MEASURING
&
MONITORING
Plant Populations

AUTHORS:

Caryl L. Elzinga Ph.D.
Alderspring Ecological Consulting
P.O. Box 64
Tendoy, ID 83468

Daniel W. Salzer
Coordinator of Research and Monitoring
The Nature Conservancy of Oregon
821 S.E. 14th Avenue
Portland, OR 97214

John W. Willoughby
State Botanist
Bureau of Land Management
California State Office
2135 Butano Drive
Sacramento, CA 95825

This technical reference represents a team effort by the three authors. The order of authors is alphabetical and does not represent the level of contribution.

Though this document was produced through an interagency effort, the following BLM numbers have been assigned for tracking and administrative purposes:

BLM Technical Reference 1730-1

BLM/RS/ST-98/005+1730

CHAPTER 5. Basic Principles of Sampling

A. Introduction

What is sampling? A review of several dictionary definitions led to the following composite definition:

> The act or process of selecting a part of something with the intent of showing the quality, style, or nature of the whole.

Monitoring does not always involve sampling techniques. Sometimes you can count or measure all individuals within a population of interest. Other times you may select qualitative techniques that are not intended to show the quality, style, or nature of the whole population (e.g., subjectively positioned photoplots).

What about those situations where you have an interest in learning something about the entire population, but where counting or measuring all individuals is not practical? This situation calls for sampling. The role of sampling is to provide information about the population in such a way that inferences about the total population can be made. This inference is the process of generalizing to the population from the sample, usually with the inclusion of some measure of the "goodness" of the generalization (McCall 1982).

Sampling will not only reduce the amount of work and cost associated with characterizing a population, but sampling can also increase the accuracy of the data gathered. Some kinds of errors are inherent in all data collection procedures, and by focusing on a smaller fraction of the population, more attention can be directed toward improving the accuracy of the data collected.

This chapter includes information on basic principles of sampling. Commonly used sampling terminology is defined and the principal concepts of sampling are described and illustrated. Even though the examples used in this chapter are based on counts of plants in quadrats (density measurements), most of the concepts apply to all kinds of sampling.

B. Populations and Samples

The term "population" has both a biological definition and a statistical definition. In this chapter and in Chapter 7, we will be using the term "population" to refer to the statistical population or the "sampling universe" in which monitoring takes place. The statistical population will sometimes include the entire biological population, and other times, some portion of the biological population. The *population* consists of the complete set of individual objects about which you want to make inferences. We will refer to these individual objects as *sampling units*. The sampling units can be individual plants or they may be quadrats (plots), points, or transects. The *sample* is simply part of the population, a subset of the total possible number of sampling units. These terms can be clarified in reference to an artificial plant population shown in Figure 5.1. There are a total of 400 plants in this population, distributed in 20 patches of 20 plants each. All the plants are contained within the boundaries of a 20m x 20m "macroplot." The collection of plants in this macroplot population will be referred to as the "400-plant population." A random arrangement of ten 2m x 2m quadrats positioned within the 400-plant population is shown in

Figure 5.1. Counts of plants within the individual quadrats are directed at the objective of estimating the total number of plants in the 20m x 20m macroplot. The sampling unit in this case is the 2m x 2m quadrat. The sample shown in Figure 5.1 is a set of 10 randomly selected quadrats. The population in this case is the total collection of all possible 2m x 2m quadrats that could be placed in the macroplot (N=100).

Sample information

Coordinates X	Coordinates Y	# of plants
2	2	4
6	4	0
16	4	3
12	6	2
14	6	5
6	8	10
0	12	0
2	12	6
14	12	0
2	14	20

Population parameters
Total population size:
400 plants
Mean # plants/quadrat:
$\mu = 4$
Standard deviation:
$\sigma = 5.005$

Sample statistics (n = 10)
Mean # plants/quadrat:
$\bar{x} = 5.0$
Standard deviation:
$s = 6.146$

Population estimate
Estimated population
size = 500 plants
95% confidence
interval = ± 361 plants

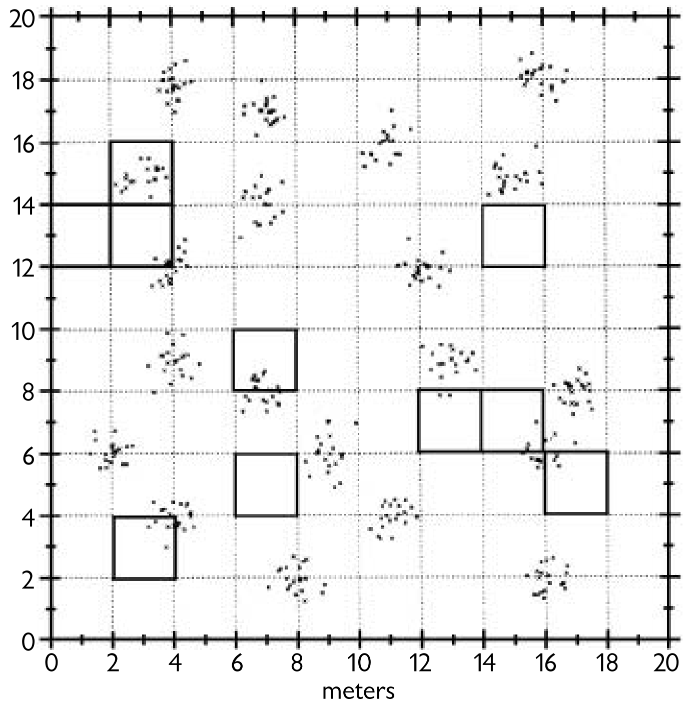


FIGURE 5.1. Population of 400 plants distributed in 20 clumps of 20 plants. This figure shows a simple random sample of ten 2m x 2m quadrats, along with sample statistics and true population parameters.

C. Population Parameters vs. Sample Statistics

Population parameters are descriptive measures which characterize the population and are assumed to be fixed but unknown quantities that change only if the population changes. Greek letters such as μ and σ are often used to denote parameters. If we count all the plants in all the quadrats that make up the 400-plant population shown in Figure 5.1 (400 plants), and divide by the total number of possible 2m x 2m quadrat locations in the macroplot (100 quadrats), we obtain the true average number of plants per quadrat (4 plants/quadrat). This, assuming we have made no errors, is the *true population mean (μ)*. If we know how much each individual quadrat differs from the true population mean, we can calculate another important population parameter, the *true population standard deviation (σ)*. The standard deviation is a measure of how similar each individual observation is to the overall mean and is the most common measure of variability used in statistics. Populations with a large amount of variation among possible sampling units will have a larger standard deviation than populations with sampling units that are more similar to one another.

Sample statistics are descriptive measures derived from a sample (e.g., 10 of the possible 100 2m x 2m quadrats). Sample statistics provide estimates of population parameters. Sample statistics will vary from sample to sample in addition to changing whenever the underlying population changes. Roman letters such as $\bar{X}$ and s are usually used for sample statistics. Consider the following simple example where a sample of three sampling units yields values of 9, 10, and 14 plants/quadrat:

The *sample mean* ($\bar{X}$) = (9+10+14)/3 = 11 plants/quadrat

We could also calculate from this sample a *sample standard deviation (s)*. The sample standard deviation describes how similar each individual observation is to the sample mean. The derivation of a standard deviation (in case you want to calculate one by hand) is provided in Appendix 8. The standard deviation is easily calculated with a simple hand calculator using the "s" or "σ_{n-1}" key.

The standard deviation (s) for the simple example above is 2.65 plants/quadrat.

Consider another simple example with sampling unit values of 2, 10, and 21 plants/quadrat.

The mean ($\bar{X}$) = (2+10+21)/3 = 11 plants/quadrat

The standard deviation (s) for this example is 9.54 plants/quadrat.

Thus, both examples have a sample mean of 11 plants/quadrat, but the second one has a higher standard deviation (9.54 plants/quadrat) than the first (2.65 plants/quadrat), because the individual quadrat values differ more from one another in the second example.

In the example shown in Figure 5.1, the true population mean is 4.00 plants/quadrat, whereas the sample mean is 5.00 plants/quadrat. The true population standard deviation is 5.005 plants/quadrat, whereas the sample standard deviation is 6.146 plants/quadrat.

D. Accuracy vs. Precision

Accuracy is the closeness of a measured or computed value to its true value. *Precision* is the closeness of repeated measurements of the same quantity. A simple example will help illustrate the difference between these two terms. Two quartz clocks, equally capable of tracking time, are sitting side-by-side on a table. Someone comes by and advances one of the clocks by 1 hour. Both clocks will be equally "precise" at tracking time, but one of them will not be "accurate."

Efficient sampling designs try to achieve high precision. When we sample to estimate some population parameter, our sample standard deviation gives us a measure of the repeatability, or precision of our sample; it does not allow us to assess the accuracy of our sample. If counts of plants within different quadrats of a sample are similar to one another (e.g., the example above with a mean of 11 and a standard deviation = 2.65) then it is likely that different independent samples from the same population will yield similar sample means and give us high precision. When quadrat counts within a sample are highly variable (e.g., the example above with a mean of 11 and a standard deviation of 9.54), individual sample means from separate independent samples may be very different from one another giving us low precision. In either case, if the counting process is biased (perhaps certain color morphs or growth forms of individuals are overlooked), results may be inaccurate.

E. Sampling vs. Nonsampling errors

In any monitoring study errors should be minimized. Two categories of errors are described next.

1. Sampling errors

Sampling errors result from chance; they occur when sample information does not reflect the true population information. These errors are introduced by measuring only a subset of all the sampling units in a population.

Sampling errors are illustrated in Figure 5.2, in which two separate, completely random samples (2A and 2B) are taken from the 400-plant population shown in Figure 5.1. In each case, ten 2m x 2m quadrats are sampled and an estimate is made of the total number of plants within the population. The sample shown in Figure 5.2A produces a population estimate of only 80 plants, whereas the sample shown in Figure 5.2B yields an estimate of 960 plants. Both estimates are poor because of sampling error (chance placement of the quadrats resulted in severe under- or overestimates of the true population total).

You can imagine the problems that can arise if you monitor the same population two years in a row and get sample information that indicates that the population shifted from 960 plants to 80 plants when it really didn't change at all. Sampling errors can lead to two kinds of mistakes: (1) missing real changes (missed-change errors), and (2) detecting apparent changes that don't really exist (false-change errors).

Sampling errors can be estimated from the sampling data. Some of the basic sampling design tools covered in Chapter 7, enable you to evaluate the effectiveness of your monitoring study by taking a closer look at the sampling data. This can be especially helpful when setting up new projects; an evaluation of pilot sampling data can point out potential sampling error problems, enabling an investigator to fix them at an early stage of the project. Good sampling designs can reduce sampling errors without increasing the cost of sampling.

2. Nonsampling errors

Nonsampling errors are errors associated with human, rather than chance, mistakes. Examples of nonsampling errors include:

- Using biased selection rules, such as selecting "representative samples" by subjectively locating sampling units, or by substituting sampling units that are "easier" to measure.
- Using vegetation measurement or counting techniques within sampling units in which attributes cannot be accurately counted or measured. For example, counts of grass stems within a quadrat with counts in the hundreds may lead to numerous counting errors.
- Inconsistent field sampling effort. Nonsampling errors can be introduced if different investigators use different levels of effort (e.g., one investigator makes counts from "eye-level," whereas another counts by kneeling next to the quadrat).
- Transcription and recording errors. Nonsampling errors can be introduced if the data recorder's "7's" look like "1's" to the person entering the data.
- Incorrect or inconsistent species identification. This category also includes biases introduced by missing certain size classes or color morphs.

A

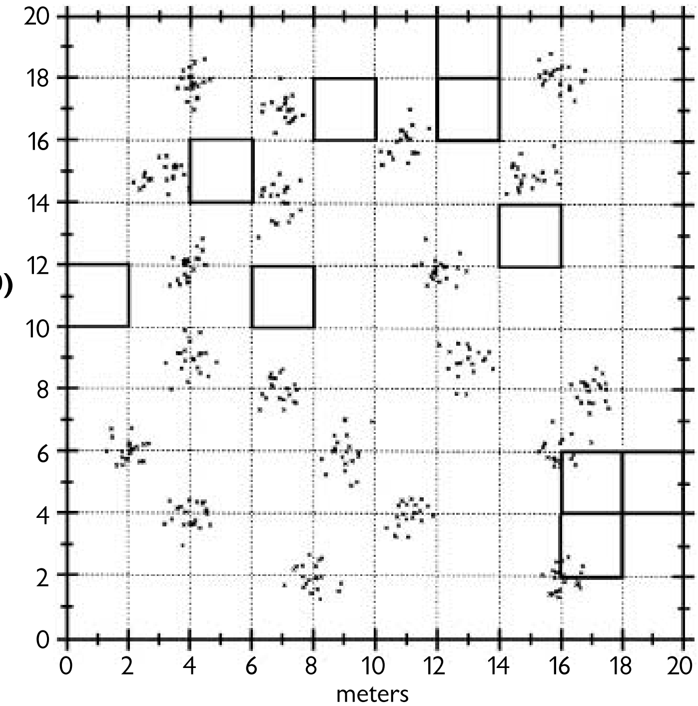
Sample information

Coordinates X	Coordinates Y	# of plants
16	2	5
16	4	3
18	4	0
0	10	0
6	10	0
14	12	0
4	14	0
8	16	0
12	16	0
12	18	0

Population parameters
Total population size:
400 plants
Mean # plants/quadrat:
$\mu = 4$
Standard deviation:
$\sigma = 5.005$

Sample statistics (n = 10)
Mean # plants/quadrat:
$\bar{x} = 0.8$
Standard deviation:
$s = 1.75$

