CHAPTER 2. SAMPLING DESIGN

2.1. INTRODUCTION

This chapter discusses recommended methods for designing sampling programs to track and evaluate the implementation of nonpoint source control measures. This chapter does not address whether the management measures (MMs) or best management practices (BMPs) are effective since no water quality sampling is done. Because of the variation in forestry practices and related nonpoint source control measures implemented throughout the United States, the approaches taken by various states to track and evaluate nonpoint source control measure implementation will differ. Nevertheless, all approaches should be based on sound statistical methods for selecting sampling strategies, computing sample sizes, and evaluating data. EPA recommends that states consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

As described in Chapter 1, implementation monitoring is the focus of this guidance. Effectiveness monitoring is the focus of another guidance prepared by EPA, the *Nonpoint Source Monitoring and Evaluation Guide* (USEPA, 1996). Dissmeyer (1994) also provides substantial information regarding QA/QC, statistical considerations, BMP effectiveness monitoring, and monitoring methods. The recommendations and examples in this chapter address two primary monitoring goals:

- Determine the extent to which MMs and BMPs are implemented in accordance with design standards and specifications.
- Determine whether there has been a change in the extent to which MMs and BMPs are being implemented.

For example, State forestry agencies might be interested in whether streamside management areas (SMAs) at harvest sites associated with all types of forest ownerships (industrial, private nonindustrial, federal, and state) are in compliance with design standards. State forestry agencies might also be interested in whether the percentage of owners of nonindustrial private forest land that are correctly implementing the BMPs specified in a voluntary implementation program.

2.1.1. Study Objectives

To develop a study design, clear, quantitative monitoring objectives must be developed. For example, the objective might be to estimate to within ±5 percent the percent of harvest sites that have adequate SMAs. Or perhaps a state is getting ready to implement new administrative procedures to ensure that purchasers of timber have been advised of needed work. In this case, detecting a 10 percent change in the number of operators that implement the work specified in the timber sale administration file might be of interest. In the first example, summary statistics are developed to describe the current status, whereas in the second example, some sort of statistical analysis (hypothesis testing) is performed to determine whether a significant change has really occurred. This choice has an impact on how the data are collected. As an example, balanced designs (e.g., two sets of data with the same number of observations in each set) are more typical for hypothesis testing, whereas summary statistics might require unbalanced sample allocations to account for variability such as site size, type, and ownership.

2.1.2. Probabilistic Sampling

Most study designs that are appropriate for tracking and evaluating implementation are based on a probabilistic approach since tracking every operator is not cost-effective. In a probabilistic approach, individuals are randomly selected from the entire group. The selected individuals are evaluated, and the results provide an unbiased assessment of the entire group. Applying the results from randomly selected individuals to the entire group is *statistical inference*. Statistical inference enables one to determine, for example, the probable percentage of timber sales with adequate SMAs without visiting every tract of land. One could also determine whether the change in timber sales with appropriate streamside management is within the range of what could occur by chance or the change is large enough to indicate a real modification of operator practices.

The group about which inferences are made is the population or *target population,* which consists of *population units*. The *sample population* is the set of population units that are directly available for measurement. For example, if the objective is to determine the degree to which adequate SMAs have been established, silvicultural operations for which SMAs are an appropriate BMP (e.g., timber sales with nearby streams) would be the sample population. Statistical inferences can be made only about the target population available for sampling. For example, if implementation of erosion control is being assessed and only public lands can be sampled, inferences cannot be made about the management of private lands.

The most common types of probabilistic sampling that can be used for implementation monitoring are summarized in Table 2-1. In general, probabilistic approaches are preferred. However, there might be circumstances under which targeted sampling should be used. Targeted sampling refers to using best professional judgement for selecting sample locations. For example, state foresters deciding to evaluate all timber sales in a given watershed would be targeted sampling. The choice of a sampling plan depends on study objectives, patterns of variability in the target population, cost-effectiveness of alternative plans, types of measurements to be made, and convenience (Gilbert, 1987).

Simple random sampling is the most basic type of sampling. Each unit of the target population has an equal chance of being selected. This type of sampling is appropriate when there are no major trends, cycles, or patterns in the target population (Cochran, 1977). Random sampling can be applied in a variety of ways including operator or timber sale selection. Random samples can also be taken at different times at a single site. Figure 2-1 provides an example of simple random sampling from a listing of harvest sites and from a map.

Table 2-1. Applications of four sampling designs for implementation monitoring.

Sampling Design	Comment
Simple Random Sampling	Each population unit has an equal probability of being selected.
Stratified Random Sampling	Useful when a sample population can be broken down into groups, or strata, that are internally more homogeneous than the entire sample population. Random samples are taken from each stratum although the probability of being selected might vary from stratum to stratum depending on cost and variability.
Cluster Sampling	Useful when there are a number of methods for defining population units and when individual units are clumped together. In this case, clusters are randomly selected and every unit in the cluster is measured.
Systematic Sampling	This sampling has a random starting point with each subsequent observation a fixed interval (space or time) from the previous observation.

If the pattern of MM and BMP implementation is expected to be uniform across the state, simple random sampling is appropriate to estimate the extent of implementation. If, however, implementation is homogeneous only within certain categories (e.g., federal, state, or private lands), stratified random sampling should be used.

In *stratified random sampling*, the target population is divided into groups called strata for the purpose of obtaining a better estimate of the mean or total for the entire population. Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables to group observations into more units, thereby reducing the variability of observations within each unit. For example, in a state with federal, state, and private forests, there might be different patterns of BMP implementation. Lands in the state could be divided into federal, state, and private as separate strata from which samples would be taken. In general, a larger number of samples should be taken in a stratum if the stratum is more variable, larger, or less costly to sample than other strata. For example, if BMP implementation is more variable on private lands, a greater number of sampling sites might be needed in that stratum to increase the precision of the overall estimate. Cochran (1977) found that stratified random sampling provides a better estimate of the mean for a population with a trend, followed in order by systematic sampling (discussed later) and simple random sampling. He also noted that stratification typically results in a smaller variance for the estimated mean or total than that which results from comparable simple random sampling.

Harvest Site No.	Water Type	Ownership	County Code
1	Stream	Industry	14
2	Stream	Private Non-industrial	3
3	Lake/Pond	Industry	12
4	Stream	Industry	11
5	Lake/Pond	Private Non-industrial	7
6	Lake/Pond	Private Non-industrial	4
7	Lake/Pond	Industry	7
8	Stream	State	10
9	Stream	Federal	10
10	Lake/Pond	Private Non-industrial	5
11	Stream	Industry	3
12	Stream	State	6
•••	•••	•••	•••
142	Lake/Pond	Industry	11
143	Lake/Pond	State	7
144	Stream	State	15
145	Lake/Pond	Industry	15
146	Stream	Private Non-industrial	8
147	Stream	Private Non-industrial	9
148	Lake/Pond	Industry	1
149	Stream	Industry	12
150	Lake/Pond	Federal	11

Figure 2-1a. Simple random sampling from a listing of harvest sites. In this listing, all harvest sites are presented as a single list and sites are selected randomly from the entire list. Shaded harvest sites represent those selected for sampling.

