FRPA Resource Evaluation Program
Scientifically Valid Evaluations of Forest Practices under the *Forest and Range Practices Act*

The FRPA Evaluator
Technical Note #3
October 2004

Why the Units We Evaluate Should be Randomly Selected

Prepared by Wendy Bergerud, Research Branch, BC Ministry of Forests

Introduction

An initial step in the design of any evaluation is to clearly specify what in the physical world is to be studied. Some examples include:

- Cutblocks harvested 10–15 years ago under specific management regimes and located within certain districts and biogeoclimatic zones;
- S4 stream reaches in areas used by cattle within a particular district; or
- User maintained Forest Service recreation sites that can be accessed by vehicle.

These three examples begin to describe the populations under study for three different evaluations. They also describe the basic units that are to be measured and studied. The units within these three populations[1] are quite different – cutblocks, S4 stream reaches and recreation sites.

In all three of these examples, the number of units within each population is quite large. It is easy to see that studying and measuring all of the units is often not practical. Thus, a subset of the units will need to be selected for measurement.[2] This extension note will discuss why the units we assess should be selected in a random manner.

Homogeneous Population?

If all of the study units under consideration were very uniform, then it would not matter which units we chose to evaluate or how we chose them. As Cochran (1963) states: "Laboratory diagnoses about the state of our health are made from a few drops of blood. This procedure is based on the assumption that the circulating blood is always well mixed and that one drop tells the same story as another... But when the material is far from uniform, as is often the case, the method by which the sample is obtained is critical, and the study of techniques that ensure a trustworthy sample becomes important."

1 As effectiveness evaluation teams work on defining their particular populations, these examples would become more detailed. Also note that unit definition is not always as straightforward as these examples suggest.

2 It is 'well-known' within sampling circles that samples often give a more accurate picture of the population than would an attempted census of all the units. Two reasons for this are: 1) actually sampling ALL of the units is often not possible, and 2) the measuring and recording process for all the units is often so onerous that the accuracy and quality control of the data collection is significantly impacted.

The FRPA Evaluator is a regular publication of the FRPA Resource Evaluation Program designed to inform stakeholders on program development and implementation, and report on the results of evaluation projects.

The objective of the FRPA Resource Evaluation Program is to determine if forest and range policies and practices in British Columbia are achieving government's objectives for the resource values identified in FRPA, with a priority on environmental outcomes and consideration for social and economic parameters, where appropriate.

Sample as a Mirror of the Population

In forestry, most of the populations we study are not homogeneous, so we would like to draw a sample of units that represents the entire population. As Lohr (1999) states: "A perfect sample would be ... a scaled-down version of the population, mirroring every characteristic of the whole population." It is particularly important that the characteristics that we are most interested in be mirrored as well as possible within our sample.

Clearly, the more units we select to include in our sample, the better chance the sample has to mirror the whole population. A sample of one unit would provide some idea of the mean value of a chosen indicator, but is unable to provide any estimate of the variability in the population around that indicator's mean value. A sample of 10 units not only provides a better estimate of the mean, but also provides an estimate of the variability around that indicator's mean value.[3]

Desirable Qualities of our Sampling Methodology

As Cochran stated, we need a technique or methodology for selecting a sample to ensure that we obtain one that is trustworthy. What properties of the samples obtained from a sampling methodology will allow them to be trustworthy and effectively 'mirror' a population? Stuart (1984) states: "We are all accustomed to the idea of sampling in everyday life. The housewife visually samples the quality of the fruit and vegetables she intends to buy... If the greengrocer puts the best on display and sells us inferior qualities, we protest at the biased sample or change our supplier... Even if bias of this kind is excluded, we know that some greengrocers are more reliable than others in the consistency of the quality they offer... These qualities, of freedom from bias and of reliability, are what we require from samples in general, and from sample surveys in particular..."

In Stuart's quote, bias and reliability are defined in terms of repeated experience with greengrocers. After several experiences with the first greengrocer, the consumer realizes that he or she consistently provides poorer quality food than what is on display. This is an example of bias. On the other hand, the second greengrocer consistently provides a similar quality of food time and time again. This is a reliable greengrocer since there is little variability in his or her produce. But clearly, if the second greengrocer's produce was poor, even though reliably so, we wouldn't continue to buy from him or her. We would prefer a greengrocer who does both – reliably provides us with produce of high quality.

