Statistical Sampling Methods for Soils Monitoring

Ann M. Abbott, Moscow Forestry Sciences Laboratory, Rocky Mountain Research Station, Moscow, ID.

Abstract—Development of the best sampling design to answer a research question should be an interactive venture between the land manager or researcher and statisticians, and is the result of answering various questions. A series of questions that can be asked to guide the researcher in making decisions that will arrive at an effective sampling plan are described, and a case study is used to explain how the sampling effort was designed for the Forest Soil Monitoring Protocol.

Introduction

In general, the goal of statistics is to be able to make inferences about a population based on information gathered from a sample of units from that population. For the inferences to have meaning, then, the sample must be representative of the entire population in question.

To appropriately draw a sample from a population, there are several considerations that must be met.

- The sampling must be done in such a way that it will meet the objectives of the research.
- The sample itself must be representative of the population.
- The sampling plan must be feasible, and the plan must be cost effective.

An appropriate sampling plan is the result of answering a series of questions, and it is the answers to the questions that lead to the best sampling design, data analysis methods, and subsequent interpretation of the analysis. Knowing the questions to ask, therefore, is the key to designing a good sampling plan. Some of questions that must be asked include:

- What are the objectives of the research?
- What is the population about which inferences will be made?
- What are the sampling units?
- What is the translation of the objectives into specific questions that can be answered with measurements from the sampling units?
- What preliminary information is available about the population?
- What choice of sampling design will be used?
- What sample size is necessary to answer the research questions with acceptable accuracy?
- Are there any auxiliary variables that can provide additional information?
- How will the randomization be performed?
- How will the results of the sampling effort be recorded?
- How will the data be analyzed?

This paper will explain how to design a sampling plan using the Forest Soil Monitoring Protocol as a Case Study.

Questions

What Are the Objectives of the Research?

It is imperative that clear objectives be stated prior to beginning any data collection. They must be clearly and explicitly stated, and the reasons for undertaking the research must also be documented. A subsequent step in defining the objectives of the research is to translate the objectives into precise questions that the sample measurements can answer. The translation of the objective into precise questions is the link between the initial research question and a question that may be answered with sampling and statistics. To answer a question about a population with statistics, the question must be asked in terms of measurements that may be taken on individuals within a population.

An example of defining and translating objectives is as follows:

- The amount of soils disturbed by management activities must be documented.
- The objective, then, is to quantify the disturbance. While this is a good objective, it is not, as stated, something that can immediately be answered with sampling.
- A precise question is "What proportion of points within a transect is compacted to 10 cm?"

What Is the Population About Which Inferences Will Be Made?

Once the initial objectives or research questions have been formulated, the population about which information is desired must be defined, which can be a somewhat circular process. Often, the definition step will refine the overall population about which information is desired into a population from which a sample may be drawn. When considering the population, it is helpful to think of it as a collection of individuals or sampling units that can be listed. Such a list may also be used as the sampling frame, from which the sample will be drawn. Defining the population as a collection of individuals that may be listed will determine whether any constraints are present that will limit the overall population into a smaller segment that can be sampled. This will also help ensure that the sampling units are representative of the population. It is possible that the population from which the sample is drawn may be different from the population as a whole if the entire population cannot be sampled.

What Are the Sampling Units?

Sampling units are defined as 'non-overlapping collections of elements from the population that cover the entire population.' A successful sampling scheme includes the selection of an appropriate sampling unit. The sampling unit is the individual within the population on which measurements and inferences will be made, so it is critical that the unit be carefully defined and possible to measure, as well as meet the objectives of the study. The sampling unit is also the subject of the randomization scheme for the study. Some examples of sampling units include quadrats, leaves of a plant, individual organisms, belt transects, or points.

Further questions that should be asked when considering the choice of sampling unit include:

- Are the sampling units naturally defined?
- If not, how will they be defined?
- Is the number of sampling units finite?
- If it is finite, is the total number of units in the population large enough to ignore finite sampling considerations?
- Is the definition of the sampling units appropriate to the objectives?

There are some important considerations that should be made when choosing the unit for sampling. The sampling unit must be the unit upon which you wish to make inferences and estimates. It is the subject of the randomization process used in the sampling design. Although it is common that the measurements taken in the study are performed on the sampling unit, it is not a requirement and usually occurs when the measurements cannot be performed on the randomization units. When the objects upon which the measurements are taken are not those which were randomly selected, however, the analysis is performed on the randomized units.