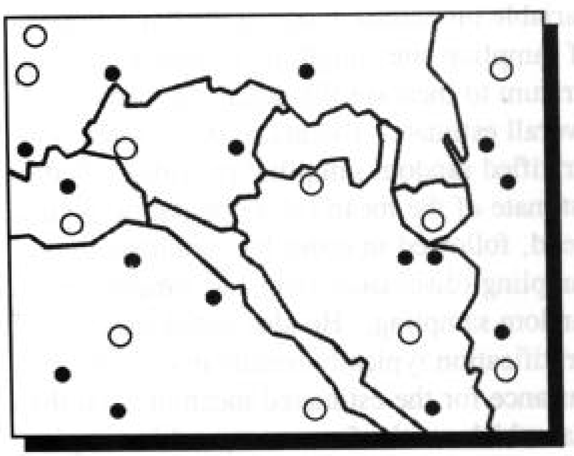

Figure 2-1b. Simple random sampling from a map. Dots represent harvest sites. All harvest sites of interest are represented on the map, and the sites to be sampled (open dots—○) were selected randomly from all harvest sites on the map. The shaded lines on the map could represent county, watershed, hydrologic, or some other boundary, but they are ignored for the purposes of simple random sampling.

If the state believes that there will be a difference between two or more subsets of the sites, such as between types of ownership or region, the sites can first be stratified into these subsets and a random sample taken within each subset (McNew, 1990). States with silviculture implementation monitoring programs commonly divide the sites by ownership and county and/or region before selecting survey sites. The goal of stratification is to increase the accuracy of the estimated mean values over what could have been obtained using simple random sampling of the entire population. The method makes use of prior information to divide the target population into subgroups that are internally homogeneous. There are a number of ways to "select" sites, or sets of sites, to be certain that important information will not be lost, or that MM or BMP use will not be misrepresented as a result of treating all potential survey sites as equal. Figure 2-2 provides an example of stratified random sampling from a listing of harvest sites and from a map.

Where data are available, it might be useful to compare the relative percentages of harvested timberland that is classified as having high, medium, and low erosion potentials. In cases where sediment is impacting water quality, highly erodible land might be responsible for a larger share of sediment delivery and would therefore be an important target for tracking the implementation of erosion controls. A stratified random sampling procedure could be used to estimate the percentage of total harvested timberland with different erosion potentials that have erosion controls in place. For other water quality problems (e.g., spawning habitat in decline), other stratification parameters (e.g., stream classification) might be more appropriate.

Cluster sampling is applied in cases where it is more practical to measure randomly selected groups of individual units than to measure randomly selected individual units (Gilbert, 1987). In cluster sampling, the total population is divided into a number of relatively small subdivisions, or clusters, and then some of the subdivisions are randomly selected for sampling. In one-stage cluster sampling, the selected clusters are sampled totally. In two-stage cluster sampling, random sampling is performed within each cluster (Gaugush, 1987). For example, this approach might be useful if a state wants to estimate the proportion of harvest sites that are following state-approved MMs or BMPs. All harvest sites in a particular county can be regarded as a single cluster. Once all clusters have been identified, specific clusters can be randomly chosen for sampling. Freund (1973) notes that estimates based on cluster sampling are generally not as good as those based on simple random samples, but they are more cost-effective. As a result, Gaugush (1987) believes that the difficulty associated with analyzing cluster samples is compensated for by the reduced sampling requirements and cost. Figure 2-3 provides an example of cluster sampling from a listing of harvest sites and from a map.

Systematic sampling is used extensively in water quality monitoring programs because it is relatively easy to do from a management perspective. In systematic sampling the first sample has a random starting point and each

Harvest Site No.	Water Type	Ownership	County Code
1	Stream	Industry	14
3	Lake/Pond	Industry	12
4	Stream	Industry	11
...	...	...	...
148	Lake/Pond	Industry	1
149	Stream	Industry	12
2	Stream	Private Non-industrial	3
5	Lake/Pond	Private Non-industrial	7
6	Lake/Pond	Private Non-industrial	4
...	...	...	...
146	Stream	Private Non-industrial	8
147	Stream	Private Non-industrial	9
9	Stream	Federal	10
21	Stream	Federal	3
...	...	...	...
150	Lake/Pond	Federal	11
8	Stream	State	10
12	Stream	State	6
...	...	...	...
143	Lake/Pond	State	7
144	Stream	State	15

Figure 2-2a. Stratified random sampling from a listing of harvest sites. Within the listing, harvest sites were subdivided by the type of ownership. Then, considering only the private ownership category, harvest sites were selected randomly from within the category. The process of random site selection was then repeated separately for each ownership category. Shaded harvest sites within each category represent those selected for sampling.

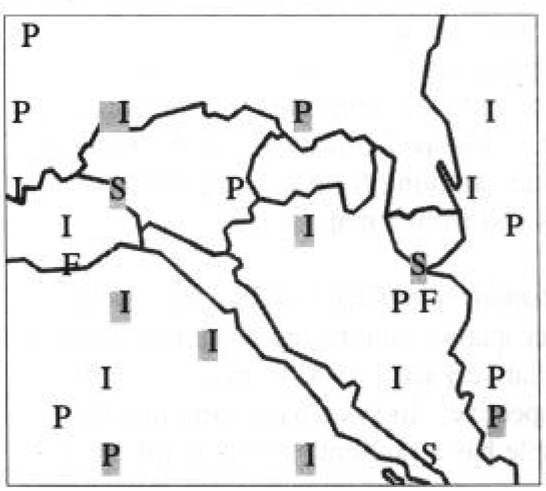

Figure 2-2b. Stratified random sampling from a map. Letters represent harvest sites, subdivided by type of ownership (P = private nonindustrial, I = industrial, F = federal, S = state). All harvest sites of interest are represented on the map. From all of the sites in one ownership category, sites were randomly selected for sampling (highlighted sites). The process was repeated for each ownership category. The shaded lines on the map could represent county, soil type, or some other boundary, and could have been used as a means for separating the harvest sites into categories for the sampling process.

Harvest Site No.	Water Type	Ownership	County Code
148	Lake/Pond	Industry	1
2	Stream	Private Non-industrial	3
11	Stream	Industry	3
6	Lake/Pond	Private Non-industrial	4
10	Lake/Pond	Private Non-industrial	5
12	Stream	State	6
5	Lake/Pond	Private Non-industrial	7
7	Lake/Pond	Industry	7
143	Lake/Pond	State	7
146	Stream	Private Non-industrial	8
147	Stream	Private Non-industrial	9
8	Stream	State	10
9	Stream	Federal	10
4	Stream	Industry	11
142	Lake/Pond	Industry	11
150	Lake/Pond	Federal	11
3	Lake/Pond	Industry	12
149	Stream	Industry	12
1	Stream	Industry	14
144	Stream	State	15
145	Lake/Pond	Industry	15

Figure 2-3a. One-stage cluster sampling from a listing of harvest sites. Within the listing, harvest sties were subdivided by the county in which they were located. Some of these counties were then randomly selected, and all harvest sites within the selected counties were chosen for sampling. Shaded harvest sites represent those located in the counties selected (i.e., counties 3, 5, 8, 9, 11, and 15).

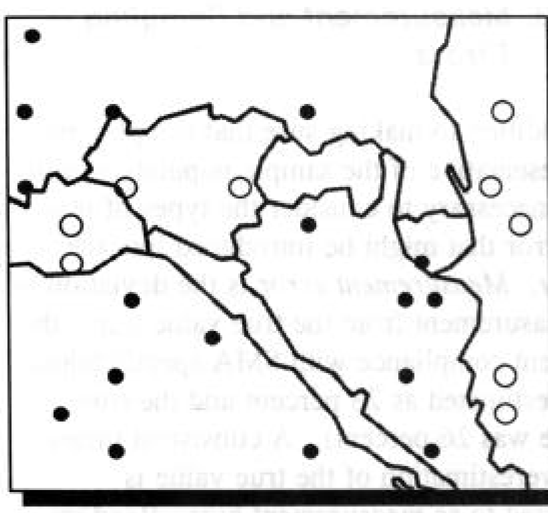

Figure 2-3b. Cluster sampling from a map. All harvest sites in the area of interest are represented on the map (closed {•} and open {O} dots). The shaded lines on the map represent county boundaries. Some of the counties were randomly selected, and all harvest sites within those counties (open dots - O) were selected for sampling. Some other type of boundary, such as soil type or watershed, could have been used to separate the harvest sites for the sampling process.

subsequent sample has a constant distance from the previous sample. For example, if a sample size of 70 is desired from a mailing list of 700 operators, the first sample would be randomly selected from among the first 10 people, say the seventh person. Subsequent samples would then be based on the 17th, 27th, ..., 697th person. In comparison, a stratified random sampling approach might be to sort the mailing list by county and then to randomly select operators from each county. Figure 2-4 provides an example of systematic sampling from a listing of harvest sites and from a map.