Another way of looking at this is to consider target practice. With each shot or sample we are trying to reach the target centre or bull's eye. The target centre represents the true value of the indicator we are investigating, while the shooter represents the method we use to select a sample of study units. A biased shooter or sampling methodology will tend to hit off target in one direction, for instance, maybe most of the shots will be above the target. A reliable shooter will be precise and all of his or her shots will hit the target close together. If this reliable shooter is also unbiased then the cloud of shots will be centred on the target centre.

Thus, we want a sampling methodology that would repeatedly select samples[4] that are:

- **Unbiased** – the samples do not consistently produce lower or higher estimates from the population value for the indicator we are measuring.
- **Reliable or Precise** – sample estimates of the population value for the indicator we are measuring do not vary greatly.
- **Feasible** – the samples are practical and feasible to collect at a reasonable cost.

3 The variability around the mean value is usually estimated by the standard error which, in turn, is used to construct confidence limits around the indicator's mean value. In general, increasing the sample size decreases the standard error, but there are diminishing returns if the sample size gets too large – rarely a problem in forestry.

4 In most cases we can't practically test a sampling methodology by repeated sampling, so we must use sampling theory instead to determine an appropriate sample design for our specific situation.

While the precision or confidence limits of our sample can generally be reduced by simply increasing the sample size, any potential bias can only be controlled by understanding the subject matter under investigation and designing a sampling plan appropriately. Of course, a well-designed plan might also reduce the confidence limits without increasing the sample size.

Sampling Frame

Leaving aside the question of how many units to sample, we next ask: how should we select the units to be included in our sample? That is, what will our sampling plan be? To begin to answer this question we need basic information about the units within the population: how many are there and where can we find them? In other words, we need a 'frame' or list of all the units in the population along with information about how to find those units (their addresses if you will).

As you can imagine, obtaining a good sampling frame may be quite a challenge. After all, the frame should include every unit in the population while not containing any units that don't belong. Clearly, if there are significant problems with the frame then the units selected for the sample may provide biased results. This would occur if some units in the population were not listed in the frame, and so did not have a chance to be in the sample, while other units, not in the population, were inappropriately included in the frame, thus having a chance to be in the sample.

Seed Orchard Example

Let us also leave this issue aside and work with an example where the frame for a population is at least conceptually easy to obtain: a seed orchard. Suppose that our units are the trees within the seed orchard and that there is an accurate and complete map with the location of the 1000 trees arranged in 50 rows of 20 trees each. The indicator variable under investigation is an attribute of the whole tree, such as its height, health or size of its cone crop.

Convenience Sample

Suppose that we have already determined that a sample size of 10 trees[5] will be sufficient for the purposes of our study. How will we select these 10 trees? We might be tempted to select a *convenience* sample.[6] For instance, suppose that there is a road along one side of the orchard just across from the office. It would be convenient then to simply cross the road and pick the closest 10 trees. What could be wrong with this?

There are a number of possibilities. Maybe these same trees get sampled all the time, and they have already lost several branches to destructive sampling; if it is a dirt road maybe their needles are covered with dust so that they are unable to grow as well as trees deeper inside the seed orchard; maybe trees within rows were each cloned from the same genetic stock so that the selected 10 trees all belong to one clone. There are many other possibilities (I'm sure that you can think of some!) including ones that we may not be able to think of. But in the cases given, we can see that these 10 trees may be different in some important way from the other trees in the seed orchard. So, how can we feel confident that what we observe for them would also hold for the other trees? Instead we have a possible case of **selection bias** – where we have or may have biased the results because of the way in which we chose the sample trees. Unless the population we are sampling is homogeneous, a sample of convenience is likely to be biased.

If a convenience sample is used as if it were a random sample "the fact that this assumption is being made needs to be recognized, and the validity of making statistical inferences as if a random sample had been selected needs to be critically assessed…" (Hahn and Meeker, 1993). Further, any report of the results should include a discussion of this issue.

5 While we are not discussing sample sizes here, it is important to know that when considering survey precision, so long as the sample size is a small proportion of the population size (say, less than 5%), the number of units in the sample is more important than the proportion of the population sampled.

6 This is especially true in situations where the frame is hard to define. A good example of the pitfalls of convenience sampling is described in Lohr (1993), pages 7 and 8. This same example is described differently in Schwarz (2004), chapter 4, pages 12–14.