Population estimate
Estimated population
size = 80 plants
95% confidence
interval = ± 119 plants



B

Sample information

Coordinates X	Coordinates Y	# of plants
16	0	5
10	2	11
16	4	3
14	6	5
12	8	18
6	8	10
4	16	9
2	12	6
10	16	9
14	14	20

Population parameters
Total population size:
400 plants
Mean # plants/quadrat:
$\mu = 4$
Standard deviation:
$\sigma = 5.005$

Sample statistics (n = 10)
Mean # plants/quadrat:
$\bar{x} = 9.6$
Standard deviation:
$s = 5.58$

Population estimate
Estimated population
size = 960 plants
95% confidence
interval = ± 379 plants

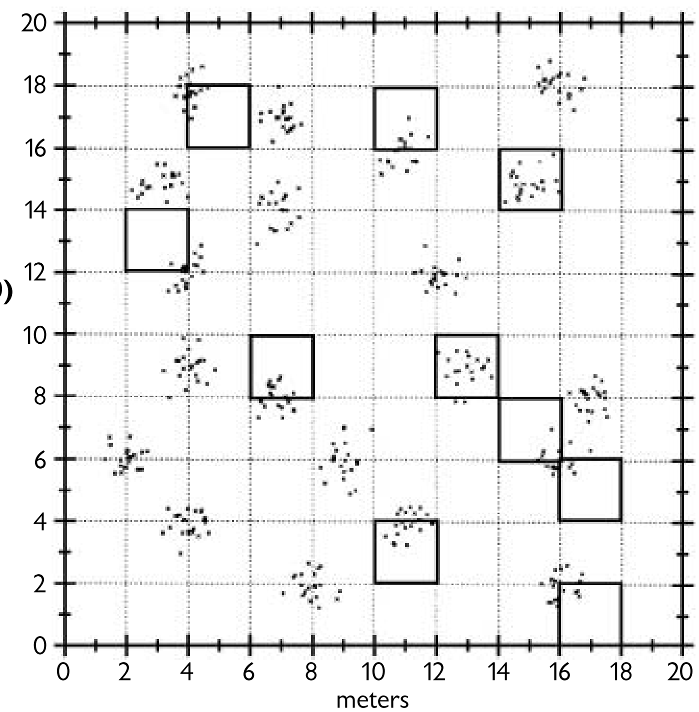


FIGURE 5.2. Examples of sampling errors from sampling the 400-plant population. The population estimates of 80 plants and 960 plants are far from the true population of 400 plants.

Because sampling designs are based on the assumption that nonsampling errors are zero, the number of nonsampling errors needs to be minimized. Ensure that your sampling unit makes sense for the type of vegetation measurement technique you have selected. When different personnel are used in the same monitoring study, conduct rigorous training and testing to ensure consistency in counts or measurements. Design field data forms (Chapter 9) that are easy to use and easy for data transcribers to interpret. Proof all data entered into computer programs to ensure that entered numbers are correct. In contrast to sampling errors, the probability of nonsampling errors occurring cannot be assessed from pilot sample data.

F. Sampling Distributions

One way of evaluating the risk of obtaining a sample value that is vastly different than the true value (such as population estimates of 80 or 960 plants when the true population is 400 plants) is to sample a population repeatedly and look at the differences among the repeated population estimates. If almost all the separate, independently derived population estimates are similar, then you know you have a good sampling design with high precision. If many of the independent population estimates are not similar, then you know your precision is low.

The 400-plant population can be resampled by erasing the 10 quadrats (as shown in either Figure 5.1 or Figure 5.2) and putting 10 more down in new random positions. We can keep repeating this procedure, each time writing down the sample mean. Plotting the results of a large number of individual sample means in a simple histogram graph yields a *sampling distribution*. A sampling distribution is a distribution of many independently gathered sample statistics (most often a distribution of sample means). Under most circumstances, this distribution of sample means fits a normal or bell-shaped curve.

A distribution of population size estimates from 10,000 separate random samples using ten 2m x 2m quadrats from the 400-plant population is shown in Figure 5.3A. The x-axis shows the range of different population estimates, and the y-axis shows the relative and actual frequency of the different population estimates. Think of this as the results of 10,000 different people sampling the same population on the same day, each one setting out 10 randomly positioned 2m x 2m quadrats (somehow without negatively impacting the population) and coming up with their own independent population estimate. The highest population estimate out of the 10,000 separate samples was 960 plants and the lowest population estimate was zero (four of the 10,000 samples yielded a population estimate of zero). The shape of this distribution indicates the magnitude of likely sampling errors. Wide distributions could yield population estimates that are "far" from the true population value. A sampling design that led to the type of sampling distribution depicted in Figure 5.3A would not be useful since few of the estimates approach the true population size of 400 plants. *One of the principal objectives in sampling design is to make the shape of sampling distributions as narrow as possible*.

Fortunately, you do not have to repeatedly sample your population and see how wide your sampling distribution is to determine if you need to change anything. There are some simple statistical tools that provide a convenient shortcut for evaluating the precision of your sampling effort from a single sample. These tools involve calculating standard errors and confidence intervals to estimate sampling precision levels.

1. Standard error

A *standard error* is the standard deviation of a large number of independent sample means. It is a measure of precision that you derive from a single sample. The formula for calculating a standard error is as follows:

Formula for standard error:

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

Where: SE = Standard error
s = Standard deviation
n = Sample size

To paraphrase the earlier statement regarding an important objective of sampling design, *one of the principal objectives in sampling design is to reduce the size of the standard error*.

This formula demonstrates that there are only two ways of minimizing standard errors, either: (1) increase sample size (n), or (2) decrease the standard deviation (s).

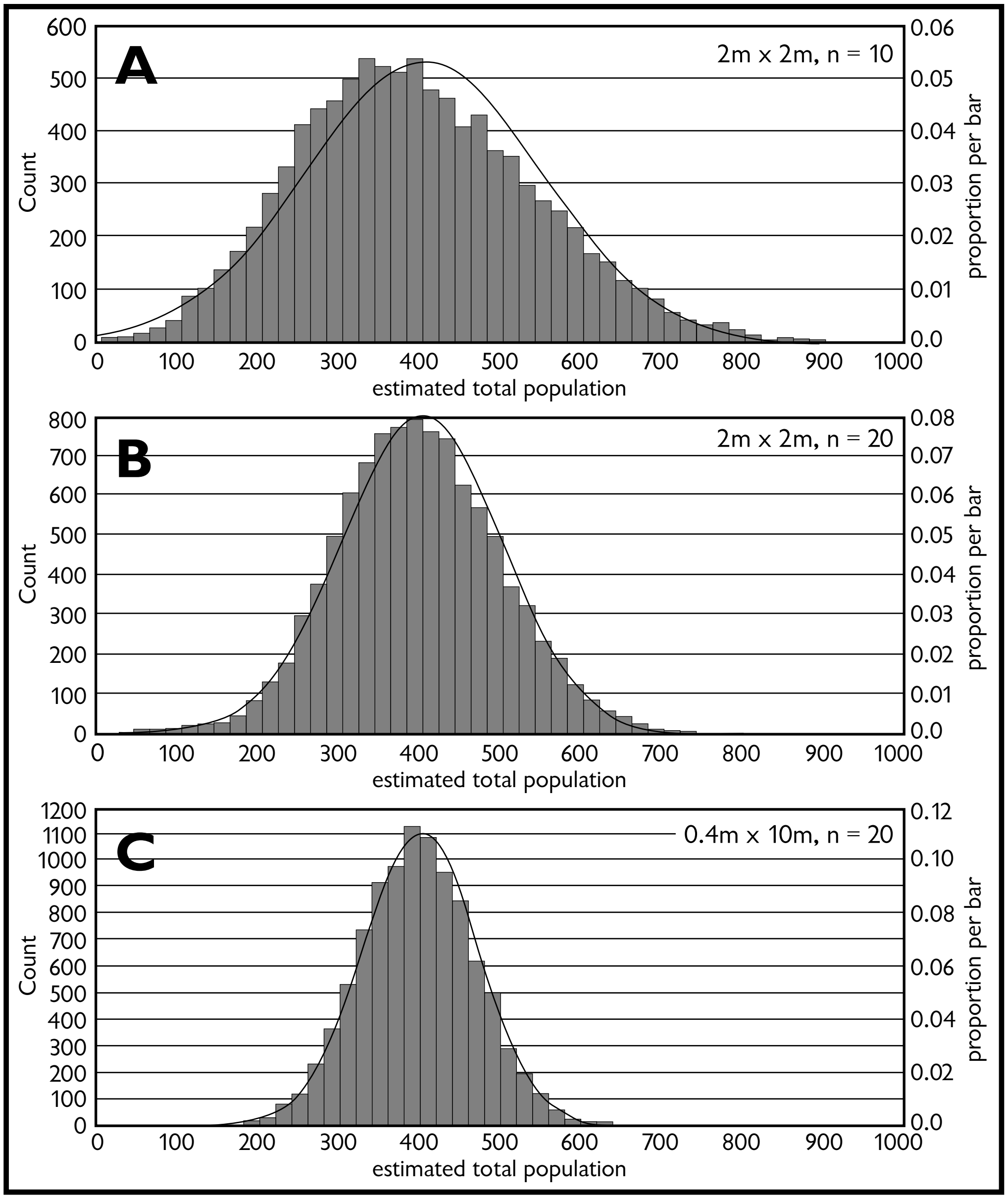


FIGURE 5.3. Sampling distributions from three separate sampling designs used on the 400-plant population. All distributions were created by sampling the population 10,000 separate times. The smooth lines show a normal bell-shaped curve fit to the data. Figure 3A shows a sampling distribution where ten 2m x 2m quadrats were used. Figure 3B shows a sampling distribution where twenty 2m x 2m quadrats were used. Figure 3C shows a sampling distribution where twenty 0.4m x 10m quadrats were used.

- **Increase sample size**. A new sampling distribution of 10,000 separate random samples drawn from our example population is shown in Figure 5.3B. This distribution came from randomly drawing samples of *twenty* 2m x 2m quadrats instead of the *ten* quadrats used to create the sampling distribution in Figure 5.3A. This increase in sample size from 10 to 20 provides a 29.3% improvement in precision (as measured by the reduced size of the standard error).

- **Decrease sample standard deviation**. Another sampling distribution of 10,000 separate random samples drawn from our 400-plant population is shown in Figure 5.3C. The sampling design used to create this distribution of population estimates is similar to the one used to create the sampling distribution in Figure 5.3B. The only difference between the two designs is in quadrat shape. The sampling distribution shown in Figure 5.3B came from using twenty 2m x 2m quadrats; the sampling distribution shown in Figure 5.3C came from using twenty 0.4m x 10m quadrats. This change in quadrat shape reduced the true population standard deviation from 5.005 plants to 3.551 plants. This change in quadrat shape led to a 29.0% improvement in precision over the 2m x 2m design shown in Figure 5.3B (as measured by the reduced size of the standard error). This 29.0% improvement in precision came without changing the sampling unit size ($4m^2$) or the number of quadrats sampled (n=20); only the quadrat shape (from square to rectangular) changed. When compared to the original sampling design of ten 2m x 2m quadrats, the twenty 0.4m x 10m quadrat design led to a 49.8% improvement in precision. Details of this method and other methods of reducing sample standard deviation are covered in Chapter 7.

How is the standard error most often used to report the precision level of sampling data? Sometimes the standard error is reported directly. You may see tables with standard errors reported or graphs that include error bars that show ± 1 standard error. Often, however, the standard error is multiplied by a coefficient that converts the number into something called a confidence interval.

2. Confidence intervals

A confidence interval provides an estimate of precision around a sample mean, a sample proportion, or an estimate of total population size that specifies the likelihood that the interval includes the true value. The vertical lines marked with the "95%" in Figure 5.4 indicate that 95% of all the samples (9,500 out of the 10,000) fit between these two lines. Five percent of the samples (2.5% in each tail of the distribution) fall outside the vertical lines. These lines

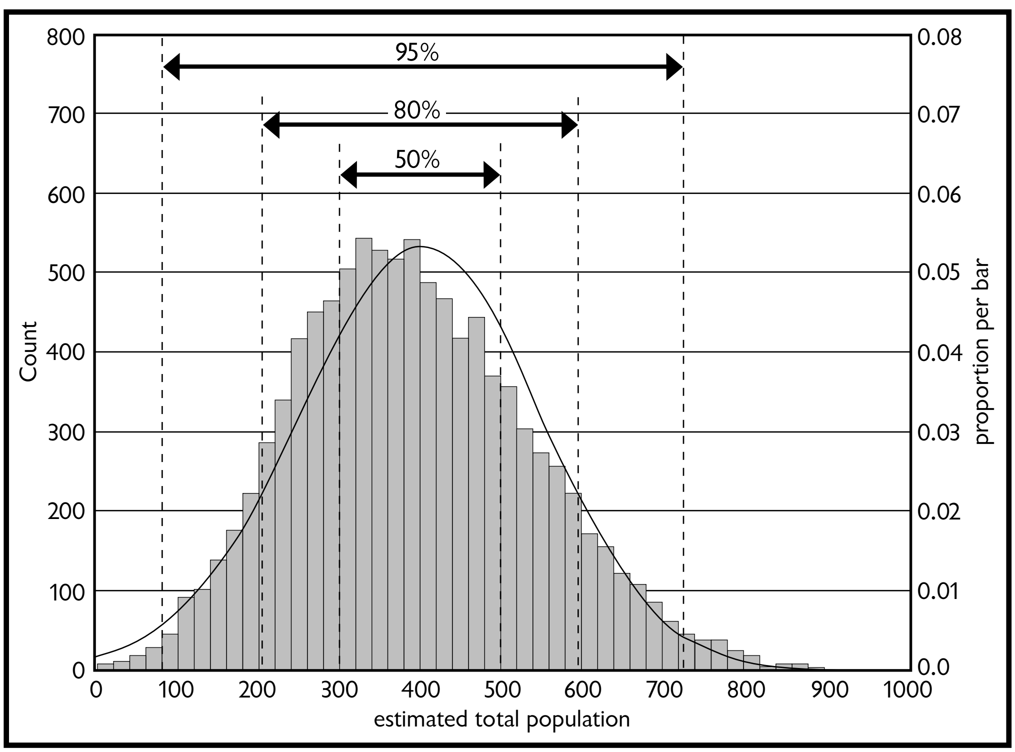


FIGURE 5.4. Distribution from sampling the 400-plant population 10,000 times using ten samples of 2m x 2m quadrats. The 95%, 80%, and 50% confidence intervals around the true population of 400 plants are shown. The smooth line shows a normal, bell-shape curve fit to the data.

are positioned equally from the center of the sampling distribution, approximately 320 plants away from the center of 400 plants. Thus, 95% of all samples are within ± 320 plants of the true population size.