In general, systematic sampling is superior to stratified random sampling when only one or two samples per stratum are taken for estimating the mean (Cochran, 1977) or when there is a known pattern of management measure implementation. Gilbert (1987) reports that systematic sampling is equivalent to simple random sampling in estimating the mean if the target population has no trends, strata, or correlations among the population units. Cochran (1977) notes that on the average, simple random sampling and systematic sampling have equal variances. However, Cochran (1977) also states that for any single population for which the number of sampling units is small, the variance from systematic sampling is erratic and might be smaller or larger than the variance from simple random sampling.

Gilbert (1987) cautions that any periodic variation in the target population should be known before establishing a systematic sampling program. Sampling intervals equal to or multiples of the target population's cycle of variation might result in biased estimates of the population mean. Systematic sampling can be designed to capitalize on a periodic structure if that structure can be characterized sufficiently (Cochran, 1977). A simple or stratified random sample is recommended, however, in cases where the periodic structure is not well known or whether the randomly selected starting point is likely to have an impact on the results (Cochran, 1977).

Gilbert (1987) notes that assumptions about the population are required in estimating population variance from a single systematic sample of a given size. There are, however, systematic sampling approaches that do support unbiased estimation of population variance. They include multiple systematic sampling, systematic stratified sampling, and two-stage sampling (Gilbert, 1987). In multiple systematic sampling, more than one systematic sample is taken from the target population. Systematic stratified sampling involves the collection of two or more systematic samples within each stratum.

2.1.3. Measurement and Sampling Errors

In addition to making sure that samples are representative of the sample population, it is also necessary to consider the types of bias or error that might be introduced into the study. *Measurement error* is the deviation of a measurement from the true value (e.g., the percent compliance with SMA specifications was estimated as 23 percent and the true value was 26 percent). A consistent under- or overestimation of the true value is referred to as *measurement bias*. Random

Harvest Site No.	Water Type	Ownership	County Code
1	Stream	Industry	14
2	Stream	Private Non-industrial	3
3	Lake/Pond	Industry	12
4	Stream	Industry	11
5	Lake/Pond	Private Non-industrial	7
6	Lake/Pond	Private Non-industrial	4
7	Lake/Pond	Industry	7
8	Stream	State	10
9	Stream	Federal	10
10	Lake/Pond	Private Non-industrial	5
11	Stream	Industry	3
12	Stream	State	6
•••	•••	•••	•••
142	Lake/Pond	Industry	11
143	Lake/Pond	State	7
144	Stream	State	15
145	Lake/Pond	Industry	15
146	Stream	Private Non-industrial	8
147	Stream	Private Non-industrial	9
148	Lake/Pond	Industry	1
149	Stream	Industry	12
150	Lake/Pond	Federal	11

Figure 2-4a. Systematic sampling from a listing of harvest sites. From a listing of all harvest sites of interest, an initial site (Harvest Site No. 2) was chosen randomly from among the first ten sites on the list. Every fifth site listed subsequently was then selected for sampling.

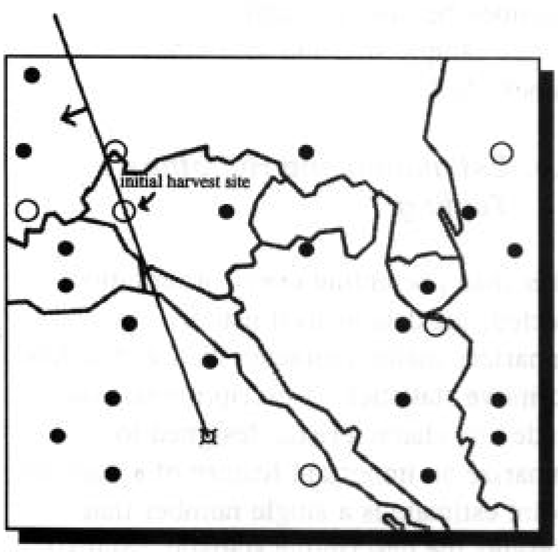

Figure 2-4b. Systematic sampling from a map. Dots (● and ○) represent harvest sites of interest. A single point on the map (¤) and one of the harvest sites were randomly selected. A line was stretched outward from the point to (and beyond) the selected harvest site. The line was then rotated about the map and every fifth dot that it touched was selected for sampling (open dots—○). The direction of rotation was determined prior to selection of the point of the line's origin and the beginning harvest site. The shaded lines on the map could represent county boundaries, soil type, watershed, or some other boundary, but were not used for the sampling. process.

sampling error arises from the variability from one population unit to the next (Gilbert, 1987), explaining why the proportion of operators using a certain BMP differs from one survey to another.

The goal of sampling is to obtain an accurate estimate by reducing the sampling and measurement errors to acceptable levels, while explaining as much of the variability as possible to improve the precision of the estimates (Gaugush, 1987). *Precision* is a measure of how close an agreement there is between individual measurements of the same population. The *accuracy* of a measurement refers to how close the measurement is to the true value. If a study has low bias and high precision, the results will have high accuracy. Figure 2-5 illustrates the relationship between bias, precision, and accuracy.

As suggested earlier, numerous sources of variability should be accounted for in developing a sampling design. Sampling errors are introduced by virtue of the natural variability within any given population of interest. Since sampling errors relate to MM or BMP implementation, the most effective method for reducing such errors is to carefully determine the target population and to stratify the target population to minimize the nonuniformity in each stratum.

Measurement errors can be minimized by ensuring that site inspections are well designed. If data are collected by sending staff out to inspect randomly selected harvest sites, the approach for inspecting the harvest sites should be consistent. For example, how do field personnel determine the percent of adequate SMAs, or what is the basis for determining whether a BMP has been properly implemented?

Reducing sampling errors below a certain point (relative to measurement errors) does not necessarily benefit the resulting analysis because total error is a function of the two types of errors. For example, if measurement errors such as response or interviewing errors are large, there is no point in taking a huge sample to reduce the sampling error of the estimate since the total error will be primarily determined by the measurement error. Measurement error is of particular concern when landowner surveys are used for implementation monitoring. Likewise, reducing measurement errors would not be worthwhile if only a small sample size were available for analysis because there would be a large sampling error (and therefore a large total error) regardless of the size of the measurement error. A proper balance between sampling and measurement errors should be maintained because research accuracy limits effective sample size and vice versa (Blalock, 1979).