Judgement Sample

A slightly less convenient sample in the case of the seed orchard might be to select 10 'representative' trees based on the judgement of one of the team members. Maybe this team member has worked closely with all of these trees for years and feels that he or she is qualified to pick a 'representative' sample. This raises a number of interesting questions. Does the person have to know the whole population well in order to pick representative units? Will this person be aware of all the variables that could impact the indicator variable under study so that he or she can choose trees in such a way that none of these variables would bias the results from the units selected? Would someone else with the same qualifications pick a 'similar' sample (whatever similar might mean in this circumstance)? Would other people accept this person's judgement? As Deming (1975) states: "Use of a judgement-sample of material and environmental conditions ... is worth no more than the reputation of the [person who selected those units]."

Another problem with judgement samples is the tendency to pick units that have 'average' values of the indicator variables. If this happens, then the selected units will not vary much from each other and it will not be possible to properly estimate the variability within the population as a whole. Confidence limits calculated for the observed mean would be smaller than that of the population.

In most cases, it is best to avoid judgement samples. It is simply too easy for unconscious biases to creep in so that, for instance, trees that are slightly larger or healthier might be selected. This can happen with even the best intention and diligence of the person selecting the units. Furthermore, the choice of sample may be difficult to defend when challenged by others, as quite reasonably, many people will not accept the argument 'trust me.'

Random Sample

Instead, it is better that each tree in the seed orchard have a known and positive chance of being included in the sample. If these chances are all equal then every tree in the orchard, whether it is alongside the dusty road or surrounded by clean air in the middle of the orchard, would have an equal opportunity to be included in the sample. If there are more trees in the middle of the orchard than alongside the dusty road, then the sample is more likely to contain some of those trees than the dusty ones.

If there are systematic differences between units within the population, these can either be ignored or included in the sampling design as strata. Stratification is one technique that can be used to increase the precision of the final estimates without necessarily increasing the sample size. For instance, if differences between clones in the seed orchard are believed to strongly impact the indicator variable, then random samples within each clone might be an appropriate design (although this may be difficult with our example sample of 10 trees and 50 different clones).

Arkin (1984) lists a number of advantages of using a random sample. These include:

- The sample result is objective and defensible.
- The sampling error can be estimated so that confidence limits can be calculated.
- The method usually provides a more accurate and cheaper way of drawing conclusions about the population than could a full census.
- Determination of the required sample size can be done on an objective basis.

Systematic Sample

Sometimes, a random sample is difficult to undertake. Finding randomly located units may involve substantial travel time and work. It may be hard to identify which units are the ones that were randomly selected. If the same units need to be measured again, relocation of the units may be quite difficult. In this case, it may be as important to correctly find and relocate the chosen study units as to have a fully random sample. Or, it may seem that the units to be studied are so homogeneous and/or randomly distributed already, that a truly random sample doesn't seem necessary. Under these circumstances, it may be tempting to take a systematic sample instead.

While random sampling the example seed orchard does not seem particularly arduous, one could take a systematic sample. For instance, we might take the 8th tree in every 5th row starting at the 2nd row so that a tree is chosen from 10 different rows: the 2nd, 7th, 12th, 17th, 22nd, 27th, … rows. This allows us to select our trees from across the whole seed orchard and to do so 'objectively' in the sense that the rule we use is very clear and not open to interpretation (i.e., subjective bias). We should add a random twist to this systematic sample by randomly choosing which of the first 5 rows to start with and which of the 20 trees along each row will be sampled.

If we do take a systematic sample in this way, we redefine the way we sample the orchard (our sample design). The sample is no longer chosen from every possible group of 10 trees,[7] but from only those groups of trees that are 5 rows apart and at the same position in the row. Thus, there are only 5 rows × 20 trees/per row = 100 possible samples for us to choose from. Millions of possible 10-tree samples now have a zero probability of being selected. Effectively we are now sampling groups of trees[8] instead of individual trees. Since we selected one group of trees from a possible 100 groups, we have an effective sample size of one even though we measure the attributes of 10 trees.

We could increase our sample size by taking two systematic samples of 5 each. Each sample could start from a row randomly chosen from between 1 and 10 and then a randomly chosen tree location for each 10th row. For example, we might start with the 2nd row and take every 8th tree in the 2nd, 12th, 22nd, 32nd, and 42nd rows. The second systematic sample might start with the 3rd row and take the 1st tree in the 3rd, 13th, 23rd, 33rd, and 43rd rows. We are now taking two samples from a possible 10 rows × 20 trees per row = 200 samples.