Sampling units for estimates of characteristics of a particular area can be either point samples or area samples. For either a point sample or an area sample, they should be sampled without replacement to ensure that any particular sampling unit is only sampled once. Point samples allow inferences to be made on the number of observations in the sample, and the inferences are often made on the means or percentages from the sample observations. Area samples are generally measured with densities of percent of area covered, and inferences are made by extrapolating the sample density to the entire area. Area samples can yield more detailed information but can also be more time consuming to carry out.

Translating the Objectives

The translation of the research objectives into specific questions that can be answered with measurements from the sampling units is often the most challenging step and a good time to consult with a statistician. The translation is the integration of the research question into the quantitative question "What exactly is to be estimated or tested?"

Part of the translation step will identify whether the required estimates are proportions, totals, means, totals or means over subpopulations, or some other quantitative estimate. Constructing the blank data sheet for recording observations will assist in the translation step as that step will clearly identify the measurements that will be taken. It is critical that once the data sheets have been constructed and the measurements to be taken are identified that one revisit the research question to ensure that the observations and resulting summaries will, in fact, answer the research question.

What Preliminary Information Is Available About the Population?

Information that can be gathered about the population of interest prior to sampling can help ensure that the sampling design will be successful in providing the necessary information to answer the research question. Such information includes whether estimates of the likely variability are available. If variability estimates are available, they can be used to determine the necessary sample size to provide estimates within specified confidence levels.

If there are no variability estimates, then one should determine whether a pilot study is desirable and/or feasible. A pilot study can be used to determine variability estimates as well as to test the sampling methods.

If there are factors within the population that affect the results of the observations, it is possible that such factors can be used to stratify the population into separate groups for randomization. This information, if available prior to sampling and when used to develop a stratified sampling design, can reduce the variability around the estimates, thus improving the statistical efficiency of the estimates.

Accounting for Variability

The variation that is inherent in soils data must be accounted for during the design phase of a soil sampling plan, including the sampling design, data collection procedures, and data analysis. Researchers have long been cautioned about failing to consider the variability in soil sampling when dealing with any study of the soils system (Cline 1944).

Variability can be accounted for by ensuring that the sample adequately covers the entire population, by reporting the variability estimates along with central tendency estimates and by reporting interval estimates.

A good sampling design will use an interactive approach to balance the data quality needs and resources with designs that will either control variation, stratify to reduce variation, or reduce the influence of variation on the decision process.

Precision, Bias, and Accuracy

Precision is a measure of the reproducibility of the measurements of a particular soil condition or constituent. Precision is increased as the variability around the estimates is decreased. If the variability in the observations is constant, precision can also be increased by taking a larger number of measurements (increasing the sample size). The statistical techniques seen in soil sampling are designed to measure precision and not accuracy.

Bias is a systematic error that contributes to the difference between the mean of a large number of test results and an accepted reference value. Bias is often the result of an imperfect measurement technique (the characteristic measured does not match the characteristic in question in a systematic way) or an imperfect measurement instrument (the measurement tool must be calibrated).

Accuracy is the correctness of the measurement and cannot be directly measured. It is the sum of precision and bias, and can be improved by taking care that the estimates are as precise as is required and that bias is as small as possible.

Sampling Designs

The choice of a sampling design often depends on what is available for a sampling frame, whether the population can be divided into a natural grouping in terms of the measurement variables, variability within the population, and the cost of sampling. There are three initial questions that can be posed when considering the four commonly used sampling designs:

- Does the population contain a natural grouping in terms of the variables that will be measured?
- Does the grouping variable affect the results of the measurement variable?
- Can the efficiency of the sampling effort be improved by separating the population into such groups?

If there are no natural groupings, then two possible sampling designs are Simple Random Sampling and Systematic Random Sampling. The choice between these two designs depends on the answer to the question "Is a comprehensive list of sampling units available?"

Simple Random Sampling—Simple Random Sampling is the basis for most other sampling designs. It is used when a comprehensive list of all population units is available and either no information is known about the population or a natural grouping does not exist. A randomization scheme is used to select individuals for measurement in which each element in the population has an equal probability of being selected. Simple random sampling is the basis for all probability sampling techniques and is the point of reference from which modifications to increase sampling efficiency may be made. Alone, simple random sampling may not give the desired precision.