2.1.4. Estimation and Hypothesis Testing

Rather than presenting every observation collected, the data analyst usually summarizes major characteristics with a few descriptive statistics. Descriptive statistics include any characteristic designed to summarize an important feature of a data set. A point estimate is a single number that represents the descriptive statistic. Statistics common to implementation monitoring is useful to estimate the *confidence interval*. The confidence interval indicates the range in which the true value lies. For example, if it is

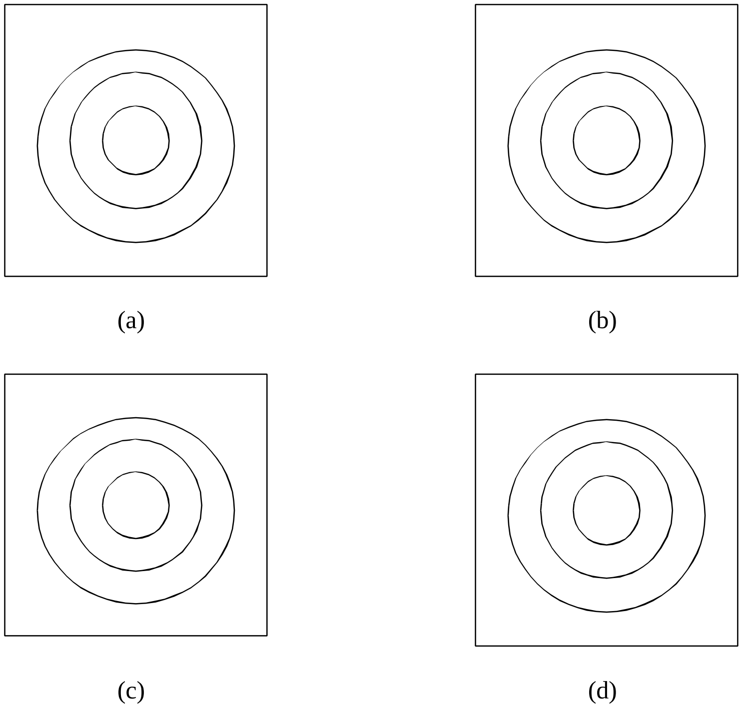

Figure 2-5. Graphical presentation of the relationship between bias, precision, and accuracy (after Gilbert, 1987). (a): high bias + low precision = low accuracy; (b): low bias + low precision = low accuracy; (c): high bias + high precision = low accuracy; and (d): low bias + high precision = high accuracy.

include proportions, means, medians, totals, and others. When estimating parameters of a population, such as the proportion or mean, it estimated that 65 percent of waterbars on skid trails were installed in accordance with design standards and specifications and the 90 percent confidence limit is ±5 percent, there is a 90 percent chance that between 60 and 70 percent of the waterbars were installed correctly.

Hypothesis testing should be used to determine whether the level of MM and BMP implementation has changed over time. The *null hypothesis* (H_o) is the root of hypothesis testing. Traditionally, H_o is a statement of no change, no effect, or no difference; for example, “the proportion of properly installed waterbars after operator training is equal to the proportion of properly installed waterbars before operator training." The *alternative hypothesis (H_a)* is counter to H_o, traditionally being a statement of change, effect, or difference, for example. If H_o is rejected, H_a is accepted. Regardless of the statistical test selected for analyzing the data, the analyst must select the *significance level (α)* of the test. That is, the analyst must determine what error level is acceptable based on the needs of decision makers. There are two types of errors in hypothesis testing:

Type I: H_o is rejected when H_o is really true.

Type II: H_o is accepted when H_o is really false.

Table 2-2 depicts these errors, with the magnitude of Type I errors represented by α and the magnitude of Type II errors represented by β. The probability of making a Type I error is equal to the α of the test and is selected by the data analyst. In most cases, managers or analysts will define *1-α* to be in the range of 0.90 to 0.99 (e.g., a confidence level of 90 to 99 percent), although there have been applications where *1-α* has been set to as low as 0.80. Selecting a 95 percent confidence level implies that the analyst will reject the H_o when H_o is true (i.e., a false positive) 5 percent of the time. The same notion applies to the confidence interval for point estimates described above: α is set to 0.10, and there is a 10 percent chance that the true percentage of properly installed waterbars is outside the 60 to 70 percent range. This implies that if the decisions to be made based on the analysis are major (i.e., affect many people in adverse or costly ways) the confidence level needs to be greater. For less significant decisions (i.e., low cost ramifications) the confidence level can be lower.

Type II error depends on the significance level, sample size, and variability, and which alternative hypothesis is true. *Power (1-β)* is defined as the probability of correctly rejecting H_o when H_o is false. In general, for a fixed sample size, α and β vary inversely. For a fixed α, β can be reduced by increasing the sample size (Remington and Schork, 1970).

Table 2-2. Errors in hypothesis testing.

Decision	**State of Affairs in the Population**	
	H_o is True	**H_o is False**
Accept H_o	1-α (Confidence level)	β (Type II error)
Reject H_o	α (Significance level) (Type I error)	1-β (Power)

2.2. SAMPLING CONSIDERATIONS

In a document of this brevity, it is not possible to address all the issues that face technical staff who are responsible for developing and implementing studies to track and evaluate the implementation of nonpoint source control measures. For example, when is the best time to implement a survey or do on-site visits? In reality, it is difficult to pinpoint a single time of the year. Some BMPs can be checked any time of the year, whereas others have a small window of opportunity. If the goal of the study is to determine the effectiveness of an operator education program, sampling should be timed to ensure that there was sufficient time for outreach activities and for the operators to implement the desired practices. Furthermore, field personnel must have approval to perform a site visit on each tract of land to be sampled. Where access is denied, a randomly selected replacement site is needed.

2.2.1. Site Selection

From a study design perspective, all of these issues must be considered together when determining the sampling strategy. Site selection criteria will differ from state to state depending on the type of forestry practiced in the state, physical landscape, and intended purposes for the information obtained from the implementation monitoring. The following list indicates the typical site selection criteria culled from existing state implementation monitoring programs. (The corresponding state postal code is presented in parentheses.)

- *Site size:* minimum of 5 or 10 acres, depending on the region of the state (MN); minimum 10 acres (SC); minimum 5 acres (MT); minimum 20 acres (ID).

- *Proximity to a stream* (perennial or intermittent): within 300 feet of a stream, or a lake of at least 10 acres surface area (FL); within 200 feet of a stream (MT); sites did not have to be associated with streams or wetlands (SC); within 150 feet of a class II stream (ID).

- *Time of harvest:* within the past 1 year (SC); 1-3 years prior to the audit (MT); within 2 years of harvest (FL).

- *Site preparation:* only sites that had not been site prepared (SC); either slash piled and burned or waiting burning, or slash broadcast and scheduled to be burned (MT).

- *Volume harvested:* at least 7 MBF/ac (MT).

- *Compatibility with previous surveys:* sales had to meet the selection criteria of a previous study for comparability purposes (MT).

Other criteria that might be considered include erosion risk (e.g., more sampling sites could be placed in high-erosion-risk areas than in low-risk erosion areas) and beneficial use (bias sampling toward high use and/or sensitive areas).

2.2.2. Data to Support Site Selection

A list of harvest sites from which to choose those to be surveyed can be created from information obtained from timber harvesters. Depending on the state, the information is often in the form of harvest plans and timber sale contracts. These sources of information normally include:

- U.S. Forest Service offices.
- The state forestry agency, department, or division (for state lands and nonindustrial private).
- Private timber companies (Ehinger and Potts, 1991).

In addition, the Bureau of Land Management (BLM) manages a significant acreage of federal property and may have valuable information (IDDHW, 1993). Aerial photographs of the areas to be surveyed can be used to identify recent harvest sites as well, and this method of identifying the sites tends to remove site selection biases due to the distance of sites from roads or other forms of inconvenience that otherwise might make them less apt to be chosen for a survey.