A confidence interval includes two components: (1) the *confidence interval width* (e.g., ± 320 plants); and (2) the *confidence level* (e.g., 90%, 95%). The confidence level indicates the probability that the interval includes the true value. Confidence interval width decreases as the confidence level decreases. This relationship is shown in Figure 5.4 where three different confidence levels are graphed on the sampling distribution obtained by sampling the 400-plant population with ten 2m x 2m quadrats. These same three confidence intervals are shown again in Figure 5.5A, where they are graphed in a format commonly used to report confidence intervals. There is no gain in precision associated with the narrowing of confidence interval width as you go from left to right in Figure 5.5A (i.e., from 95% confidence, to 80% confidence, to 50% confidence); only the probability that the confidence interval includes the true value is altered. Another set of three confidence intervals is shown in Figure 5.5B. Like Figure 5.5A, confidence intervals get narrower as we move from left to right in the graph, but this time the confidence level is the same (95%) and the narrower widths came from using different sampling designs. There is a gain in precision associated with the narrowing of confidence interval width as you go from left to right in Figure 5.5B (i.e., from the ten 2m x 2m design to the twenty 2m x 2m design to the twenty 0.4m x 10m design) because we have reduced the uncertainty of our population estimate by tightening the confidence interval width at the same confidence level.

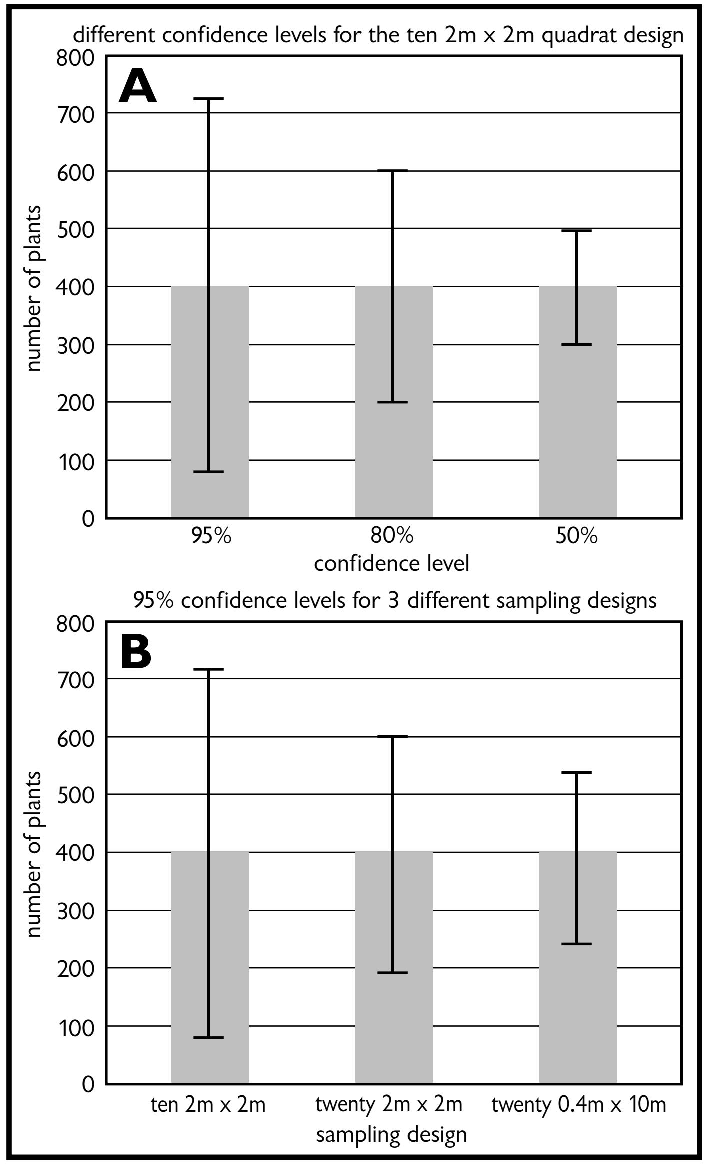


FIGURE 5.5. Comparison of confidence intervals and confidence levels for different sampling designs from the 400-plant population. Figure 5A shows three different confidence levels (95%, 80%, and 50%) for the same data set based upon sampling ten 2m x 2m quadrats. Figure 5B shows 95% confidence intervals for three different sampling designs that differ in the level of precision of the population estimates.

In order to calculate confidence intervals for sample means, we need two values: (1) the standard error calculated according to the above formula ($SE = s/\sqrt{n}$), and (2) the corresponding value from a table of critical values of the *t* distribution (see Appendix 8 for instructions on calculating confidence intervals around proportions). The confidence interval half-width, extending an equal distance on both sides of the mean, is the standard error × the critical *t* value (except when sampling from finite populations, see below). The appropriate critical value of *t* depends on the level of confidence desired and the number of sampling units *(n)* in the sample. A table of critical values for the *t* distribution (Zar 1996) is found in Appendix 5. To use this table, you must first select the appropriate confidence level column. If you want to be 95% confident that your confidence interval includes the true mean, use the column headed $\alpha(2) = 0.05$. For 90% confidence, use the column headed $\alpha(2) = 0.10$. You use $\alpha(2)$ because you are interested in a confidence interval on both sides of the mean. You then use the row indicating the number of degrees of freedom *(v)*, which is the number of sampling units minus one (n-1).

For example, if we sample 20 quadrats and come up with a mean of 4.0 plants and a standard deviation of 5.005, here are the steps for calculating a 95% confidence interval around our sample mean:

The standard error ($SE = s/\sqrt{n}$) = $\frac{5.005}{\sqrt{20}}$ = 5.005/4.472 = 1.119.

The appropriate *t* value from Appendix 5 for 19 degrees of freedom *(v)* is 2.093.

One-half of our confidence interval width is then SE × *t*-value = 1.119 × 2.093 = 2.342.

Our 95% confidence interval can then be reported as 4.0 ± 2.34 plants/quadrat or we can report the entire confidence interval width from 1.66 to 6.34 plants/quadrat. This indicates a 95% chance that our interval from 1.66 plants/quadrat to 6.34 plants/quadrat includes the true value.

Another way to think of 95% confidence intervals calculated from sampling data is that the interval specifies a range that should include the true value 95% of the time. If you are calculating 95% confidence intervals and independently randomly sample a population 100 different times, you should see that approximately 95 of the intervals will include the true mean and 5 will miss it. This relationship is shown in Figure 5.6 where 100 independent population estimates are graphed with 95% confidence intervals from the 400-plant populations using samples of twenty 0.4m x 10m quadrats. You will notice that the solid dots, used to show each of the 100 population estimates, fluctuate around the true population value of 400 plants. You will also notice that 96 out of 100 confidence intervals shown in Figure 5.6 include the true value. If the confidence level had been set at 80%, then approximately 20 of the intervals would have failed to include the true value. A 99% confidence level would have led to approximately only one interval out of the 100 that did not include the true population size (in order to capture the true value more often, the individual confidence interval widths for a 99% confidence level are wider than the confidence interval widths for a 95% confidence level).

G. Finite vs. Infinite Populations

If we are sampling with quadrats, there is a finite number of quadrats that can be placed in the area to be sampled, assuming that no two quadrats overlap (this is called sampling without replacement). If the sampled area is large, then the number of quadrats placed in the area may be very large as well, but nonetheless finite. On the other hand, an infinite number of points or lines could be placed in the area to be sampled. This is because points, at least theoretically, are dimensionless, and lines are dimensionless in one direction. This means, at least for all practical purposes, that a population of points or of lines is infinite.

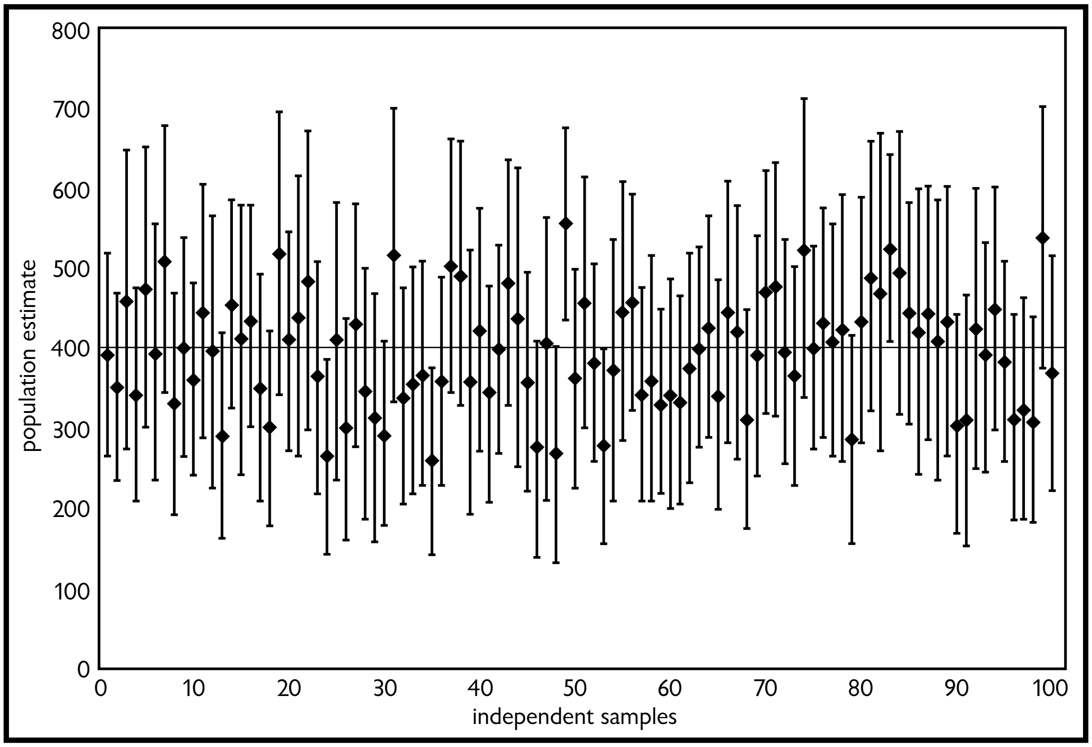


FIGURE 5.6. Population estimates from 100 separate random samples from the 400-plant population. Each sample represents the population estimate from sampling twenty 0.4m x 10m quadrats. The horizontal line through the graph indicates the true population of 400 plants. Vertical bars represent 95% confidence intervals. Four of the intervals miss the true population size.

If the area to be sampled is large relative to the area that is actually sampled, the distinction between finite and infinite is of only theoretical interest. When, however, the area sampled makes up a significant portion of the area to be sampled, we can apply the *finite population correction factor*, which reduces the size of the standard error. The most commonly used finite population correction factor is shown to the right:

Formula for the finite population correction:

$$\text{FPC} = \sqrt{\frac{N - n}{N}}$$

Where: N = total number of potential quadrat positions
n = number of quadrats sampled

When n is small relative to N, the equation is close to 1, whereas when n is large relative to N, the value approaches zero. The standard error ($s/\sqrt{n}$) is multiplied by the finite population correction factor to yield a corrected standard error for the finite population.

Consider the following example. The density of plant species X is estimated within a 20m x 50m macroplot (total area = 1000m^2). This estimate is obtained by collecting data from randomly selected 1m x 10m quadrats (10m^2). Sampling without replacement, there are 100 possible quadrat positions. $\frac{\text{total area}}{\text{quadrat area}} = \frac{1000\text{m}^2}{10\text{m}^2}$

Thus, our population, N, is 100. Let's say we take a random sample, n, of 30 of these quadrats and calculate a mean of eight plants per quadrat and a standard deviation of four plants per quadrat. Our standard error is thus: $s/\sqrt{n} = 4/\sqrt{30} = 0.73$. Although our sample mean is an unbiased estimator of the true population mean and needs no correction, the standard error should be corrected by the finite population correction factor shown on the top of page 72:

Example of applying the finite population correction factor:

$$SE' = (SE)\sqrt{\frac{N - n}{N}} \qquad SE' = (0.73)\sqrt{\frac{100 - 30}{100}} = 0.61$$

Where: SE' = corrected standard error
SE = uncorrected standard error
N = total number of potential quadrat positions
n = number of quadrats sampled

Since the standard error is one of the factors used to calculate confidence intervals (the other is the appropriate value of *t* from a *t* table), correcting the standard error with the finite population correction factor makes the resulting confidence interval narrower. It does this, however, *only* if *n* is sufficiently large relative to *N*. A rule of thumb is that unless the ratio *n/N* is greater than .05 (i.e., you are sampling more than 5% of the population area), there is little to be gained by applying the finite population correction factor to your standard error.