A formal definition of simple random sampling is:

> If a sample of size n is drawn from a population in such a way that every possible sample of size n has the same chance of being selected, the sampling procedure is called *simple random sampling* and the sample thus obtained is called a *simple random sample*.

To draw a simple random sample, all of the possible elements in the population are listed to form a sampling frame. A randomization scheme, often from a random number table, is used to draw elements from the sampling frame without replacement.

Common estimators calculated for continuous variables are the estimator of the population mean, the variance of the population mean (to evaluate the goodness of the estimated mean), a confidence interval around the estimated mean, and a required sample size to estimate the population mean. For binomial variables (those with either a yes or no response), the estimator of the proportion of the population possessing the yes response is often of interest, along with it's variance, confidence interval and sample size.

To calculate the estimators, we use the following equations:
Estimator of the population mean:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad [1]$$

where y_i is the observation from element *i* and *n* is the number of elements sampled.
Estimator of the population variance:

$$\hat{V}(\bar{y}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) \qquad [2]$$

where s^2 is the sample standard deviation.

$\left(\frac{N-n}{N}\right)$ is the finite population correction factor (fpc). When *n* is small relative the population size *N*, the fpc is close to unity.

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \qquad [3]$$

Confidence interval around the estimated mean:

$$\bar{y} \pm t_{\alpha/2, n-1}\sqrt{\hat{V}(\bar{y})} \qquad [4]$$

where $t_{\alpha/2, n-1}$ is the coefficient from *t* distribution with *n* -1 degrees of freedom.
Estimator of the population proportion:

$$\hat{p} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad [5]$$

Estimated variance of the population proportion:

$$\hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}\left(\frac{N-n}{N}\right) \qquad [6]$$

Confidence interval around the population proportion:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{p})} \qquad [7]$$

Systematic Random Sampling—Systematic Random Sampling is an alternative to simple random sampling. It is used when a comprehensive list of sampling units is not available but an estimate of the total number of units within the population can be obtained. The randomization aspect occurs in the starting point. When systematic random sampling is performed, the sample size must be determined so that a sampling interval may be computed. When a random start is selected within the first sampling interval, then each subsequent element from the following intervals are also random by default.

The goal of systematic random sampling is to provide better coverage of the study area or population than that provided by a simple random sample or from a stratified random sample, and is a simple random sample based on spatial distribution over the population. To use systematic random sampling, some estimate of the total number of sampling units in the population must be estimated. The required sample size must also be known so that the interval for sampling can be calculated.

Systematic random sampling is a useful alternative to simple random sampling:

1. Systematic sampling is easier to perform in the field and hence is less subject to selection errors by field workers than either simple random sampling or stratified random sampling.
2. Systematic random sampling can provide greater information per unit cost than simple random sampling can provide.

Transect sampling is a version of systematic random sampling, and when using transects, they should be randomly oriented or the starting point should be randomly chosen.

A danger in systematic random sampling is that if the sampling interval is chosen in such a way that it matches any periodicities in the population, the resulting estimates could be biased. Knowledge of the population is useful to avoid this danger so that care can be taken to avoid sampling along any periodicities.

Estimators from systematic random sampling can be calculated using the same equations as those used for simple random sampling.

If there are natural groups within the population, then the question becomes "Are the groups likely to be similar to each other in terms of the measurement variables or are the groups different?" An alternative phrasing for this question is "Is the variability within groups larger than the variability between groups?"

Stratified Random Sampling—Stratified random sampling is used when the groups are different from each other, or when the variability is larger between the groups compared to variability within groups. Each group (stratum) is sampled individually, using either a simple random sample or a systematic random sample.

Prior knowledge of the sampling area and information obtained from background data are required for stratified random sampling. The goal is to increase precision and control sources of variability in the data, and a potential result is that the overall sample size may be reduced. For stratified random sampling to be efficient (the overall variability estimates from a stratified design are smaller than those from simple random sampling), the variability between strata must be larger than variability within strata.

The advantages of stratified random sampling include obtaining estimates for subgroups, potentially more precise estimates than those from simple random sampling, and can be more convenient to implement. Disadvantages are that prior information about the population is necessary and the computations are more complex.