The data necessary to select sites for BMP tracking will naturally depend on the site selection criteria. For instance, if sites must meet a minimum of board feet harvested, it will be necessary to know harvest volumes in order to select appropriate sites. The amount of data needed will increase as the number of site selection criteria increase, and this should be taken into account when deciding on the criteria, especially given the possibility that some of the data or types of data collected might be unavailable or unreliable.

2.2.3. Example State and Federal Programs

Several states and federal agencies have implemented programs for developing their implementation monitoring programs. This section describes those implemented by Florida, Montana, Idaho, and the USDA Forest Service.

2.2.3.1. Florida

In Florida, the following site selection criteria are used (Vowell and Gilpin, 1994):

- All ownership classes are included.
- Only the northernmost 37 counties are included because most forestry activities occur in these counties.
- Timber harvesting, site preparation, tree planting, or some combination must have occurred within the past 2 years and within 300 feet of an intermittent or perennial stream or a lake 10 acres or larger.
- Each county has a predetermined number of survey sites based on the level of timber removal reported by the Forest Service.
- Sites are selected from fixed-wing aircraft using a random, predetermined flight pattern in each county.

County foresters randomly select qualifying sites along the flight pattern until they have located the number of survey sites assigned to their county.

This approach is a type of stratified random sampling. The entire population (entire state) is first divided into strata containing the northernmost 37 counties based on prior information that indicated most forestry operations occur in those counties. These strata are still too large to conduct random sampling; therefore, the criteria described above are used to reduce the strata to a manageable number given available resources.

2.2.3.2. Montana and Idaho

Montana is interested in certain types of information related to BMP implementation, so they stratify their sample before selecting sites. They follow these steps:

- Information on the sites (e.g., ownership, erosion hazard) is compiled by watershed or basin.
- A list of all sales and information on them is compiled for each basin for the time period of interest (usually within 1-2 years of harvest date).
- Sites that do not meet the selection criteria are eliminated.
- Sites that do meet the selection criteria are ground-truthed.

This is a stratified random approach: Within drainage basins, sites are stratified first by ownership and then by erosion hazard (Ehinger and Potts, 1991; Schultz, 1992).

Idaho also uses this approach, stratifying sites by geographic region and administrative category. This ensures that differences in MM and BMP implementation among different soil, geologic, and administrative groupings are not lost as would be the case if simple random sampling were used (IDDHW, 1993).

2.2.3.3. U.S. Forest Service

The U.S. Forest Service (USDA, 1992) has developed a monitoring system for Region 5 of the Forest Service, Best Management Practice Evaluation Program (BMPEP), with the following objectives:

- Assess the degree of implementation of BMPs.
- Determine which BMPs are effective.
- Determine which BMPs need improvement or development.
- Fulfill Forest Land and Resource Management Plan BMP monitoring commitments.
- Provide a record of performance for management of nonpoint source pollution in Region 5 of the Forest Service.

These objectives are met through three evaluation phases: Administrative, on-site, and in-channel. In general, the first two phases deal with issues related to implementation monitoring, with administrative evaluation primarily addressing programmatic evaluation and on-site evaluation dealing primarily with individual practices. In-channel evaluation consists primarily of effectiveness monitoring.

In the BMPEP, forests are assigned the number and types of evaluations to be completed each year. To support statistical inference, the evaluations assigned to each forest must be performed at randomly identified sites. Sites to be evaluated are identified in two ways: Randomly and by selection ("selected" sites).

Randomly identified sites are essential for making statistical inferences regarding the implementation and effectiveness of BMPs. Random sites are picked from a pool of sites that meet specified criteria.

Selected sites are identified in various ways:

- Identified as part of a monitoring plan prescribed in an environmental assessment, environmental impact study, or land management plan.
- Identified as part of a Settlement of Negotiated Agreement.
- Part of a routine site visit.
- Follow-up evaluations upstream or In-channel Evaluation Sites, to discover sources of problems.
- Sites that are of particular interest to site administrators, specialists, and/or management due to their sensitivity, uniqueness, and other factors.
- Selected for a particular reason specific to local needs.

It is important to note that for statistical inference, the sample pool can only contain the randomly identified sites, not the "selected" sites. Selected sites must be clearly identified and kept separate from the random sites during data storage and analysis. Because on-site evaluation addresses a range of practices, corresponding methods are provided for developing sample pools for randomly selected sites. For example, the sample pool for SMAs should be developed using a Sale Area Map from the Pool of Timber Sales and counting the number of units that have designated SMAs. This constitutes the SMA sample pool.

The data obtained from the sources discussed above may not be precisely what are required by the state conducting the implementation monitoring survey. Ehinger and Potts (1991) report the following difficulties encountered in using data from the National Forest Service database:

- Flawed assumptions concerning the age of a harvest (some were found to be too old to meet the survey criteria).
- Uncertain age of roads.
- Units within 200 feet of stream on paper were in fact farther than 200 feet from a stream.

The following difficulties were associated with private nonindustrial forest units:

- Inadequate database to identify sales meeting the survey criteria.

- Permission to access landowner property not granted.

Landowners who did grant permission were interested primarily in demonstrating their BMP efforts to the state forest department, so this class of ownership was statistically biased.

On private industrial forest sites, a backlog of slash burnings on sale units was found, preventing their use (because of survey criteria), and state forests sites were mostly found to be farther than 200 feet from a stream, making them ineligible for the survey. States must be aware that these kinds of limitations will be encountered.

2.3. SAMPLE SIZE CALCULATIONS

This section describes methods for estimating sample sizes to compute point estimates such as proportions and means, as well as detecting changes with a given significance level. Usually, several assumptions regarding data distribution, variability, and cost must be made to determine the sample size. Some assumptions might result in sample size estimates that are too high or too low. Depending on the sampling cost and cost for not sampling enough data, it must be decided whether to make conservative or "best-value" assumptions. Because the cost of visiting any individual site or group of sites is relatively constant, it is probably cheaper to collect a few extra samples the first time than to realize later that additional data are needed. In most cases, the analyst should probably consider evaluating a range of assumptions regarding the impact of sample size and overall program cost. To maintain document brevity, some terms and definitions used in the remainder of this chapter are summarized in Table 2-3. These terms are consistent with those in most introductory-level statistics texts, and more information can be found there. Those with some statistical training will note that some of these definitions include an additional term referred to as the *finite population correction term (1-N)*, where N is equal to *n/N*. In many applications, the number of population units in the sample population (*N*) is large in comparison to the number of population units sampled (*n*), and (*1*-N) can be ignored. However, depending on the number of units (harvest sites for example) in a particular population, *N* can become quite small. *N* is determined by the definition of the sample population and the corresponding population units. If N is greater than 0.1, the finite population correction factor should not be ignored (Cochran, 1977).

Applying any of the equations described in this section is difficult when no historical data set exists to quantify initial estimates of proportions, standard deviations, means, or coefficients of variation. To estimate these parameters, Cochran (1977) recommends four sources:

- Existing information on the same population or a similar population.

Table 2-3. Definitions used in sample size calculation equations.