The finite population correction factor is also important in sample size determination (Chapter 7) and in adjusting test statistics (Chapter 11). The finite population correction factor works, however, only with finite populations, which we will have when using quadrats, but will not have when using points or lines.

H. False-Change Errors and Statistical Power Considerations

These terms relate to situations where two or more sample means or proportions are being compared with some statistical test. This comparison may be between two or more places or the same place between two or more time periods. These terms are pertinent to both planning and interpretation stages of a monitoring study. Consider a simple example where you have sampled a population in two different years and now you want to determine whether a change took place between the two years. You usually start with the assumption, called the null hypothesis, that no change has taken place. There are two types of decisions that you can make when interpreting the results of a monitoring study: (1) you can decide that a change took place, or (2) you can decide that no change took place. In either case, you can either be right, or you can be wrong (Figure 5.7).

monitoring for change — possible errors

	no change has taken place	there has been a real change
monitoring system detects a change	false-change error (Type I) α	no error (Power) $1 - \beta$
monitoring system detects no change	no error $(1 - \alpha)$	missed-change error (Type II) β

FIGURE 5.7. Four possible outcomes for a statistical test of some null hypothesis, depending on the true state of nature.

1. The change decision and false-change errors

The conclusion that a change took place may lead to some kind of action. For example, if a population of a rare plant is thought to have declined, a change in management may be needed. If a change did not actually occur, this constitutes a false-change error, a sort of false alarm. Controlling this type of error is important because taking action unnecessarily can be expensive (e.g., a range permittee is not going to want to change the grazing intensity if a decline in a rare plant population really didn't take place). There will be a certain probability of concluding that a change took place even if no difference actually occurred. The probability of this occurring is usually labeled the *P*-value, and it is one of the types of information that comes out of a statistical analysis of the data. The *P*-value reports the likelihood that the observed difference was due to chance alone. For example, if a statistical test comparing two sample means yields a *P*-value of 0.24 this indicates that there is a 24% chance of obtaining the observed result even if there is no true difference between the two sample means.

Some threshold value for this false-change error rate should be set in advance so that the *P*-value from a statistical test can be evaluated relative to the threshold. *P*-values from a statistical test that are smaller than or equal to the threshold are considered statistically "significant," whereas *P*-values that are larger than the threshold are considered statistically "nonsignificant." Statistically significant differences may or may not be ecologically significant depending upon the magnitude of difference between the two values. The most commonly cited threshold for false-change errors is the 0.05 level; however, there is no reason to arbitrarily adopt the 0.05 level as the appropriate threshold. The decision of what false-change error threshold to set depends upon the relative costs of making this type of mistake and the impact of this error level on the other type of mistake, a missed-change error (see below).

2. The no-change decision, missed-change errors, and statistical power

The conclusion that no change took place usually does not lead to changes in management practices. Failing to detect a true change constitutes a *missed-change error*. Controlling this type of error is important because failing to take action when a true change actually occurred may lead to the serious decline of a rare plant population.

Statistical power is the complement of the missed-change error rate (e.g., a missed-change error rate of 0.25 gives you a power of 0.75; a missed-change error rate of 0.05 gives you a power of 0.95). High power (a value close to 1), is desirable and corresponds to a low risk of a missed-change error. Low power (a value close to 0) is not desirable because it corresponds to a high risk of a missed-change error.

Since power levels are directly related to missed-change error levels, either level can be reported and the other level can be easily calculated. Power levels are often reported instead of missed-change error levels, because it seems easier to convey this concept in terms of the certainty of detecting real changes. For example, the statement "I want to be at least 90% certain of detecting a real change of five plants/quadrat" (power is 0.90) is simpler to understand than the statement "I want the probability of missing a real change of five plants/quadrat to be 10% or less" (missed-change error rate is 0.10).

An assessment of statistical power or missed-change errors has been virtually ignored in the field of environmental monitoring. A survey of over 400 papers in fisheries and aquatic sciences found that 98% of the articles that reported nonsignificant results failed to report any power results (Peterman 1990). A separate survey, reviewing toxicology literature, found high power in only 19 out of 668 reports that failed to reject the null hypothesis (Hayes 1987). Similar surveys in other fields such as psychology or education have turned up "depressingly low" levels of power (Brewer 1972; Cohen 1988).

3. Minimum detectable change

Another sampling design concept that is directly related to statistical power and false-change error rates is the size of the change that you want to be able to detect. This will be referred to as the *minimum detectable change or MD*C. The MDC greatly influences power levels. A particular sampling design will be more likely to detect a true large change (i.e., with high power) than to detect a true small change (i.e., with low power).

Setting the MDC requires the consideration of ecological information for the species being monitored. How large of a change should be considered biologically meaningful? With a large enough sample size, statistically significant changes can be detected for changes that have no biological significance. For example, if an intensive monitoring design leads to the conclusion that the mean density of a plant population increased from 10.0 plants/m^2 to 10.1 plants/m^2, does this represent some biologically meaningful change in population density? Probably not.

Setting a reasonable MDC can be difficult when little is known about the natural history of a particular plant species. Should a 30% change in the mean density of a rare plant population be cause for alarm? What about a 20% change or a 10% change? The MDC considerations are likely to vary when assessing vegetation attributes other than density, such as cover or frequency (Chapter 8). The initial MDC, set during the design of a new monitoring study, can be modified once monitoring information demonstrates the size and rate of population fluctuations.

4. How to achieve high statistical power

Statistical power is related to four separate sampling design components by the following function equation:

Power = a function of (s, n, MDC, and α)

where: s = standard deviation
n = number of sampling units
MDC = minimum detectable change
α = false-change error rate

Power can be increased in the following four ways:

1. **Reducing standard deviation**. This means altering the sampling design to reduce the amount of variation among sampling units (see Chapter 7).

2. **Increasing the number of sampling units sampled**. This method of increasing power is straightforward, but keep in mind that increasing *n* has less of an effect than decreasing *s* since the square root of sample size is used in the standard error equation ($SE = s/\sqrt{n}$).

3. **Increasing the acceptable level of false-change errors (α)**.

4. **Increasing the MDC**.

Note that the first two ways of increasing power are related to making changes in the sampling design, whereas the other two ways are related to making changes in the sampling objective (see Chapter 6).

5. Graphical comparisons

As stated, power is driven by four different factors: standard deviation, sample size, minimum detectable change size, and false-change error rate. In this section we take a graphical look at how altering these factors changes power. The comparisons in this section are based upon sampling a fictitious plant population where we are interested in assessing plant density relative to an established threshold value of 25 plants/m^2. Any true population densities less than 25 plants/m^2 will trigger management action. We are only concerned with the question of whether the density is lower than 25 plants/m^2 and not whether the density is higher. In this example, our null hypothesis (H_O) is that the population density equals 25 plants/m^2 and our alternative hypothesis is that density is less than 25 plants/m^2. The density value of 25 plants/m^2 is the most critical single density value since it defines the lower limit of acceptable plant density.

The figures in this section are all based upon sampling distributions where we happen to know the true plant density. Recall that a sampling distribution is a bell-shaped curve that depicts the distribution of a large number of independently gathered sample statistics. A sampling distribution defines the range and relative probability of any possible sample mean. You are more likely to obtain sample means near the middle of the distribution than you are to obtain sample means near either tail of the distribution.

A sampling distribution based on sampling our fictitious population with a true mean density of 25 plants/m^2 is shown in Figure 5.8A. This distribution is based on a sampling design using thirty 1m x 1m quadrats where the true standard deviation is ± 20 plants/quadrat. If 1,000 different people randomly sample and calculate a sample mean based upon their 30 quadrat values, approximately half the individually drawn sample means will be less than 25 plants/m^2 and half will be greater than 25 plants/m^2. Approximately 40% of the samples will yield sample means less than or equal to 24 plants/m^2. A few of our 1,000 individuals will obtain estimates of the mean density that deviate from the true value by a large margin. One of the individuals will likely stand up and say, "my estimate of the mean density is 13 plants/m^2", even though the true density is actually 25 plants/m^2. As interpreters of the monitoring information, we would conclude that since 999 of the 1,000 people obtained estimates of the density that were greater than 13, the true density is probably not 13. Our best estimate of the true mean density will be the average of the 1,000 separate estimates (this average is likely to be extremely close to the actual true value).

Now that we have the benefit of 1,000 independent estimates of the true mean density, we can return to the population at a later time, take a single random sample of thirty 1m x 1m quadrats, calculate the sample mean, and then ask the question, "what is the probability of obtaining our sample mean value if the true population is still 25 plants/m^2?" If our sample mean density turns out to be 24 plants/m^2, would this lead to the conclusion that the population has crossed our threshold value? Seeing that our sample mean is lower than our target value might raise some concerns, but we have no objective basis to conclude that the true population is not, in fact, still actually 25 plants/m^2. We learned in the previous paragraph that a full 40% of possible samples are likely to yield mean densities of 24 plants/m^2 or less if the true mean is 25 plants/m^2. Thus, the probability of obtaining a single sample mean of 24 plants/m^2 or less when the true density is actually 25 plants/m^2 is approximately 0.40. Obtaining a sample mean of 24 plants/m^2 is consistent with the hypothesis that the true population density is actually 25 plants/m^2.

How small a sample mean do we need to obtain to feel confident that the population has indeed dropped below 25 plants/m^2? What will our interpretation be if we obtained a sample mean of 22 plants/m^2? Based upon our sampling distribution from the 1,000 people, the probability of obtaining an estimate of 22 plants/m^2 or less is around 20%, which represents a 1-in-5 chance that the true mean is still actually 25 plants/m^2. Based upon the sampling distribution from our 1,000 separate samplers, we can look at the likelihood of obtaining other different sample means. The probability of obtaining a sample of 20 plants/m^2 is 8.5%,

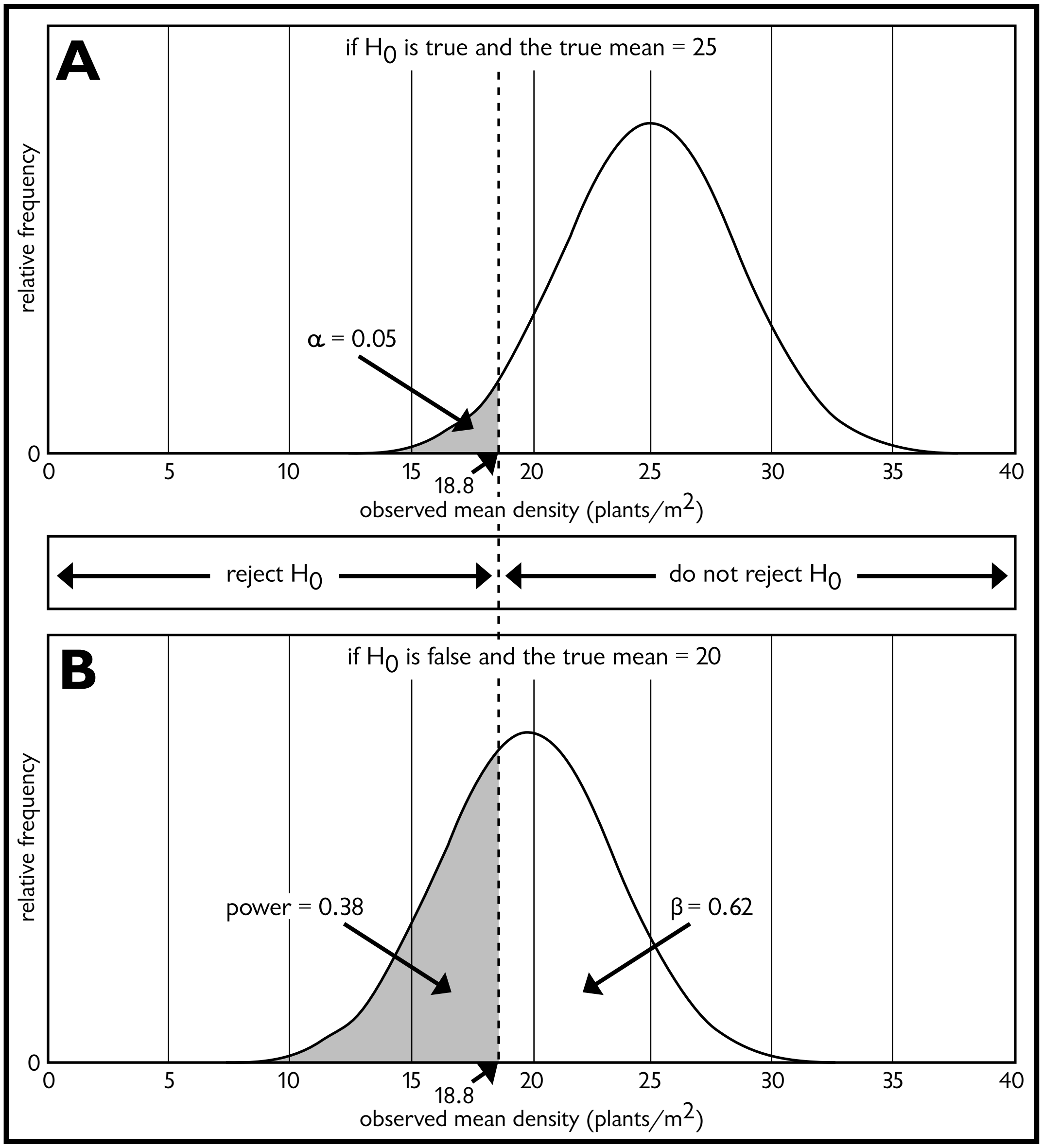


FIGURE 5.8. Example of sampling distributions for mean plant density in samples of 30 permanent quadrats where the among-quadrat standard deviation is 20 plants/m². Part A is the sampling distribution for the case in which the null hypothesis, H_0, is true and the true population mean density is 25 plants/m². The shaded area in part A is the critical region for $\alpha = 0.05$ and the vertical dashed line is at the critical sample mean value, 18.8. Part B is the sampling distribution for the case in which the H_0 is false and the true mean is 20 plants/m². In both distributions, a sample mean to the left of the vertical dashed line would reject H_0, and to the right of it, would not reject H_0. Power and β values in part B, in which H_0 is false and the true mean = 20, are the proportion of sample means that would occur in the region in which H_0 was rejected or not rejected, respectively (adapted from Peterman 1990).

and the probability of obtaining a sample of 18 plants/m² is 2.9% if the true mean density is 25 plants/m².