Some additional notation is required for computational formulas for stratified random sampling.

L = number of strata
N_i = number of sampling units in stratum i
N = number of sampling units in the population = $N_1 + N_2 + \ldots + N_L$

Estimator of the population mean from a stratified random sample:

$$\bar{y}_{st} = \frac{1}{N}[N_1\bar{y}_1 + N_2\bar{y}_2 + \cdots + N_L\bar{y}_L] = \frac{1}{N}\sum_{i=1}^{L} N_i\bar{y}_i \qquad [8]$$

Estimated variance of the mean from a stratified random sample:

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2}[N_1^2\hat{V}(\bar{y}_1) + N_2^2\hat{V}(\bar{y}_2) + \cdots + N_L^2\hat{V}(\bar{y}_L)] \quad [9]$$

Confidence interval around the estimated mean:

$$\bar{y}_{st} \pm t_{\alpha/2, n-1}\sqrt{\hat{V}(\bar{y}_{st})} \quad [10]$$

The goal of the allocation scheme to divide the overall sample size into the different strata depends on three factors:

3. The total number of elements in each stratum.
4. The variability of observations in each stratum.
5. The cost of obtaining an observation from each stratum.

The number of elements in each stratum affects the quality of information in the sample. A sample size 20 from a population of 200 elements should contain more information than a sample of 20 from 20,000 elements. Thus, larger sample sizes should be assigned to strata containing larger numbers of elements. Variability must be considered because a larger sample is needed to obtain a good estimate of a population when the observations are less homogeneous. If the cost of obtaining a sample varies from stratum to stratum, smaller samples from strata with higher costs is advisable when the goal is to keep the cost of sampling at a minimum.

An approximate allocation that minimizes cost for a fixed value of $\hat{V}(\bar{y}_{st})$ or that minimizes $\hat{V}(\bar{y}_{st})$ for a fixed cost:

$$n_i = n\left(\frac{N_i s_i / \sqrt{c_i}}{N_1 s_1/\sqrt{c_1} + N_2 s_2/\sqrt{c_2} + \cdots + N_L s_L/\sqrt{c_L}}\right) \quad [11]$$

where N_i denotes the size of the ith stratum, s_i^2 is the estimated variance from the ith stratum, and c_i is the cost of obtaining a single observation from the ith stratum.
Estimator for the population proportion from stratified random sample:

$$\hat{p}_{st} = \frac{1}{N}\left(N_1\hat{p}_1 + N_2\hat{p}_2 + \cdots + N_L\hat{p}_L\right) \quad [12]$$

Estimator of the variance of the estimated proportion from a stratified random sample:

$$\hat{V}(\hat{p}_{st}) = \frac{1}{N^2} N_1^2\hat{V}(\hat{p}_1) + N_2^2\hat{V}(\hat{p}_2) + \cdots + N_L^2\hat{V}(\hat{p}_L) \quad [13]$$

Confidence interval around a proportion from a stratified random sample:

$$\hat{p}_{st} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{p}_{st})} \quad [14]$$

Approximate allocation that minimizes cost for a fixed value of $\hat{V}(\hat{p}_{st})$ or minimizes $\hat{V}(\hat{p}_{st})$ for a fixed cost:

$$n_i = n\frac{N_i\sqrt{\hat{p}_i(1-\hat{p}_i)/c_i}}{N_1\sqrt{\hat{p}_1(1-\hat{p}_1)/c_1} + N_2\sqrt{\hat{p}_2(1-\hat{p}_2)/c_2} + \cdots + N_L\sqrt{\hat{p}_L(1-\hat{p}_L)/c_L}} \quad [15]$$

Post-stratification can be used when stratification is appropriate for some key variable, but cannot be done until after the sample is selected. This is often appropriate when a simple random sample is not properly balanced according to major groupings. While the mean from a post stratification scheme is calculated in the same way as for a

designed stratified random sample, the variance must be estimated differently since the stratification was not designed into the plan.
Estimated variance of the mean from post stratification:

$$\hat{V}_p(\bar{y}_{st}) = \frac{N-n}{Nn}\sum_{i=1}^{L} W_i s_i^2 + \frac{1}{n^2}(1-W_i)s_i^2 \qquad [16]$$

where W_i is the weight proportion for each stratum.