N	=	total number of population units in sample population
n	=	number of samples
n_0	=	preliminary estimate of sample size
a	=	number of successes
p	=	proportion of successes
q	=	proportion of failures (1-p)
x_i	=	i^{th} observation of a sample
$\bar{x}$	=	sample mean
s^2	=	sample variance
s	=	sample standard deviation
$N\bar{x}$	=	total amount
μ	=	population mean
σ^2	=	population variance
σ	=	population standard deviation
C_v	=	coefficient of variation
$s^2(\bar{x})$	=	variance of sample mean
ϕ	=	n/N (unless otherwise stated in text)
$s(\bar{x})$	=	standard error (of sample mean)
1-ϕ	=	finite population correction factor
d	=	allowable error
d_r	=	relative error
Z_α	=	value corresponding to cumulative area of 1-α using the normal distribution (see Table A1).
$t_{\alpha,df}$	=	value corresponding to cumulative area of 1-α using the student t distribution with df degrees of freedom (see Table A2).

$$p = a/n \qquad q = 1 - p$$

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

$$s = \sqrt{s^2} \qquad C_v = s/\bar{x}$$

$$d = |\bar{x} - \mu| \qquad d_r = \frac{|\bar{x} - \mu|}{\mu}$$

$$s^2(\bar{x}) = \frac{s^2}{n}(1-\phi) \qquad s(\bar{x}) = \frac{s}{\sqrt{n}}(1-\phi)^{0.5}$$

$$s(N\bar{x}) = \frac{Ns}{\sqrt{n}}(1-\phi)^{0.5} \qquad s(p) = \sqrt{\frac{pq}{n}}(1-\phi)^{0.5}$$

- A two-step sample. Use the first-step sampling results to estimate the needed factors, for best design, of the second step. Use data from both steps to estimate the final precision of the characteristic(s) sampled.

- A "pilot study" on a "convenient" or "meaningful" subsample. Use the results to estimate the needed factors. Here the results of the pilot study generally cannot be used in the calculation of the final precision because often the pilot sample is not representative of the entire population to be sampled.

- Informed judgment, or an educated guess.

It is important to note that this document only addresses estimating sample sizes with traditional parametric procedures. The methods described in this document should be appropriate in most cases, considering the type of data expected. If the data to be sampled are skewed, as with much water quality data, the analyst should plan to transform the data to something symmetric, if not normal, before computing sample sizes (Helsel and Hirsch, 1995). Kupper and Hafner (1989) also note that some of these equations tend to underestimate the necessary sample because power is not taken into consideration. Again, EPA recommends that if the analyst lacks a background in statistics, he/she should consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

2.3.1. Simple Random Sampling

In simple random sampling, it is presumed that the sample population is relatively homogeneous and a difference in sampling costs or variability is not expected. If the cost or variability of any group within the sample population were different, it might be more appropriate to consider a stratified random sampling approach.

> *What sample size is necessary to estimate to within ±5 percent the proportion of harvest sites that have adequate SMAs?*
>
> *What sample size is necessary to estimate the proportion of harvest sites that have adequate SMAs so that the relative error is less than 5 percent?*

To estimate the proportion of harvest sites implementing a certain BMP or MM, such that the allowable error, *d*, meets the study precision requirements (i.e., the true proportion lies between *p-d* and *p+d* with a *1-*″ confidence level), a preliminary estimate of sample size can be computed as (Snedecor and Cochran, 1980)

$$n_o = \frac{(Z_{1-\alpha/2})^2 p\,q}{d^2} \qquad \textbf{(2-1)}$$

If the proportion is expected to be a low number, using a constant allowable error might not be appropriate. Ten percent plus/minus 5 percent has a 50 percent relative error. Alternatively, the relative error, d_r, can be specified (i.e., the true proportion lies between $p-d_r\,p$ and $p+d_r\,p$ with a *1-*″ confidence level) and a preliminary estimate of sample size can be computed as (Snedecor and Cochran, 1980)

$$n_o = \frac{(Z_{1-\alpha/2})^2 q}{d_r^2\, p} \qquad \textbf{(2-2)}$$

In both equations, the analyst must make an initial estimate of *p* before starting the study. In the first equation, a conservative sample size can be computed by assuming *p* equal to 0.5. In the second equation the sample size gets larger as *p* approaches 0 for constant d_r, and thus an informed initial estimate of *p* is needed. Values of ″ typically range from 0.01 to 0.10. The final sample size is then estimated as (Snedecor and Cochran, 1980)

$$n = \begin{cases} \dfrac{n_0}{1+\phi} & \text{for } \phi > 0.1 \\ n_o & \text{otherwise} \end{cases} \qquad \textbf{(2-3)}$$

where *N* is equal to n_o/N. Table 2-4 demonstrates the impact on *n* of selecting *p*, ″, *d*, d_r, and *N*. For example, 278 random

Table 2-4. Comparison of sample size as a function of p, α, d, d_r, and N for estimating proportions using equations 2-1 through 2-3.

Probability of Success, p	Significance level, α	Allowable error, d	Relative error, d_r	Preliminary sample size, n_o	Sample Size, n				
					Number of Population Units in Sample Population, N				
					500	750	1,000	2,000	Large N
0.1	0.05	0.050	0.500	138	108	117	121	138	138
0.1	0.05	0.075	0.750	61	55	61	61	61	61
0.5	0.05	0.050	0.100	384	217	254	278	322	384
0.5	0.05	0.075	0.150	171	127	139	146	171	171
0.1	0.10	0.050	0.500	97	82	86	97	97	97
0.1	0.10	0.075	0.750	43	43	43	43	43	43
0.5	0.10	0.050	0.100	271	176	199	213	238	271
0.5	0.10	0.075	0.150	120	97	104	107	120	120

samples are needed to estimate the proportion of 1,000 harvest sites with adequate SMAs to within ±5 percent (*d*=0.05) with a 95 percent confidence level, assuming roughly one-half of harvest sites have adequate SMAs.

Suppose the goal is to estimate the average acreage per harvest site where erosion controls are used. The number of random samples required to achieve a desired margin of error when estimating the mean (i.e., the true mean lies between $\bar{x}-d$ and $\bar{x}+d$ with a *1-"* confidence level) is (Gilbert, 1987)

> *What sample size is necessary to estimate the average number of acres per harvest site using erosion controls to within ±25 acres?*
>
> *What sample size is necessary to estimate the average number of acres per harvest site using erosion controls to within ±10 percent?*

$$n = \frac{(t_{1-\alpha/2,n-1}s/d)^2}{1 + (t_{1-\alpha/2,n-1}s/d)^2/N} \qquad \textbf{(2-4)}$$

If N is large, the above equation can be simplified to

$$n = (t_{1-\alpha/2,n-1}s/d)^2 \qquad \textbf{(2-5)}$$

Since the Student's t value is a function of n, Equations 2-4 and 2-5 are applied iteratively. That is, guess at what n will be, look up $t_{1-\alpha/2,n-1}$ from Table A2, and compute a revised n. If the initial guess of n and the revised n are different, use the revised n as the new guess, and repeat the process until the computed value of n converges with the guessed value. If the population standard deviation is known (not too likely), rather than estimated, the above equation can be further simplified to

$$n = (Z_{1-\alpha/2}\sigma/d)^2 \quad \textbf{(2-6)}$$

To keep the relative error of the mean estimate below a certain level (i.e., the true mean lies between $\bar{x}-d_r\bar{x}$ and $\bar{x}+d_r\bar{x}$ with a $1-\alpha$ confidence level), the sample size can be computed with (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2,n-1}C_v/d_r)^2}{1+(t_{1-\alpha/2,n-1}C_v/d_r)^2/N} \quad \textbf{(2-7)}$$

C_v is usually less variable from study to study than are estimates of the standard deviation, which are used in Equations 2-4 through 2-6. Professional judgment and experience, typically based on previous studies, are required to estimate C_v. Had C_v been known, $Z_{1-\alpha/2}$ would have been used in place of $t_{1-\alpha/2,n-1}$ in Equation 2-7. If N is large, Equation 2-7 simplifies to:

$$n = (t_{1-\alpha/2,n-1}C_v/d_r)^2 \quad \textbf{(2-8)}$$

For Company X, harvest sites range in size from 20 to 400 acres although most are less than 80 acres in size. The goal of the sampling program is to estimate the average number of harvested acres using erosion controls. However, the investigator is concerned about skewing the mean estimate with the few large sites. As a result, the sample population for this analysis is the 430 harvested sites with less than 80 total acres. The investigator also wants to keep the relative error under 15 percent (i.e., $d_r = 0.15$) with a 90 percent confidence level.