Since in most circumstances we will only have the results from a single sample (and not the benefit of 1,000 independently gathered sample means), another technique must be used to determine whether the population density has dropped below 25 plants/m². One method is to run a statistical test that compares our sample mean to our density threshold value (25 plants/m^2). The statistical test will yield a *P*-value that defines the probability of obtaining our sample mean if the true population density is actually 25 plants/m^2. As interpreters of our monitoring information, we will need to set some probability threshold *P*-value to guide our interpretation of the results from the statistical test. This *P*-value threshold defines our acceptable false-change error rate. If we run a statistical test that compares our sample mean to our density threshold value (25 plants/m^2) and the *P*-value from the test is lower than our threshold value, then we conclude that the population density has, in fact, declined below 25 plants/m^2. Thus, if we set our *P*-value threshold to 0.05 and the statistical test yields a *P*-value of 0.40, then we *fail to reject* the null hypothesis that the true population density is 25 plants/m^2. If, however, the statistical test yields a *P*-value of 0.022, this is lower than our threshold *P*-value of 0.05, and we would reject the null hypothesis that the population is 25 plants/m^2 in favor of our alternative hypothesis that the density is lower than 25 plants/m^2.

The relationship between the *P*-value threshold of 0.05 and our sampling distribution based upon sampling thirty 1m x 1m quadrats is shown in Figure 5.8A. The threshold density value corresponding to our *P*-value threshold of 0.05 is 18.8 plants/m^2, which is indicated on the sampling distribution by the dashed vertical line. Thus, if we obtain a mean density of 18 plants/m^2, which is to the left of the vertical line, we reject the null hypothesis that the population density is 25 plants/m^2 in favor of an alternative hypothesis that density is lower than 25 plants/m^2. If we obtain a mean density of 21 plants/m^2, which is to the right of the vertical line, then we fail to reject the null hypothesis that the population density is really 25 plants/m^2.

So far we have been discussing the situation where the true population density is right at the threshold density of 25 plants/m^2. Let's look now at a situation where we know the true density has declined to 20 plants/m^2. What is the likelihood of our detecting this true density difference of 5 plants/m^2? Figure 5.8B shows a new sampling distribution based upon the true density of 20 plants/m^2 (standard deviation is still ± 20 plants/m^2). We know from our previous discussion that sample means to the right of the vertical line in Figure 5.8A lead to the conclusion that we can't reject the null hypothesis that our density is 25 plants/m^2. If our new sample mean turns out to exactly match the new true population mean (i.e., 20 plants/m^2), will we reject the idea that the sample actually came from a population with a true mean of 25 plants/m^2? No, at least not at our stated *P*-value (false-change error) threshold of 0.05. A sample mean value of 20 plants/m^2 falls to the right of our dashed threshold line in the "do not reject H_O" portion of the graph and we would have failed to detect the true difference that actually occurred. Thus, we would have committed a missed-change error.

What is the probability of missing the true difference of five plants/m^2 show in Figure 5.8B? This probability represents the missed-change error rate (β) and it is defined by the nonshaded area under the sampling distribution in Figure 5.8B, which represents 62% of the possible sample mean values. Recall that the area under the whole curve defines the entire range of possible values that you could obtain by sampling the population with the true mean = 20 plants/m^2. If we bring back our 1,000 sampling people and have each of them sample thirty 1m x 1m quadrats in our new population, we will find that approximately 620 of them will obtain estimates of the mean density that are greater than the 18.8 plants/m^2 threshold value that is shown by the vertical dashed line. What about the other 380 people? They will obtain population estimates fewer than the critical threshold of 18.8 plants/m^2 and they will reject the null hypothesis that the population equals 25 plants per quadrat. This proportion of 0.38 (380 people out of 1,000 people sampling) represents the *statistical power* of our sampling design and it is represented by the shaded area under the curve in Figure 5.8B. If the true population mean is indeed 20 plants/m^2 instead of 25 plants/m^2, then we can be 38%

(power = 0.38) sure that we will detect this true difference of five plants/m². With this particular sampling design (thirty 1m x 1m quadrats) and a false-change error rate of α=0.05, we run a 62% chance (β=0.62) that we will commit a missed-change error (i.e., fail to detect the true difference of five plants/m²). If the difference of five plants/m² is biologically important, a power of only 0.38 would not be satisfactory.

We can improve the low-power situation in four different ways: (1) increase the acceptable false-change error rate; (2) increase the acceptable MDC; (3) increase sample size; or (4) decrease the standard deviation. New paired sampling distributions illustrate the influence of making each of these changes.

a. Increasing the acceptable false-change error rate

In Figure 5.8B, a false-change error rate of α=0.05 resulted in a missed-change error rate of β=0.62 to detect a difference of five plants/m². Given these error rates, we are more than 12 times more likely to commit a missed-change error than we are to commit a false-change error. What happens to our missed-change error rate if we specify a new, higher false-change error rate? Shifting our false-change error rate from α=0.05 to α=0.10 is illustrated in Figure 5.9

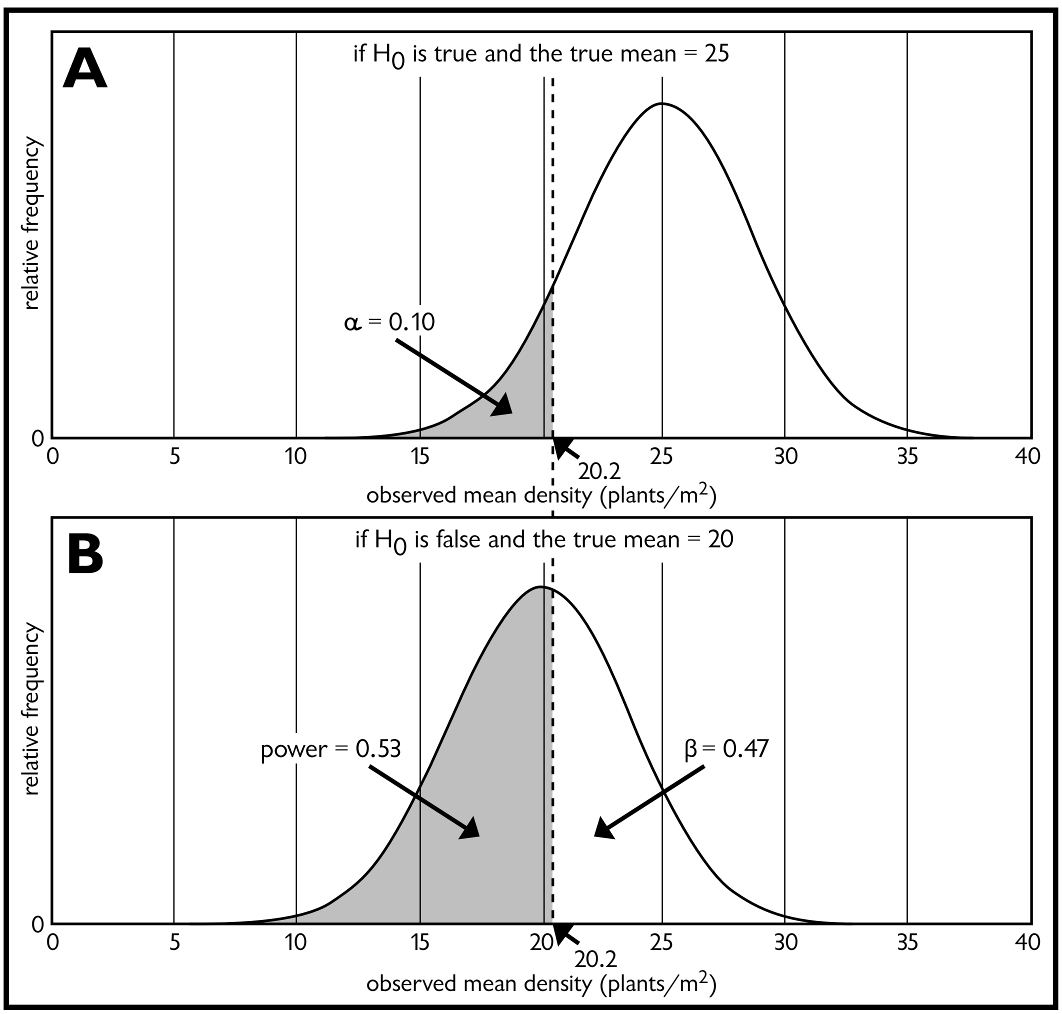


FIGURE 5.9. The critical region for the false-change error in the sampling distributions from Figure 5.8 has been increased from α = 0.05 to α = 0.10. Part B, in which the H_0 is false and the true mean = 20, shows that power is larger for α = 0.10 than for Figure 5.8 where α = 0.05 (adapted from Peterman 1990).

for the same sampling distributions shown in Figure 5.8. Our critical density threshold at the *P*=0.10 level is now 20.21 plants/m^2, and our missed-change error rate has dropped from β=0.62 down to β=0.47 (i.e., the power to detect a true five plant/m^2 difference went from 0.38 to 0.53). A sample mean of 20 plants/m^2 will now lead to the correct conclusion that a difference of five plants/m^2 between the populations does exist. Of course, the penalty we pay for increasing our false-change error rate is that we are now twice as likely to conclude that a difference exists in situations when there is no true difference and our population mean is actually 25 plants/m^2. Changing the false-change error rate even more, to α=0.20, (Figure 5.10) reduces the probability of making a missed-change error down to β=0.29 (i.e., giving us a power of 0.71 to detect a true difference of five plants/m^2).

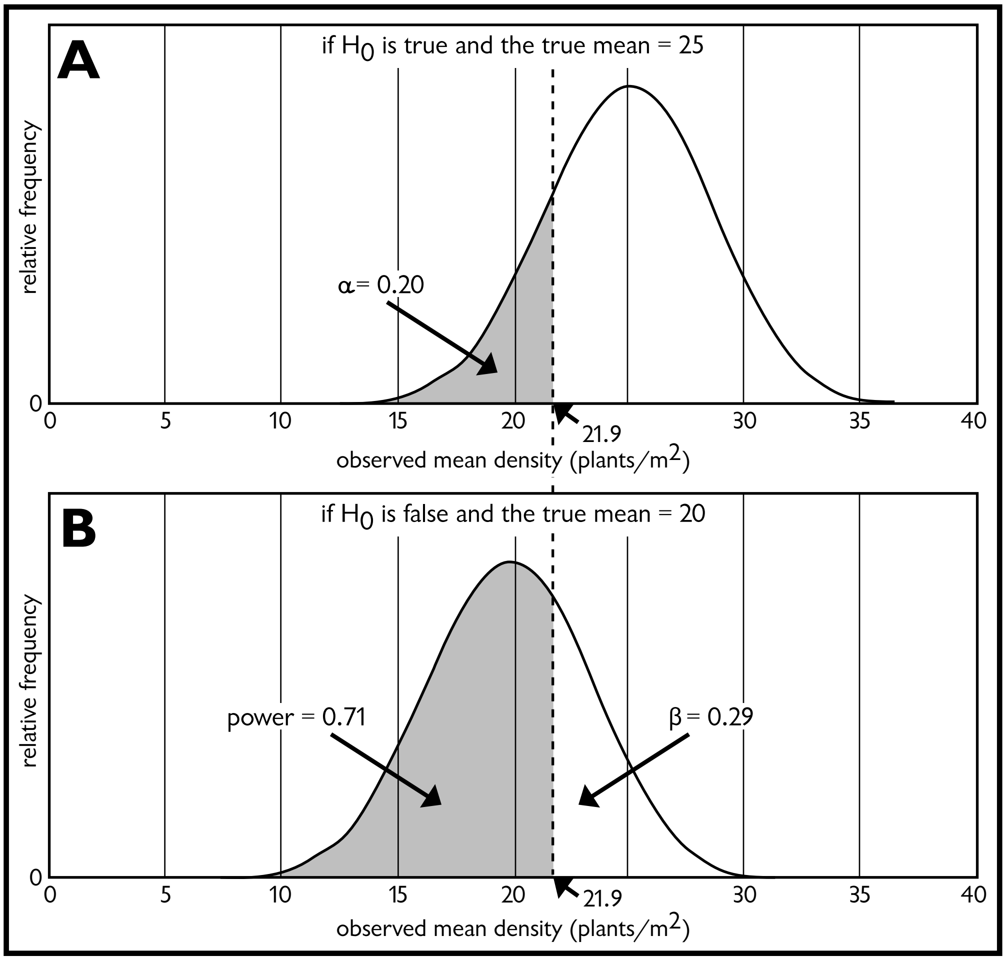


FIGURE 5.10. The critical region for the false-change error in the sampling distributions from Figure 5.8 has been increased from α = 0.05 to α = 0.20. Part B, in which the H_0 is false and the true mean = 20, shows that power is larger for α = 0.20 than for Figure 5.8 where α = 0.05 or Figure 5.9 where α = 0.10. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

b. Increasing the acceptable MDC

Any sampling design is more likely to detect a true large difference than a true small difference. As the magnitude of the difference increases, we will see an increase in the power to detect the difference. This relationship is shown in Figure 5.11B, where we see a sampling distribution with a true mean density of 15 plants/m^2, which is 10 plants/m^2 below our threshold density of 25 plants/m^2. The false-change error rate is set at α=0.05 in this example.