Cluster Sampling—Cluster sampling is used when the groups are similar to each other (there is more variability within groups than among groups). Here, the clusters themselves are randomly sampled so that not every cluster within the population is sampled. In some cases, individuals within clusters are also randomly sampled for measurement (multistage sampling), and in others every element within the cluster is sampled.

Cluster sampling can be less costly than simple or stratified random sampling if the cost of obtaining a frame that lists all population elements is very high or if the cost of obtaining observations increases as the distance separating the elements increases.

To calculate the estimators obtained in a cluster sample, we use the following equations:
Estimator for the sample mean from a cluster sample:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{m} m_i} \qquad [17]$$

where m_i is the number of elements in cluster *i*.
Estimated variance of the sample mean from a cluster sample:

$$\hat{V}(\bar{y}) = \frac{N-n}{Nn\bar{M}^2} s_r^2 \qquad [18]$$

where

$$s_r^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}m_i)^2}{n-1} \qquad [19]$$

and $\bar{M} = \frac{\sum_{i=1}^{n} m_i}{N}$ (the average cluster size for the population), and can be estimated with $\bar{m}_i = \frac{1}{n}\sum_{i=1}^{n} m_i$ (the average cluster size for the sample).

Confidence interval around the estimated mean for a cluster sample:

$$\bar{y} \pm t_{\alpha/2,n-1}\sqrt{\hat{V}(\bar{y})} \qquad [20]$$

Estimate of the population proportion from a cluster sample**:**

$$\hat{p} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} \qquad [21]$$

Estimated variance of the estimated population proportion from a cluster sample:

$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\overline{M}^2}\right)s_p^2 \qquad [22]$$

where $s_p^2 = \dfrac{\sum_{i=1}^{n}(y_i - \hat{p}m_i)^2}{n-1}$

Confidence interval around a proportion from a cluster sample:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{p})} \qquad [23]$$

To aid in the selection of the best sampling design, a series of questions can be asked that will help make the best choice:

- If there is no information on population groupings, will simple random sampling or systematic random sampling better meet the objectives?
- Is simple random sampling likely to be effective?
- If not, have the reasons for not using simple random sampling been clearly stated?
- If systematic random sampling is chosen, what interval will separate the sampling units?
- Is there a likelihood that the interval will coincide with periodicity in the data?
- If so, what steps will be taken to avoid the resulting bias in the estimates?
- If there is a grouping in the population, will stratification improve the precision of the estimates?
- Has the efficiency of the stratification been calculated?
- What is the basis of the stratification?
- How will the sampling units be allocated?
- If there is a grouping in the population, is there an advantage to cluster sampling?
- Has the efficiency of using clusters been calculated?

What Sample Size Is Necessary to Answer the Research Questions With Acceptable Accuracy?

Once the sampling design has been chosen, the number of observations (sample size) must be calculated. Sample calculations are based on the variability within the population and the desired precision of the estimate (the confidence level). In order to calculate sample size, one must obtain an estimate of the variability within the population to be sampled, either from prior data or from a pilot study. One must then decide what level of confidence is required for the estimates, realizing that as the confidence level increases, so does the number of observations required to make the estimate.

Sample sizes for the four sampling designs described here are as follows:

Sample size to estimate the population mean with an interval width w for a simple random sample:

$$n = \frac{t^2_{\alpha/2,n-1}s^2}{D} \qquad [24]$$

where

$$D = \frac{w^2}{4}$$

Sample size to estimate the population proportion with an interval width *w* for a simple random sample:

$$n = \frac{N\hat{p}(1-\hat{p})}{(N-1)D + \hat{p}(1-\hat{p})} \quad [25]$$

Sample size for estimating the population mean from a stratified random sample:

$$n = \frac{\sum_{i=1}^{L} N_i^2 s_i^2 / w_i}{N^2 D + \sum_{i=1}^{L} N_i s_i^2} \quad [26]$$

where w_i is the fraction of observations allocated to stratum *i*.
Sample size to estimate a population proportion from a stratified random sample with interval width *w*:

$$n = \frac{\sum_{i=1}^{L} N_i^2 \hat{p}_i (1-\hat{p}_i) / w_i}{N^2 D + \sum_{i=1}^{L} N_i \hat{p}_i (1-\hat{p}_i)} \quad [27]$$

Sample size to estimate the population mean from a cluster sample with interval width *w*:

$$n = \frac{N s_r^2}{ND + s_r^2} \quad [28]$$

where $D = (w^2 \overline{M}^2)/4$.
Sample size to estimate the proportion from a cluster sample:

$$n = \frac{N s_p^2}{ND + s_p^2} \quad [29]$$

where $D = w^2 \overline{M}^2 / 4$ and $s_p^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{p} m_i)^2}{n-1}$

How Will the Randomization be Carried Out?