Unfortunately, this is the first study that Company X has done and there is no information about C_v or s. The investigator, however, is familiar with a recent study done by another company. Based on that study, the investigator estimates the C_v as 0.6 and s equal to 30. As a first-cut approximation, Equation 2-6 was applied with $Z_{1-\alpha/2}$ equal to 1.645 and assuming N is large:

$$n = (1.645 * 0.6/0.15)^2 = 43.3 \approx 44 \ \textit{samples}$$

Since n/N is greater than 0.1 and C_v is estimated (i.e., not known), it is best to reestimate n with Equation 2-7 using 44 samples as the initial guess of n. In this case, $t_{1-\alpha/2,n-1}$ is obtained from Table A2 as 1.6811.

$$n = \frac{(1.6811 \times 0.6/0.15)^2}{1+(1.6811 \times 0.6/0.15)^2/430} = 40.9 \approx 41 \ \textit{samples}$$

Notice that the revised sample is somewhat smaller than the initial guess of n. In this case it is recommended to reapply the Equation 2-7 using 41 samples as the revised guess of n. In this case, $t_{1-\alpha/2,n-1}$ is obtained from Table A2 as 1.6839.

$$n = \frac{(1.6839 \times 0.6/0.15)^2}{1+(1.6839 \times 0.6/0.15)^2/430} = 41.0 \approx 41 \ \textit{samples}$$

Since the revised sample size matches the estimated sample size on which $t_{1-\alpha/2,n-1}$ was based, no further iterations are necessary. The proposed study should include 41 harvested sites randomly selected from the 430 sites with less than 80 total acres.

When interest is focused on whether the level of BMP implementation has changed, it is necessary to estimate the extent of implementation at two different time periods.

> *What sample size is necessary to determine whether there is a 20 percent difference in BMP implementation before and after an operator training program?*
>
> *What sample size is necessary to detect a 30-acre increase in average harvested acreage per site using erosion controls when comparing private and public timber sales?*

Alternatively, the proportion from two different populations can be compared. In either case, two independent random samples are taken and a hypothesis test is used to determine whether there has been a significant change in implementation. (See Snedecor and Cochran (1980) for sample size calculations for matched data.) Consider an example in which the proportion of waterbars that effectively divert water from the skid trail will be estimated at two time periods. What sample size is needed?

To compute sample sizes for comparing two proportions, p_1 and p_2, it is necessary to provide a best estimate for p_1 and p_2, as well as specifying the significance level and power (*1-$*). Recall that power is equal to the probability of rejecting H_o when H_o is false. Given this information, the analyst substitutes these values into (Snedecor and Cochran, 1980)

$$n_o = (Z_{\alpha} + Z_{2\beta})^2 \frac{(p_1 q_1 + p_2 q_2)}{(p_2 - p_1)^2} \qquad \textbf{(2-9)}$$

where $Z_{"}$ and $Z_{2\$}$ correspond to the normal deviate. Although this equation assumes that *N* is large, it is acceptable for practical use (Snedecor and Cochran, 1980). Common values of $(Z_{"} + Z_{2\$})^2$ are summarized in Table 2-5. To account for p_1 and p_2 being estimated, *Z* should be replaced with *t*. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute n_o using Equation 2-9; (2) round n_o up to the next highest integer, *f*; and (3) multiply n_o by $(f+3)/(f+1)$ to derive the final estimate of *n*.

To detect a difference in proportions of 0.20 with a two-sided test, " equal to 0.05, *1-$* equal to 0.90, and an estimate of p_1 and p_2 equal to 0.4 and 0.6, n_o is computed as

$$n_o = 10.51 \frac{[(0.4)(0.6) + (0.6)(0.4)]}{(0.6 - 0.4)^2} = 126.1$$

Rounding 126.1 to the next highest integer, *f* is equal to 127, and *n* is computed as 126.1 x 130/128 or 128.1. Therefore, 129 samples in each random sample, or 258 total samples, are needed to detect a difference in proportions of 0.2. Beware of other sources of information that give significantly lower estimates of sample size. In some cases the other sources do not specify *1-$*; in all cases, it is important that an "apples-to-apples" comparison is being made.

To compare the average from two random samples to detect a change of * (i.e., $\bar{x}_2 - \bar{x}_1$), the following equation is used:

$$n_o = (Z_{\alpha} + Z_{2\beta})^2 \frac{(s_1^2 + s_2^2)}{\delta^2} \qquad \textbf{(2-10)}$$

Table 2-5. Common values of $(Z_{\alpha} + Z_{2\beta})^2$ for estimating sample size for use with equations 2-9 and 2-10.

Power, 1-β	α for One-sided Test			α for Two-sided Test		
	0.01	0.05	0.10	0.01	0.05	0.10
0.80	10.04	6.18	4.51	11.68	7.85	6.18
0.85	11.31	7.19	5.37	13.05	8.98	7.19
0.90	13.02	8.56	6.57	14.88	10.51	8.56
0.95	15.77	10.82	8.56	17.81	12.99	10.82
0.99	21.65	15.77	13.02	24.03	18.37	15.77

Common values of $(Z_{"} + Z_{2\$})^2$ are summarized in Table 2-5. To account for s_1 and s_2 being estimated, *Z* should be replaced with *t*. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute n_o using Equation 2-10; (2) round n_o up to the next highest integer, *f*; and (3) multiply n_o by *(f+3)/(f+1)* to derive the final estimate of *n*.

Continuing the Company X example above, where *s* was estimated as 30 acres, the investigator will also want to compare the average number of harvested acres that used erosion controls to the average number of harvested acres that used erosion controls in a few years. To demonstrate success, the investigator believes that it will be necessary to detect a 20-acre increase. Although the standard deviation might change after the operator training program, there is no particular reason to propose a different *s* at this point. To detect a difference of 20 acres with a two-sided test, " equal to 0.05, *1-$* equal to 0.90, and an estimate of s_1 and s_2 equal to 30, n_o is computed as

$$n_o = 10.51 \frac{(30^2 + 30^2)}{20^2} = 47.3 \qquad \textbf{(2-11)}$$

Rounding 47.3 to the next highest integer, *f* is equal to 48, and *n* is computed as (47.3)•(51/49) or 49.2. Therefore 50 samples in each random sample, or 100 total samples, are needed to detect a difference of 20 acres.