This figure shows that the statistical power to detect this larger difference of 10 plants/m^2 (25 plants/m^2 to 15 plants/m^2) is 0.85 compared with the original power value of 0.38 to detect the difference of five plants/m^2 (25 plants/m^2 to 20 plants/m^2). Thus, with a false-change error rate of 0.05, we can be 85% certain of detecting a difference from our 25 plants/m^2 threshold of 10 plants/m^2 or greater. If we raised our false-change error from α=0.05 to α=0.10 (not shown in Figure 5.11), our power value would rise to 0.92, which creates a sampling situation where our two error rates are nearly equal (α=0.10, β=0.08).

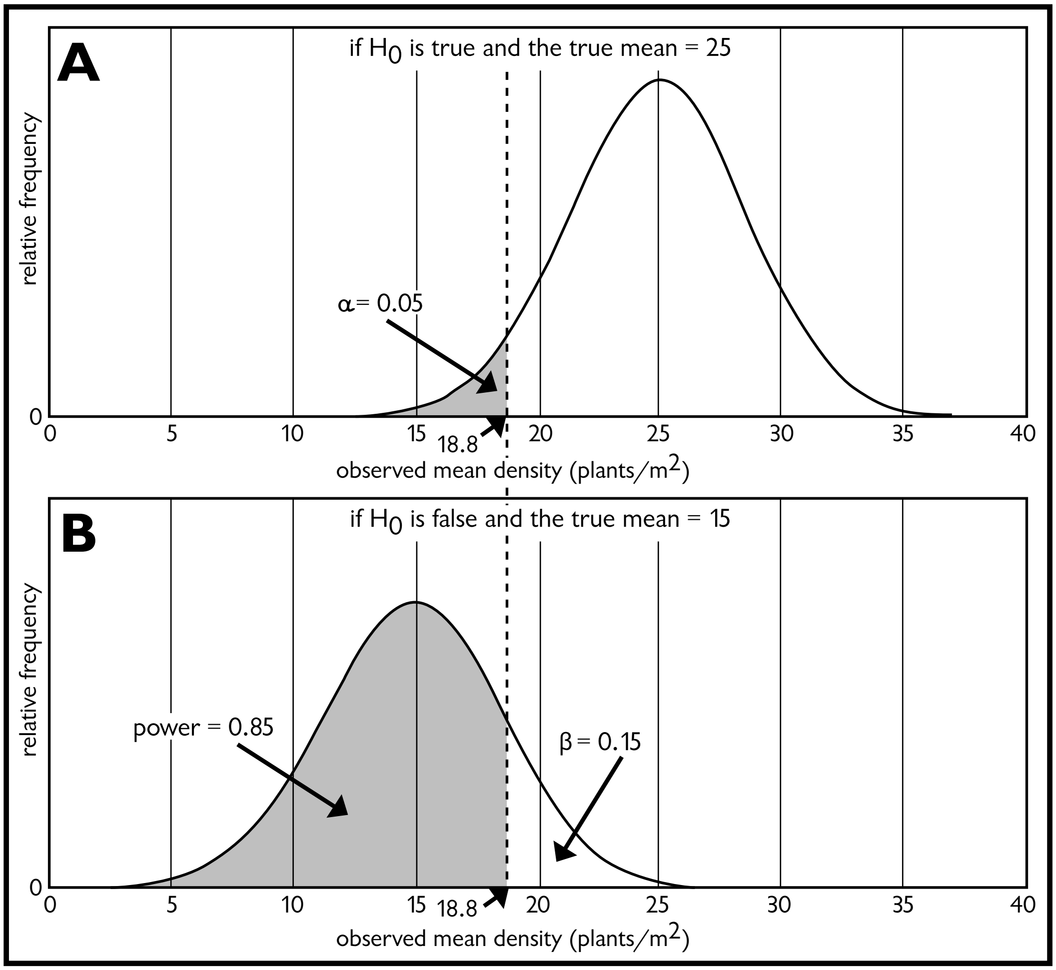


FIGURE 5.11. Part A is the same as Figure 5.8; in part B, the true population mean is 15 plants/m^2 instead of the 20 plants/m^2 shown in Figure 5.8. Note that power increases (and β decreases) when the new true population mean gets further from the original true mean of 25 plants/m^2. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

c. Increasing the sample size

The sampling distributions shown in Figures 5.8 to 5.11 were all created by sampling the populations with n=30 1m x 1m quadrats. Any increase in sample size will lead to a subsequent increase in power to detect some specified minimum detectable difference. This increase in power results from the sampling distributions becoming narrower. Sampling distributions based on samples of n=50 are shown in Figure 5.12 where the true difference between the two populations is once again five plants/m^2 with a false-change error rate threshold of α=0.05. The increase in sample size led to an increase in power from power=0.38 with n=30 to power=0.54 with n=50. Note that the critical threshold density associated with an α = 0.05 is now 20.3 plants/m^2 as compared to the 18.8 plants/m^2 threshold when n = 30.

d. Decreasing the standard deviation

The sampling distributions shown in Figures 5.8 to 5.12 all are based on sampling distributions with a standard deviation of ±20 plants/m^2. The quadrat size used in the sampling was a square 1m x 1m quadrat. If individuals in the plant population are clumped in distribution, then it is likely that a rectangular shaped quadrat will result in a lower standard deviation (See Chapter 7 for a detailed description of the relationship between standard deviation and sampling unit size and shape). Figure 5.13 shows sampling distributions where the standard deviation was reduced from ±20 plants/m^2 to ±10 plants/m^2. Note that the critical threshold density associated with an α = 0.05 is now 21.9 plants/m^2 as compared to the 18.8 plants/m^2 threshold when the standard deviation was ± 20 plants/m^2. This reduction in the true standard deviation came from a change in quadrat shape from the 1m x 1m square shape to a 0.2m x 5m rectangular shape. Note that quadrat area (1m^2) stayed the same so that the mean densities are consistent with the previous sampling distributions shown in Figures 5.8

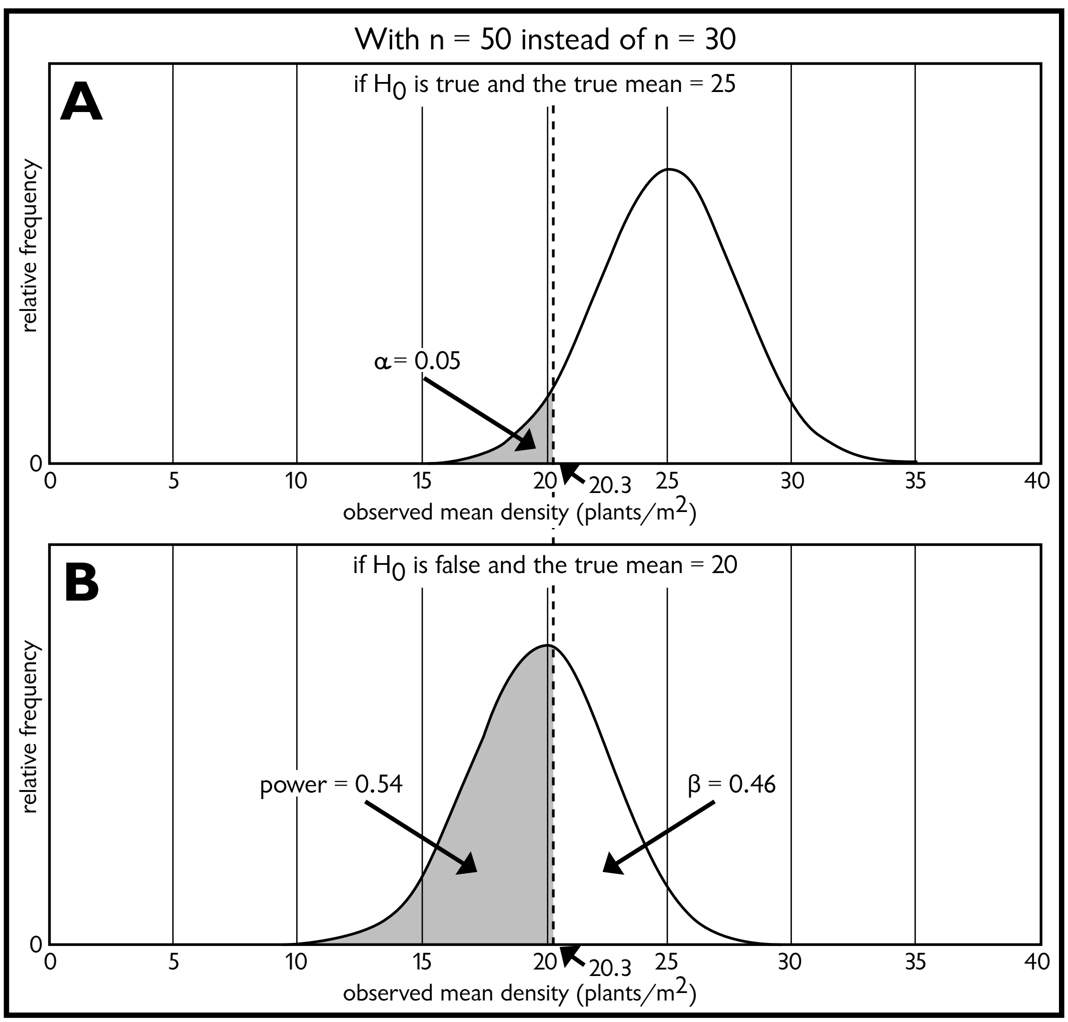


FIGURE 5.12. The sample size was increased to n = 50 quadrats from the n = 30 quadrats shown in Figure 5.8. Note that power increases (and β decreases) at larger sample sizes. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

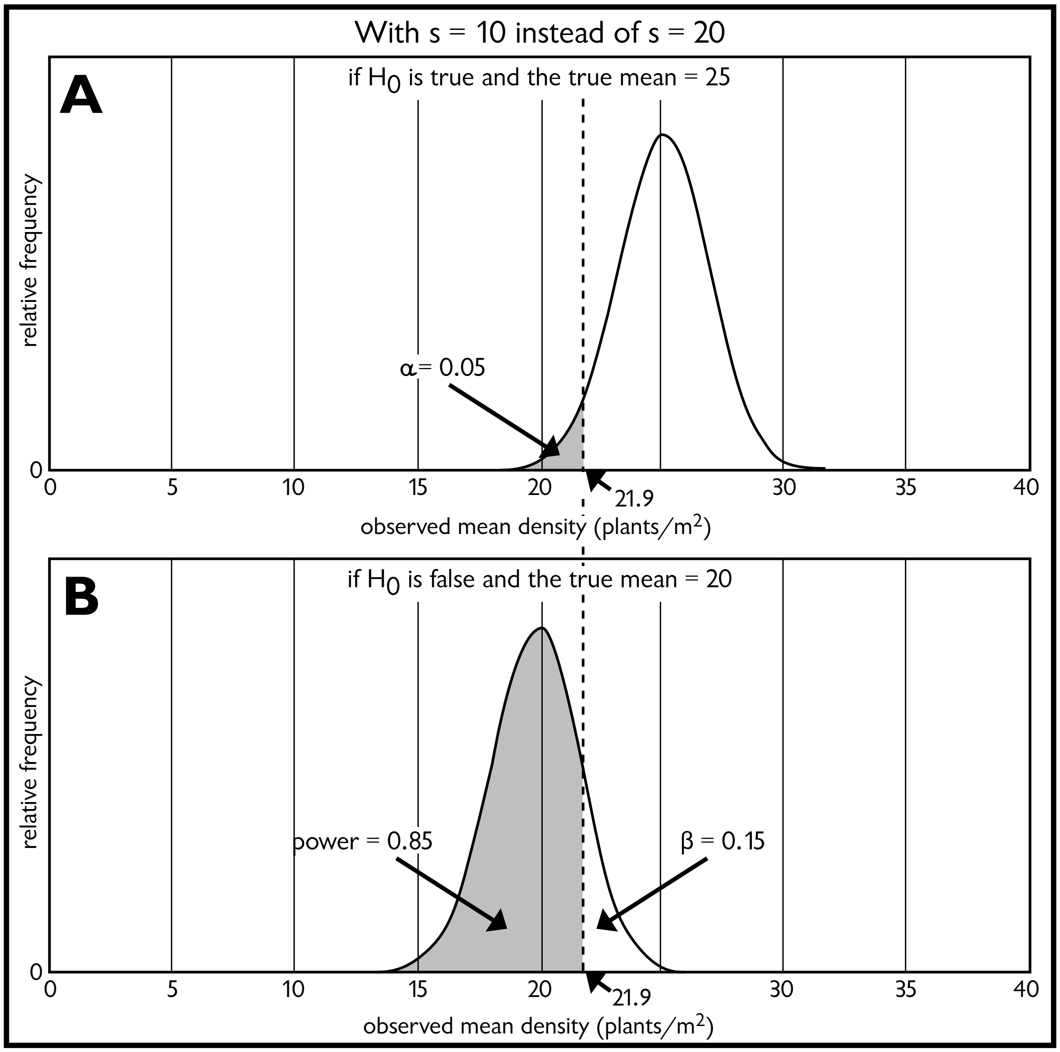


FIGURE 5.13. The standard deviation (s) of 20 plants/m² shown in Figure 5.8 is reduced to ten plants/m². Note that power increases (and β decreases), as the standard deviation decreases. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

through 5.12. This reduction in standard deviation led to a dramatic improvement in power, from 0.38 (with sd = 20 plants/m²) to 0.85 (with sd = 10 plants/m²). Reducing the standard deviation has a more direct impact on increasing power than increasing sample size, because the sample size is reduced by taking its square root in the standard error equation ($SE = s/\sqrt{n}$). Recall that the standard error provides an estimate of sampling precision from a single sample without having to enlist the support of 1,000 people who gather 1,000 independent sample means.