It is critical that the randomization be carried out according to an objective mechanism. The sampling units must be chosen by an explicit randomization procedure that should be documented in the research. Any constraints in the sampling should be documented as well.

How Will the Results of the Sampling Effort be Recorded?

It can aid the sampling design process to create the data sheets for recording results of the sampling early on in the process. This will clarify the variables that will be measured and recorded, and will guide the analysis procedures.

How Will the Data be Analyzed?

The analysis methods that will be used to answer the research questions should be determined prior to collecting the data. Once the data sheets are created, one can see the data structure, which will help with this step. It is critical to check again, to make sure that the variables collected and the analysis methods will meet the objectives of the research.

Case Study in Sampling Design: The Forest Soil Monitoring Protocol

The goal in developing a soil monitoring protocol for the Northern Region was to develop an easy-to-implement, cost effective and statistically defensible monitoring protocol for disturbance. The sampling design and analysis methods were arrived at by answering the questions illustrated in the previous section, and are described here.

Stating the objectives: The objective of the sampling effort was to characterize the activity area in terms of management related disturbance.

Defining the population: The population was defined to be all possible 'points' within the activity area.

What are the sampling units? The sampling units were defined as points along a transect where a point is a 6-inch radius. Since there are an infinite number of possible points in the population, finite sample correction factors do not need to be used.

What is the translation of the research objectives into specific questions that can be answered with measurements from the sampling units? It was decided to characterize the amount of disturbance related to management within a unit by measuring forest floor depth and observing a series of binomial (presence/absence) variables, such as presence of forest floor, displacement of topsoil, mixing of topsoil and subsoil, presence of erosion, presence of rutting, presence of burning, presence of compaction, and presence of five forest floor variables. By using the percents of observation that record 'present' for these variables, an estimate of management-related disturbance can be made.

What preliminary information about the population is available? The size and shape of the activity area is known, and in some cases soils information is available. In most cases, site specific estimates of variability are not known. Harvest history is generally available.

What sampling design will be used? Since there is not always information about groupings within activity areas, neither stratified nor cluster sampling were chosen as the first choice in sampling design. It is to be noted, however, that considerations are made for the use of both of these designs within the protocol when such information is available. For the ease of obtaining observations and to ensure that the entire activity is accounted for within the sampling, systematic random sampling was chosen as the optimal design, using a line transect to choose observation points.

What sample size is necessary to answer the research questions with acceptable accuracy? The proportions of the binomial indicator variables listed above were used to choose the sample size, all with an interval width of ±5 percent of the estimated proportion. The protocol allows for varying levels of confidence to be used with direction from the line officer. Once the sample size is computed for each of the individual variables, the largest sample size is chosen for sampling to be conservative. The first 30 observations made along the transect were used to calculate the site specific variability for each activity area.

How will the randomization be performed? The observations are randomized by choosing a random orientation for the beginning of the transect. Subsequent turns in the transects are made by choosing an angle in advance on which to turn when the transect reaches the activity area boundary.

How will the results of the sampling effort be recorded? An Excel spreadsheet was developed to record the results of sampling. Observations can either be recorded on a paper sheet or directly into an electronic data recorder.

How will the data be analyzed? Confidence intervals are computed for each of the indicator variables, along with the estimated proportions. Summaries for multiple areas can be calculated using the methods for stratified random sampling.

References

Devore, Jay L. 1995. Probability and statistics for engineering and the sciences. 4th Edition. Duxbury Press. 743 p.

Raj, D. 1968. Sampling theory. New York: McGraw-Hill. 302 p.

Scheaffer, Richard L.; Mendenhall, William; Ott, Lyman. 1996. Elementary survey sampling. 5th Edition. Duxbury Press. 501 p.

The content of this paper reflects the views of the authors, who are responsible for the facts and accuracy of the information presented herein.