What are some of the pitfalls of using a systematic sample? Well, it is particularly important that the pattern we use does not match some pattern in the population as a whole. Since we may not always know what those patterns might be, we must think carefully about this. Suppose that the seed orchard had been systematically pruned 5 different ways so that every 5th row had received the same pruning treatment. Or maybe there is a little drainage ditch that just happens to travel by all the 8th trees in the orchard so that those trees get a little more moisture than the others. Then the systematic examples given above would pick trees that are different in a systematic way from the other trees in the orchard, and the obtained results would not likely represent the seed orchard very well.

In general, it is best to use a random sample. If, in specific cases, a simple random sample seems too expensive or onerous to implement, there are many possible sampling designs, as well as the systematic design, that may help reduce the cost of sampling while maintaining the statistical properties of a random sample. Discussion of these designs is beyond the scope of this document, but more information can be found in the references or by contacting a statistician.

Sampling Theory

Traditional sampling theory produces sampling methods and designs built on the assumption that each unit in the population will have a known and non-zero probability[9] of being included in the sample. The sampling distributions and calculations of means, standard errors and confidence limits can only be correct if this assumption is reasonable. Thus, while a convenient or judgement sample may be easier to take, there is no available statistical theory to justify the extension of the resulting estimates to the population as a whole. Any such extensions must be argued on scientific or subject matter grounds, and others in the field may be unwilling to accept these arguments. If well designed, systematic samples can be treated as a form of cluster sampling so that means and confidence limits can be calculated.

7 Of which there are 1000!/(10!990!) = 26.3×10^{21} possibilities!

8 These groups are known as 'clusters' in statistical literature. Thus, means and standard errors of a systematic sample could be calculated using cluster sampling methods. If it can be assumed that the systematic sample is as good as a simple random sample because the population is itself randomly distributed, then the data can be analysed as if it were a simple random sample. Of course, this should be clearly discussed in any report of the results.

Summary

Sampling designs with random selection of units are designed to reduce selection bias so that we can get unbiased estimates of the mean and variability for the indicators we are studying or assessing. Causes of selection bias include:

- Problems with the frame when the list of units and their 'addresses' don't completely match those in the population.
- Choosing units that are simply convenient or easy to find.
- Choosing units based on judgement as to which units are 'representative' of the population as a whole.
- Choosing a systematic sample that matches some underlying pattern in the population.
- Measurement error. For instance, in a natural regeneration study, it may be required that seedlings of any size be counted. But the smaller they are, the less likely the observer will find and count them. A minimum height requirement of 10 cm might reduce this error, but it also results in a re-definition of both the indicator and the population under study.
- Accessibility problems when some units are simply too difficult or expensive to sample. For instance, some cut-blocks in the first example at the beginning of this document may have been heli-logged or their access roads permanently deactivated so that access to that subset of cutblocks may be very difficult or expensive for data collectors. Thus, it may be decided to not sample them even though they appeared in the list of randomly selected units. This raises the question as to whether there are systematic differences between those cutblocks which are accessible versus those that are not, and how such differences might affect the results of the study. This question should be seriously examined and discussed in any reports.

Sampling plans can be designed to be practical, while minimizing the variability around the estimated means and the cost of implementing the sample design. Of course, these goals often involve trade-offs since reducing the variability of estimates usually requires more samples, which increases costs. On the other hand, there are many sampling designs available that may reduce the need for significant trade-offs in certain circumstances.

References

Arkin, Herbert. 1984. Handbook of sampling for auditing and accounting. 3rd ed., McGraw-Hill Book Co., New York, NY.

Cochran, William G., 1963. Sampling techniques, 2nd ed., John Wiley & Sons, Inc., New York, NY.

Deming, W. Edwards. 1975. On probability as a basis for action. The American Statistician, 29:146–152.

Hahn, Gerald J. and William Q. Meeker. 1993. Assumptions for statistical inference. The American Statistician, 47:1–11.

*Lohr, Sharon L., 1999, Sampling: Design and analysis, Duxbury Press, Toronto, Ont.

*Schwarz, Carl J. 2004. Course Notes for Stat-650 – Quantitative Methods for Resource Managers and Field Biologists, online at http://www.stat.sfu.ca/~cschwarz/Stat-650/.

*Stuart, Alan., 1984, The ideas of sampling. 3rd ed., Charles Griffin & Co., Ltd, London, Eng.

Yates, Frank, 1953, Sampling methods for censuses and surveys, 2nd ed., Charles Griffin & Co. Ltd, London, Eng.

* These are good sources for developing a greater understanding of sampling methodology.