e. Power curves

The relationship between power and the different sampling design components that influence power can also be displayed in power curve graphs. These graphs typically show power values on the y-axis and either sample size, MDC, or standard deviation values on the x-axis. Figure 5.14A shows statistical power graphed against different magnitudes of change for the same hypothetical data set described above and shown in Figures 5.8 to 5.11. Four different power curve lines are shown, one for each of the following four different false-change (α) error rates: 0.01, 0.05, 0.10, and 0.20. The power curves are based on sampling with a sample size of 30 quadrats and a standard deviation of 20 plants/m^2. For any particular false-change error rate, power increases as the magnitude of the minimum detectable change increases. When α=0.05, the power to detect small changes is very low (Figure 5.14A). For example, we have only a 13% chance of detecting a difference of 2 plants/m^2 (i.e., a density of 23 plants/m^2 which is 2 plants/m^2 below our threshold value of 25 plants/m^2). In contrast, we can be 90% sure of detecting a minimum difference of 11 plants/m^2. We can also attain higher power by increasing

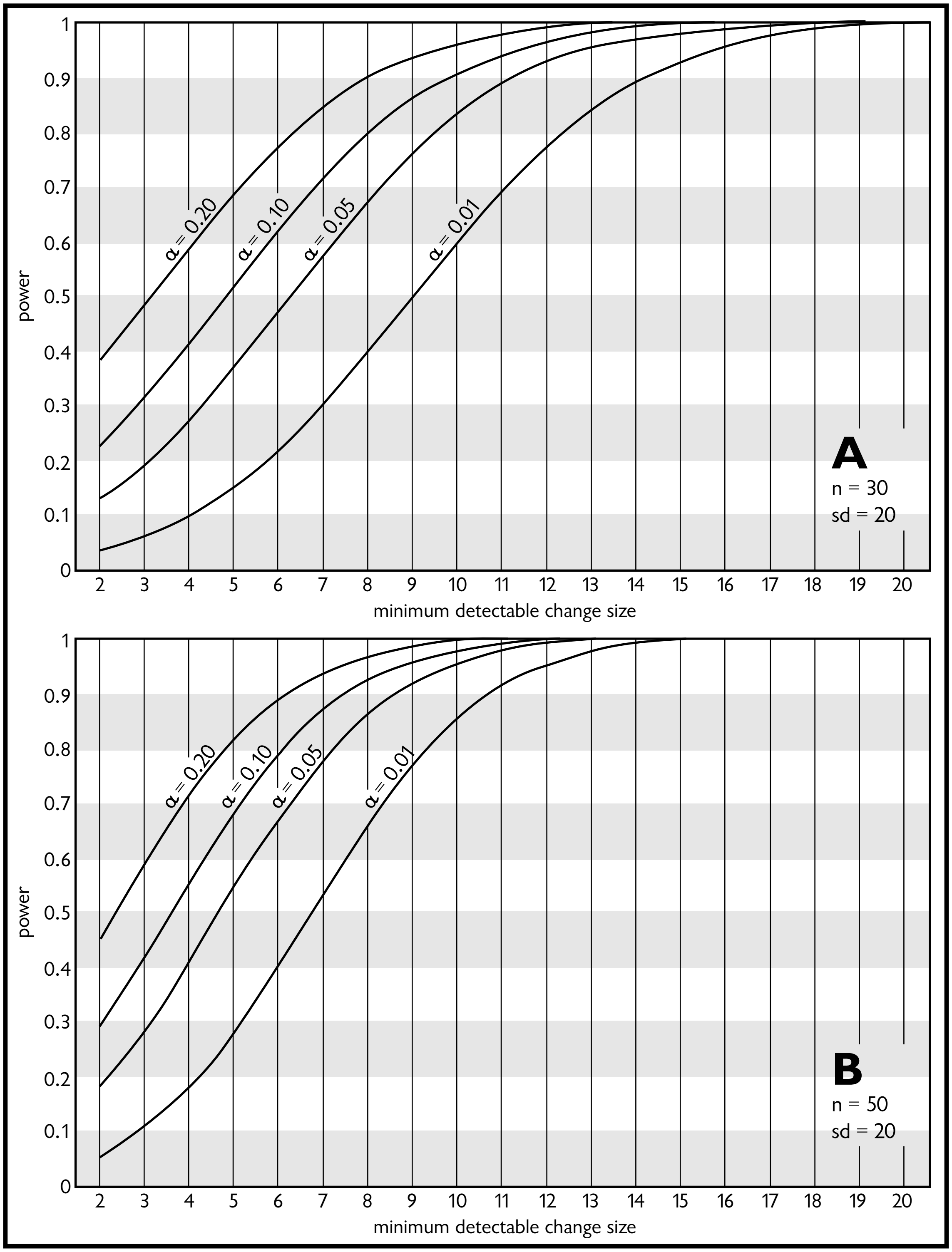


FIGURE 5.14. Power curves showing power values for various magnitudes of minimum detectable change and false-change error rates when the standard deviation is 20. Part A shows power curves with a sample size of 30. Part B shows power curves with a sample size of 50.

the false-change error rate. The power to detect a change of eight plants/m^2 is only 0.41 when α=0.01, but it increases to 0.69 at α=0.05, to 0.81 at α=0.10, and to 0.91 at α=0.20.

A different set of power curves are shown in Figure 5.14B where the sample size is n=50 instead of the n=30 shown in Figure 5.14A. This larger smaller sample size shifts all of the power curves to the left, making it more likely that smaller changes will be detected. For example, with a false-change error rate of α=0.10, the power to detect a seven plant/m^2 difference is 0.88 with a sample size of n=50 quadrats compared to the power of 0.73 with a sample size of n=30 quadrats.

Power curves that show the effect of reducing the standard deviation are shown in Figure 5.15. Figure 5.15A is the same as Figure 5.14A where the standard deviation is 20 plants/m^2. Figure 5.15B shows the same power curves except they are based on a standard deviation of 10 plants/m^2. The smaller standard deviation shifts all of the power curves to the left and results in much steeper slopes. The smaller standard deviation leads to substantially higher power levels for any particular MDC value. For example, the power to detect a change of five plants/m^2 with a false change error rate of α=0.10 is only 0.53 in Figure 5.15A as compared to the power of 0.92 in Figure 5.15B.

6. Setting false-change and missed-change error rates

Both false-change and missed-change error rates can be reduced by sampling design changes that increase sample size or decrease sample standard deviations. Missed-change and false-change error rates are inversely related, which means that reducing one will increase the other (but not proportionately) if no other changes are made. The decision of which type of error is more important should be based on the nature of the changes you are trying to determine, and the consequences of making either kind of mistake.

Because false-change and missed-change error rates are inversely related to each other, and because these errors have different consequences to different interest groups, there are different opinions as to what the "acceptable" error rates should be. The following examples demonstrate the conflict between false-change and missed-change errors.

- **Testing for a lethal disease**. When screening a patient for some disease that is lethal without treatment, a physician is less concerned about making a false diagnosis error (analogous to a false-change error) of concluding that the person has the disease when they do not than failing to detect the disease (analogous to a missed-change error) and concluding that the person does not have the disease when in fact they do.

- **Testing for guilt in our judicial system**. In the United States, the null hypothesis is that the accused person is innocent. Different standards for making judgement errors are used depending upon whether the case is a criminal or a civil case. In criminal cases, proof must be "beyond a reasonable doubt." In these situations it is less likely that an innocent person will be convicted (analogous to a false-change error), but it is more likely that a guilty person will go free (analogous to a missed-change error). In civil cases, proof only needs to be "on the balance of probabilities." In these situations, there is a greater likelihood of making a false conviction (analogous to a false-change error), but a lower likelihood of making a missed conviction (analogous to a missed-change) error.

- **Testing for pollution problems**. In pollution monitoring situations, the industry has an interest in minimizing false-change errors and may desire a very low false-change error

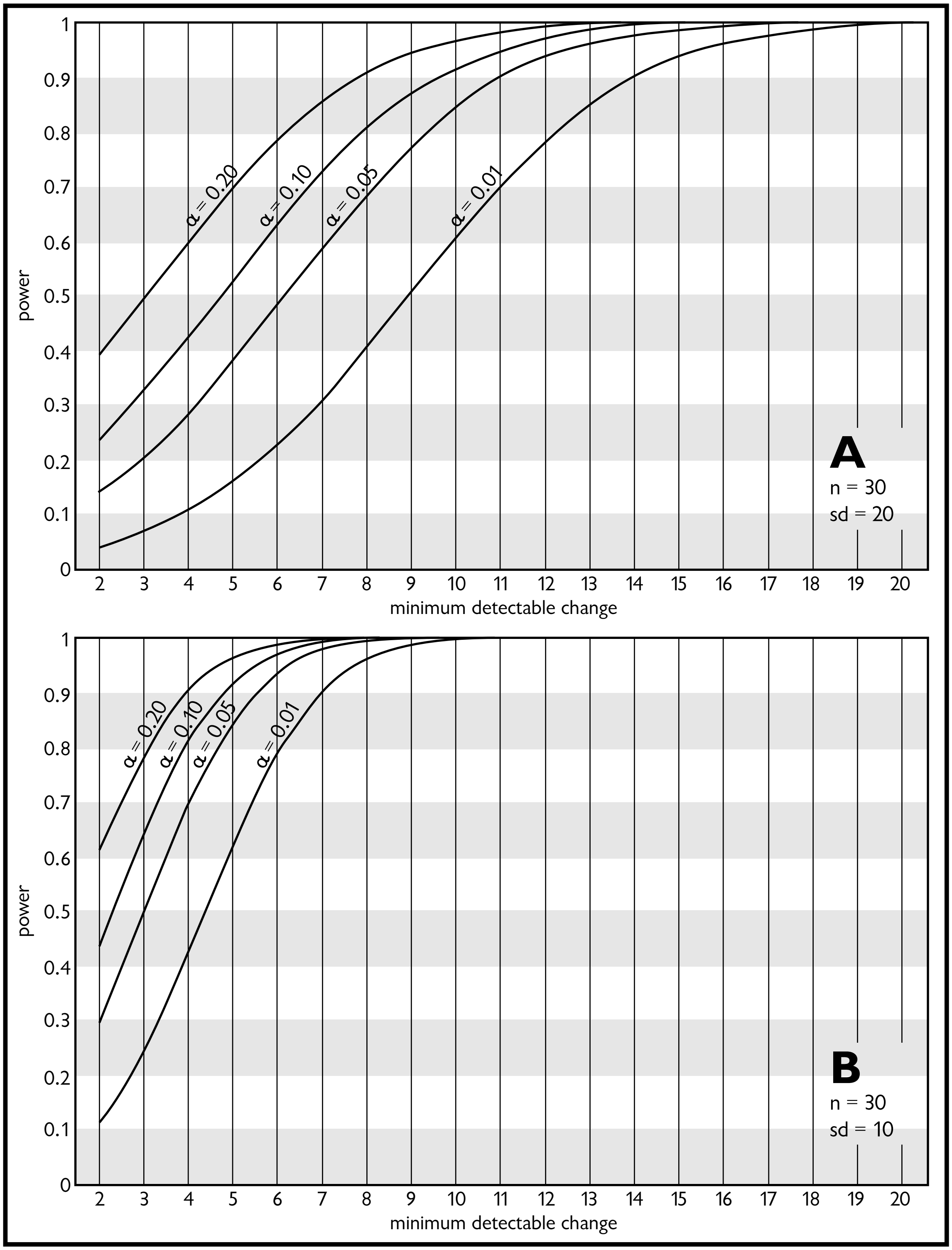


FIGURE 5.15. Power curves showing power values for various magnitudes of minimum detectable change and false-change error rates when the sample size = 30. Part A shows power curves with a standard deviation of 20 plants/m². Part B shows power curves with a standard deviation of 10 plants/m².

rate (e.g., α = 0.01 or 0.001). Companies do not want to be shut down or implement expensive pollution control procedures if a real impact has not occurred. In contrast, an organization concerned with the environmental impacts of some pollution activity will likely want to have high power (low missed-change error rate) so that they do not miss any real changes that take place. They may not be as concerned about occasional false-change errors (which would result in additional pollution control efforts even though real changes did not take place).

Missed-change errors may be as costly or more costly than false-change errors in environmental monitoring studies (Toft and Shea 1983; Peterman 1990; Fairweather 1991). A false-change error may lead to the commitment of more time, energy, and people, but probably only for the short period of time until the mistake is discovered (Simberloff 1990). In contrast, a missed-change error, as a result of a poor study design, may lead to a false sense of security until the extent of the damages are so extreme that they show up in spite of a poor study design (Fairweather 1991). In this case, rectifying the situation and returning the system to its preimpact condition could be costly. For this reason, you may want to set equal false-change and missed-change error rates or even consider setting the missed-change error rate lower than the false-change error rate (Peterman 1990; Fairweather 1991).

