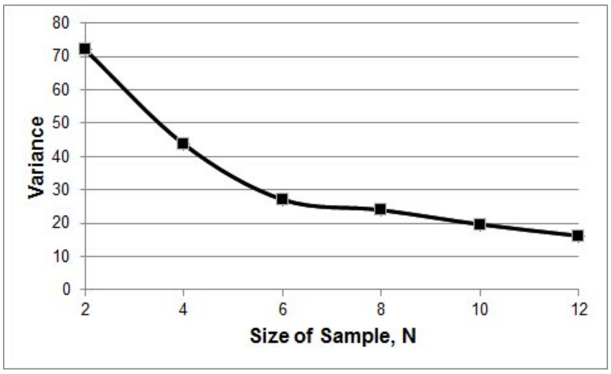

Figure 4. Running variances for the LFA nutrient cycling index in grassy swards located in a good condition woodland.

Note how the variance starts out high (72) and declines to a value of 16.3 after an N of 12. A declining variance is expected as the number of samples included in estimating the variance increases.

Statistical relationship between mean and variance

As evident from the above graphs, there is a statistical relationships between a sample mean and variance, and we can use this relationship to calculate the sample size (N) needed to achieve a defined level of accuracy about a mean. The statistical relationship between a mean (Xbar), its variance (xVar), and sample size (N) is given in the equation used to estimate a Standard Error, SE (also referred to as the Standard Deviation of the Sample Mean). This equation takes the form:

$$SE = \text{sqrt}(xVar/N).$$

Now, we square both sides of this equation and rearrange it to solve for sample size, N:

$$(SE)^2 = xVar / N$$

$$N = xVar / (SE)^2$$

In other words, N equals the sample Variance (xVar) divided by the 'Standard Deviation of the Sample Mean' squared.

Next, we define the level of accuracy about the sample mean we are willing to accept, and we call this level our "Standard Deviation of the Sample Mean", as in the above equation. For many ecological studies, a 10% level of accuracy about a sample mean would be quite acceptable. So, let's define:

Standard Deviation of the Sample Mean = 0.10 times Xbar

If a higher level of accuracy is required for estimating means (e.g., a 5% level of accuracy), then one can define the "Standard Deviation of the Sample Mean" as:

Standard Deviation of the Sample Mean = 0.05 times Xbar

Procedure to estimate Minimum Adequate Sample Size (N)

From the above, we now have the equations we need to calculate the sample size (number of observations, N) needed to estimate a mean with an accuracy of 10% or, if needed, 5%. To illustrate this calculation, let us return to our sample of 12 observations in query zones used to assess soil nutrient cycling potential for grassy sward patches located in a woodland site in good condition. In practice, one would collect what is called a 'pilot' sample of, say, 10 observations in 10 query zones in the landscape site of interest, and then check to see if 10 samples was a minimum adequate sample size for a specified patch type in this landscape.

1. Calculate the mean and variance for the pilot sample

For the woodland site, after 12 observations in grassy sward query zones along LFA transects in a woodland site, the sample mean, Xbar = 67.1, and the variance about these observations, xVar =16.3.

2. Estimate the acceptable accuracy for the mean

We define our acceptable (10%) level of accuracy about the mean for the nutrient cycling index as:

$$\text{Standard Deviation of the Sample Mean} = 0.10 \times 67.1 = 6.71$$

For a 5% level of accuracy about the mean for the nutrient cycling index:

$$\text{Standard Deviation of the Sample Mean} = 0.05 \times 67.1 = 3.35$$

3. Estimate the minimum adequate sample size

Now, for a 10% level of accuracy, we solve our minimum adequate sample size equation, N = Variance/(Standard Deviation of the Sample Mean)^2 as:

$N = xVar/(0.10 * Xbar)^2$

$N = 16.3 /(6.71)^2 = 16.3 / 45.0$

N = 0.36 for a 10% error about the sample mean, or

N = 1 (when rounding-up to next largest whole integer)

This says that, given our sample observational data to estimate a nutrient cycling index, the minimum adequate sample size (number of observations) needed to achieve a 10% level of accuracy about the sample mean is only 1, which is good news. Of course, to cover cases where grassy swards and other patch types on other sites may be more variable than this, a larger sample size of, say, 5 could be used.

For an acceptable level of accuracy about the mean of 5%, the equivalent calculations are:

$N = xVar / (0. 05 * Xbar)^2$

$N = 16.3 / (3.35)^2$

N = 16.3 / 11.2

N = 1.4, or by rounding up

N = 2 to achieve a 5% error about the sample mean

As expected, a slightly larger sample size is required to achieve a greater accuracy (smaller error) about the mean.

The minimum adequate sample sizes for all LFA indices needs to be determined, and as more site data is obtained, we recommend that sample size adequacy be frequently re-evaluated to ensure that LFA index values are being accurately estimated for different patch types and for other sites.

References

Tongway, D.J. and Ludwig, J.A. 2011. Restoring Disturbed Landscapes: Putting Principles into Practice. Island Press, Washington, D.C. USA.

Quinn, G.P. and Keough, M.J. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge, UK.