For More Information

For more information on the FRPA Resource Evaluation Program, please visit our web site at http://www.for.gov.bc.ca/hfp/frep.

9 It is not necessary that each unit be selected with equal probability. However, it is important that its probability of selection is known, so that calculations of means and standard errors can be done correctly.

Addendum:
Using EXCEL to obtain simple random samples

Prepared by Wendy Bergerud with assistance from Peter Ott and Peter Fielder, Research Branch, BC Ministry of Forests

An important step in any sampling program is to randomly select study units from the population to be sampled. A **simple random sample (SRS)** can be obtained from an EXCEL spreadsheet if each row contains information about one study unit in the population. This involves finding a way to randomly order the rows in the worksheet. The easiest way that I have found to do this is as follows:

1. **Add a column with the function RAND** in each cell. This generates a random number in each cell. [You do this by typing = RAND() into the formula bar and pressing enter.[11] Then highlight that cell and drag it down the column for all the rows for which random numbers are desired.]
2. You may notice that when any cell of the worksheet is changed, all of the numbers in the random number column change. The best way around this constant change is to highlight the column of random numbers and **copy** it. Then use a **paste special** to put the numbers into another column or on top of the original column. In the paste special window, choose to paste **only the values** and not the formula.[12]
3. Now we can **sort the whole worksheet using the column of fixed random numbers as the first sorting field**. [You do this by highlighting the whole sheet by clicking on the empty topmost left cell; clicking on the Data drop-down field; choosing the sort option; and then choosing the columns to sort by.]

You now have the rows in random order. If the sample size is to be 10, then the top 10 rows are the sample. If the sample size is to be 50 then the first 50 rows identify those study units to be measured for the sample.

There are a few additional details you can add to your worksheet:

4. Optionally, you can **create a numbered list** of 1, 2, … in another column. The numbers in this new column are easier to read than the random numbers. [You can create this column by typing 1 in the first cell, 2 in the second cell, then highlight these two cells and drag the corner to the end. EXCEL will automatically generate a numbered list.]
5. You can use conditional formatting to bold and/or add colour to those rows that will be sampled. This makes them easier to identify if you change the sort of the worksheet. [You can do this by highlighting the column containing the numbered list, then click on the format drop-down menu and choose conditional formatting, then choose a condition and a format that you want for cells satisfying that condition.]

10 **The random number generator in EXCEL 97 is not good enough if you are obtaining more than a few numbers.** See http://support.microsoft.com/default.aspx?kbid=828795&product=xl2003#appliesto. Random number tables are another possibility, as well as getting someone who uses the SAS programming language to produce some random numbers. There are also some websites that will produce random numbers that may or may not be acceptably random.

11 Another option is to use the function bar. You do this by clicking on the = sign next to the formula window, then clicking on the down arrow, then clicking on *more functions*, then choosing the category of math & trig, and finally scrolling down the list until you find rand. You can get to the math & trig list more quickly by clicking on the insert drop-down menu and choosing function.

12 Another option is to push F9 after typing in = RAND() into the formula window. Unfortunately, this must be done for each cell while the cursor is in the formula window, and so is only feasible for a small number of cells.

Stratified Random Sample

Suppose that you want to randomly sample study units within each of several strata. If you use the simple random sampling approach described above, the number of study units selected from each stratum will itself be a random variable, and you cannot be sure that each stratum will be sampled. Conducting a simple random sample within each of several strata requires only minor modifications from the methods described above assuming, of course, that each row of the worksheet represents the identifying information for each study unit in the population. This would be done by changing step 3 in the above procedure. The steps are:

1. Add a column with **= RAND()** in each cell in order to generate a random number in each cell.
2. Use a **paste special** to put the numbers on top of the original column as **values only**, thus eliminating the formula.
3. Now we can **sort** the whole worksheet using the column(s) identifying the strata as the first sorting field(s) and the column of random numbers as the next sorting field. For a sample size of 10 per stratum, the first 10 study units within each stratum form the random sample for that stratum. (This might be easier to count if you also add a column of ordered numbers within each stratum.)

Note: A different approach would use the Random Number Generation Tool in the Analysis Toolkit. It is found on the Tools drop-down menu under Data Analysis. However, this tool is not a default option and may have to be installed before it can be used.

Additional statistical information is available at: http://www.for.gov.bc.ca/hre/topics/biomet.htm.

BRITISH COLUMBIA

Ministry of Forests
Ministry of Water, Land and Air Protection
Ministry of Sustainable Resource Management