2.3.2. Stratified Random Sampling

The key reason for selecting a stratified random sampling strategy over simple random sampling is to divide a heterogeneous population into more homogeneous groups. If populations are grouped based on size (e.g., site size) when there is a large number of small units and a few larger units, a large gain in

> *What sample size is necessary to estimate the average SMA width per harvest site when there is a wide variety of stream types and site conditions?*

precision can be expected (Snedecor and Cochran, 1980). Stratifying also allows the investigator to efficiently allocate sampling resources based on cost. The stratum mean, $\bar{x}_h$, is computed using the standard approach for estimating the mean. The overall mean, $\bar{x}_{st}$, is computed as

$$\overline{x_{st}} = \sum_{h=1}^{L} W_h \overline{x_h} \quad \textbf{(2-12)}$$

where L is the number of strata and W_h is the relative size of the hth stratum. W_h can be computed as N_h/N where N_h and N are the number of population units in the hth stratum and the total number of population units across all strata, respectively. Assuming that simple random sampling is used within each stratum, the variance of $\bar{x}_{st}$ is estimated as (Gilbert, 1987)

$$s^2(\overline{x_{st}}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h} \quad \textbf{(2-13)}$$

where n_h is the number of samples in the hth stratum and s_h^2 is computed as (Gilbert, 1987)

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{h,i} - \overline{x_h})^2 \quad \textbf{(2-14)}$$

There are several procedures for computing sample sizes. The method described below allocates samples based on stratum size, variability, and unit sampling cost. If $s^2(\bar{x}_{st})$ is specified as V for a design goal, n can be obtained from (Gilbert, 1987)

$$n = \frac{\left(\sum_{h=1}^{L} W_h s_h \sqrt{c_h}\right) \sum_{h=1}^{L} W_h s_h / \sqrt{c_h}}{V + \frac{1}{N} \sum_{h=1}^{L} W_h s_h^2} \quad \textbf{(2-15)}$$

where c_h is the per unit sampling cost in the hth stratum and n_h is estimated as (Gilbert, 1987)

$$n_h = n \frac{W_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h s_h / \sqrt{c_h}} \quad \textbf{(2-16)}$$

In the discussion above, the goal is to estimate an overall mean. To apply a stratified random sampling approach to estimating proportions, p_h, p_{st}, $p_h q_h$, and $s^2(p_{st})$ should be substituted for $\bar{x}_h$, $\bar{x}_{st}$, s_h^2, and $s^2(\bar{x}_{st})$ in the above equations, respectively.

To demonstrate the above approach, consider the Company X example again. In addition to the 430 sites that are less than 80 acres, there are 100 sites that range in size from 81 to 200 acres, 50 sites that range in size from 201 to 300 acres, and 20 sites that range in size from 301 to 400 acres. Table 2-6 presents three basic scenarios for estimating sample size. In the first scenario, s_h and c_h are assumed equal among all strata. Using a design goal of V equal to 100 and applying Equation 2-15 yields a total sample size of 41.9 or 42. Since s_h and c_h are uniform, these samples are allocated proportionally to W_h, which is referred to as *proportional allocation*. This allocation can be verified by comparing the percent sample

Table 2-6. Allocation of samples.

Farm Size (acres)	Number of Farms (N_h)	Relative Size (W_h)	Standard Deviation (s_h)	Unit Sample Cost (c_h)	Sample Allocation Number	Sample Allocation %
A) Proportional allocation (s_h and c_h are constant)						
20-80	430	0.7167	30	1	31	70.5
81-200	100	0.1667	30	1	7	15.9
201-300	50	0.0833	30	1	4	9.1
301-400	20	0.0333	30	1	2	4.5
Using Equation 2-15, *n* is equal to 41.9. Applying Equation 2-16 to each stratum yields a total of 44 samples after rounding up to the next integer.						
B) Neyman allocation (c_h is constant)						
20-80	430	0.7167	30	1	35	56.5
81-200	100	0.1667	45	1	13	21.0
201-300	50	0.0833	60	1	9	14.5
301-400	20	0.0333	75	1	5	8.1
Using Equation 2-15, *n* is equal to 59.3. Applying Equation 2-16 to each stratum yields a total of 62 samples after rounding up to the next integer.						
C) Allocation where s_h and c_h are not constant						
20-80	430	0.7167	30	1.00	38	61.3
81-200	100	0.1667	45	1.25	12	19.4
201-300	50	0.0833	60	1.50	8	12.9
301-400	20	0.0333	75	2.00	4	6.5
Using Equation 2-15, *n* is equal to 60.0. Applying Equation 2-16 to each stratum yields a total of 62 samples after rounding up to the next integer.						

allocation to W_h. Due to rounding up, a total of 44 samples are allocated.

Under the second scenario, referred to as the *Neyman allocation*, the variability between strata changes, but unit sample cost is constant. In this example, s_h increases by 15 between strata. Because of the increased variability in the last three strata, a total of 59.3 or 60 samples are needed to meet the same design goal. So while more samples are taken in every stratum, proportionally fewer samples are needed in the smaller site size group. For example, using proportional allocation, more than 70 percent of the samples are taken in the 20- to 80-acre site size stratum, whereas approximately 57 percent of the samples are taken in the same stratum using the Neyman allocation.

Finally, introducing sample cost variation will also affect sample allocation. In the last scenario it was assumed that it is twice as expensive to evaluate a harvest site from the largest size stratum than to evaluate a harvest site from the smallest size stratum. In this example, roughly the same total number of samples are needed to meet the design goal, yet more samples are taken in the smaller size stratum.

2.3.3. Cluster Sampling

Cluster sampling is commonly used when there is a choice between the size of the sampling unit (e.g., skid trail versus harvest site). In general, it is cheaper to sample larger units than smaller units, but the results tend to be less accurate (Snedecor and Cochran, 1980). Thus, if there is not a unit sampling cost advantage to cluster sampling, it is probably better to use simple random sampling. To decide whether to perform cluster sampling, it will probably be necessary to perform a special investigation to quantify sampling errors and costs using the two approaches.

Perhaps the best approach to explaining the difference between simple random sampling and cluster sampling is to consider an example set of results. In this example, the investigator did an evaluation to determine whether harvest sites had adequate SMAs. Since the state had timber harvesting activities across the state, the investigator elected to inspect 10 harvest sites along each randomly selected river. Table 2-7 presents the number of harvest sites along each river that had the recommended BMPs. The overall mean is 5.6; a little more than one-half of the sites have implemented the recommended BMPs. However, note that since the population unit corresponds to the 10 sites collectively, there are only 30 samples and the standard error for the proportion of sites using recommended BMPs is 0.035. Had the investigator incorrectly calculated the standard error using the random sampling equations, he or she would have computed 0.0287, nearly a 20 percent error.

Since the standard error from the cluster sampling example is 0.035, it is possible to estimate the corresponding simple random sample size to obtain the same precision using

$$n = \frac{p\,q}{s(p)^2} = \frac{(0.56)(0.44)}{0.035^2} = 201 \qquad \textbf{(2-17)}$$

Is collecting 300 samples using a cluster sampling approach cheaper than collecting about 200 simple random samples? If so, cluster sampling should be used; otherwise, simple random sampling should be used.

2.3.4. Systematic Sampling

It might be necessary to obtain an estimate of the proportion of harvest sites where cable yarding was implemented using site inspections. Assuming a record of harvest

Table 2-7. Number of harvest sites (out of 10) implementing recommended BMPs.

3	9	5	7	6	4	6	3	5	5
5	7	7	4	7	5	3	8	4	6
8	4	7	4	5	3	3	9	9	7

Grand Total = 168
$\bar{x}$ = 5.6 p = 5.6/10=0.560
s = 1.923 s = 1.923/10=0.1923

Standard error using cluster sampling: s(p)=0.1923/(30)$^{0.5}$=0.035
Standard error if simple random sampling assumption had been incorrectly used:
s(p)=((0.56)(1-0.56)/300)$^{0.5}$ =0.0287

sites (where cable yarding was specified in the timber sale contract or administration file) is available in a sequence unrelated to the manner in which this BMP would be implemented (e.g., in alphabetical order by the operator's name), a systematic sample can be obtained by selecting a random number *r* between 1 and *n*, where *n* is the number required in the sample (Casley and Lury, 1982). The sampling units are then *r*, *r + (N/n)*, *r + (2N/n)*, ..., *r + (n-1)(N/n)*, where *N* is total number of available records.

If the population units are in random order (e.g., no trends, no natural strata, uncorrelated), systematic sampling is, on average, equivalent to simple random sampling.

Once the sampling units (in this case, specific harvest sites) have been selected, site inspections can be made to assess the extent of compliance with cable yarding standards and specifications.