There are many historical examples of costly missed-change errors in environmental monitoring. For example, many fish population monitoring studies have had low power to detect biologically meaningful declines so that declines were not detected until it was too late and entire populations crashed (Peterman 1990). Some authors advocate the use of something they call the "precautionary principle" (Peterman and M'Gonigle 1992). They argue that, in situations where there is low power to detect biologically meaningful declines in some environmental parameter, management actions should be prescribed as if the parameter had actually declined. Similarly, some authors prefer to shift the burden of proof in situations where there might be an environmental impact from environmental protection interests to industry/development interests (Peterman 1990; Fairweather 1991). They argue that a conservative management strategy of "assume the worst until proven otherwise" should be adopted. Under this strategy, developments that may negatively impact the environment should not proceed until the proponents can demonstrate, with high power, a lack of impact on the environment.

7. Why has statistical power been ignored for so long?

It is not clear why missed-change errors, power, and minimum detectable change size have traditionally been ignored. Perhaps researchers have not been sufficiently exposed to the idea of missed-change errors. Most introductory texts and statistics courses deal with the material only briefly. Computer packages for power analysis have only recently become available. Perhaps people have not realized the potentially high costs associated with making missed-change errors. Perhaps researchers have not understood how understanding power can improve their work.

The issue of power and missed-change errors has gained a lot of attention in recent years. A literature review in the 1980's would not have turned up many articles dealing with statistical power issues. A literature review today would turn up dozens of articles in many disciplines from journals all over the world (see Peterman 1990 and Fairweather 1991 for good review papers on statistical power). Journal editors may soon start requiring that power analysis information be reported for all nonsignificant results (Peterman 1990). There may also be some departure from the strict adherence to the 0.05 significance level (Peterman 1990; Fairweather 1991).

8. Use of prior power analysis during study design

Power analysis can be useful during both the design of monitoring studies and in the interpretation of monitoring results. The former is sometimes called "prior power analysis," whereas the latter is sometimes called "post-hoc power analysis" (Fairweather 1991). Post-hoc power analysis is covered in Chapter 11.

The use of power analysis during the design and planning of monitoring studies provides valuable information that can help avoid monitoring failures. Once some preliminary or pilot data have been gathered, or if some previous years' monitoring data are available, power analysis can be used to evaluate the adequacy of the sampling design. Prior power analysis can be done in several different ways. All are based upon the power function described earlier:

Power = a function of (s, n, MDC, and α)

The power of a particular sampling design can be evaluated by plugging sample standard deviation, sample size, the desired MDC, and an acceptable false-change error rate, into equations or computer programs (Appendix 16) and solving for power. If the power to detect a biologically important change turns out to be quite low (high probability of a missed-change error), then the sampling design can be modified to try to achieve higher power.

Alternatively, a desired power level can be specified and the terms in the power function can be rearranged to solve for sample size. This will give you assurance that your study design will succeed in being able to detect a certain magnitude of change at the specified power and false-change error rate. This is the format for the sample size equations that are discussed in Chapter 7 and presented in Appendix 7.

Still another way to do prior power analysis is to specify a desired power level and a particular sample size and then rearrange the terms in the power function to solve for the MDC (Rotenberry and Wiens 1985; Cohen 1988). If the MDC is unacceptably large, then attempts should be made to improve the sampling design. If these efforts fail, then the decision must be made to either live with the large MDC or to reject the sampling design and perhaps consider an alternative monitoring approach.

The main advantage of prior power analysis is that it allows the adequacy of the sampling design to be evaluated at an early stage in the monitoring process. It is much better to learn that a particular design has a low power at a time when modifications can easily be made than it is to learn of low power after many years of data have already been gathered. The importance of specifying acceptable levels of false-change and missed-change errors along with the magnitude of change that you want to be able to detect is covered in Chapter 6 the next chapter, which introduces sampling objectives.

Literature Cited

Brewer, J. K. 1972. On the power of statistical tests in the American Educational Research Journal. American Educational Research Journal 9: 391-401.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, N. J. Lawrence Erlbaum Associates.

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring. Australian Journal of Marine and Freshwater Research 42: 555-567.

Hayes, J. P. 1987. The positive approach to negative results in toxicology studies. Ecotoxicology and Environmental Safety 14: 73-77.

McCall, C. H. 1982. Sampling and statistics handbook for research. Ames, IA: The Iowa State University Press.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences. 47: 2-15.

Peterman, R. M.; M'Gognigle. M. 1992. Statistical power analysis and the precautionary principle. Marine Pollution Bulletin. 24(5): 231-234.

Rotenberry, J. T.; Wiens J. A. 1985. Statistical power analysis and community-wide patterns. American Naturalist. 125: 164-168.

Simberloff, D. 1990. Hypotheses, errors, and statistical assumptions. Herpetelogica. 46: 351-357.

Toft, C. A.; P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist. 122: 618-625.

Zar, J. H. 1996. Biostatistical Analysis. Englewood Cliffs, N.J. Jersey: Prentice-Hall, Inc.

CHAPTER 5. Basic Principles of Sampling

A. Introduction

What is sampling? A review of several dictionary definitions led to the following composite definition:

> The act or process of selecting a part of something with the intent of showing the quality, style, or nature of the whole.

Monitoring does not always involve sampling techniques. Sometimes you can count or measure all individuals within a population of interest. Other times you may select qualitative techniques that are not intended to show the quality, style, or nature of the whole population (e.g., subjectively positioned photoplots).

What about those situations where you have an interest in learning something about the entire population, but where counting or measuring all individuals is not practical? This situation calls for sampling. The role of sampling is to provide information about the population in such a way that inferences about the total population can be made. This inference is the process of generalizing to the population from the sample, usually with the inclusion of some measure of the "goodness" of the generalization (McCall 1982).

Sampling will not only reduce the amount of work and cost associated with characterizing a population, but sampling can also increase the accuracy of the data gathered. Some kinds of errors are inherent in all data collection procedures, and by focusing on a smaller fraction of the population, more attention can be directed toward improving the accuracy of the data collected.

This chapter includes information on basic principles of sampling. Commonly used sampling terminology is defined and the principal concepts of sampling are described and illustrated. Even though the examples used in this chapter are based on counts of plants in quadrats (density measurements), most of the concepts apply to all kinds of sampling.

B. Populations and Samples

The term "population" has both a biological definition and a statistical definition. In this chapter and in Chapter 7, we will be using the term "population" to refer to the statistical population or the "sampling universe" in which monitoring takes place. The statistical population will sometimes include the entire biological population, and other times, some portion of the biological population. The *population* consists of the complete set of individual objects about which you want to make inferences. We will refer to these individual objects as *sampling units*. The sampling units can be individual plants or they may be quadrats (plots), points, or transects. The *sample* is simply part of the population, a subset of the total possible number of sampling units. These terms can be clarified in reference to an artificial plant population shown in Figure 5.1. There are a total of 400 plants in this population, distributed in 20 patches of 20 plants each. All the plants are contained within the boundaries of a 20m x 20m "macroplot." The collection of plants in this macroplot population will be referred to as the "400-plant population." A random arrangement of ten 2m x 2m quadrats positioned within the 400-plant population is shown in

Figure 5.1. Counts of plants within the individual quadrats are directed at the objective of estimating the total number of plants in the 20m x 20m macroplot. The sampling unit in this case is the 2m x 2m quadrat. The sample shown in Figure 5.1 is a set of 10 randomly selected quadrats. The population in this case is the total collection of all possible 2m x 2m quadrats that could be placed in the macroplot (N=100).

Sample information

Coordinates X	Coordinates Y	# of plants
2	2	4
6	4	0
16	4	3
12	6	2
14	6	5
6	8	10
0	12	0
2	12	6
14	12	0
2	14	20

Population parameters
Total population size:
400 plants
Mean # plants/quadrat:
$\mu = 4$
Standard deviation:
$\sigma = 5.005$

Sample statistics (n = 10)
Mean # plants/quadrat:
$\bar{x} = 5.0$
Standard deviation:
s = 6.146

Population estimate
Estimated population
size = 500 plants
95% confidence
interval = ± 361 plants

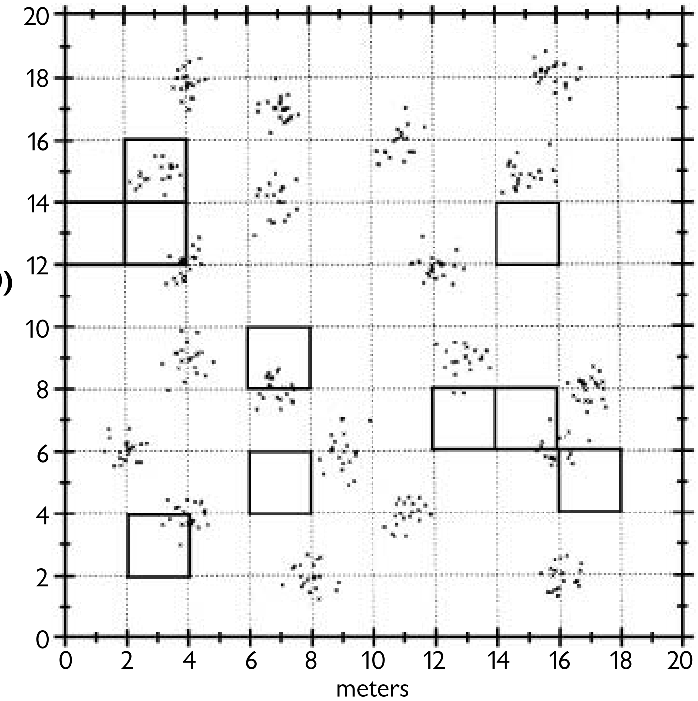


FIGURE 5.1. Population of 400 plants distributed in 20 clumps of 20 plants. This figure shows a simple random sample of ten 2m x 2m quadrats, along with sample statistics and true population parameters.

C. Population Parameters vs. Sample Statistics

Population parameters are descriptive measures which characterize the population and are assumed to be fixed but unknown quantities that change only if the population changes. Greek letters such as μ and σ are often used to denote parameters. If we count all the plants in all the quadrats that make up the 400-plant population shown in Figure 5.1 (400 plants), and divide by the total number of possible 2m x 2m quadrat locations in the macroplot (100 quadrats), we obtain the true average number of plants per quadrat (4 plants/quadrat). This, assuming we have made no errors, is the *true population mean (μ)*. If we know how much each individual quadrat differs from the true population mean, we can calculate another important population parameter, the *true population standard deviation (σ)*. The standard deviation is a measure of how similar each individual observation is to the overall mean and is the most common measure of variability used in statistics. Populations with a large amount of variation among possible sampling units will have a larger standard deviation than populations with sampling units that are more similar to one another.

Sample statistics are descriptive measures derived from a sample (e.g., 10 of the possible 100 2m x 2m quadrats). Sample statistics provide estimates of population parameters. Sample statistics will vary from sample to sample in addition to changing whenever the underlying population changes. Roman letters such as $\bar{X}$ and s are usually used for sample statistics. Consider the following simple example where a sample of three sampling units yields values of 9, 10, and 14 plants/quadrat:

The *sample mean* ($\bar{X}$) = (9+10+14)/3 = 11 plants/quadrat

We could also calculate from this sample a *sample standard deviation (s)*. The sample standard deviation describes how similar each individual observation is to the sample mean. The derivation of a standard deviation (in case you want to calculate one by hand) is provided in Appendix 8. The standard deviation is easily calculated with a simple hand calculator using the "s" or "σ_{n-1}" key.

The standard deviation (s) for the simple example above is 2.65 plants/quadrat.

Consider another simple example with sampling unit values of 2, 10, and 21 plants/quadrat.

The mean ($\bar{X}$) = (2+10+21)/3 = 11 plants/quadrat

The standard deviation (s) for this example is 9.54 plants/quadrat.

Thus, both examples have a sample mean of 11 plants/quadrat, but the second one has a higher standard deviation (9.54 plants/quadrat) than the first (2.65 plants/quadrat), because the individual quadrat values differ more from one another in the second example.

In the example shown in Figure 5.1, the true population mean is 4.00 plants/quadrat, whereas the sample mean is 5.00 plants/quadrat. The true population standard deviation is 5.005 plants/quadrat, whereas the sample standard deviation is 6.146 plants/quadrat.

D. Accuracy vs. Precision

Accuracy is the closeness of a measured or computed value to its true value. *Precision* is the closeness of repeated measurements of the same quantity. A simple example will help illustrate the difference between these two terms. Two quartz clocks, equally capable of tracking time, are sitting side-by-side on a table. Someone comes by and advances one of the clocks by 1 hour. Both clocks will be equally "precise" at tracking time, but one of them will not be "accurate."

Efficient sampling designs try to achieve high precision. When we sample to estimate some population parameter, our sample standard deviation gives us a measure of the repeatability, or precision of our sample; it does not allow us to assess the accuracy of our sample. If counts of plants within different quadrats of a sample are similar to one another (e.g., the example above with a mean of 11 and a standard deviation = 2.65) then it is likely that different independent samples from the same population will yield similar sample means and give us high precision. When quadrat counts within a sample are highly variable (e.g., the example above with a mean of 11 and a standard deviation of 9.54), individual sample means from separate independent samples may be very different from one another giving us low precision. In either case, if the counting process is biased (perhaps certain color morphs or growth forms of individuals are overlooked), results may be inaccurate.

E. Sampling vs. Nonsampling errors

In any monitoring study errors should be minimized. Two categories of errors are described next.

1. Sampling errors

Sampling errors result from chance; they occur when sample information does not reflect the true population information. These errors are introduced by measuring only a subset of all the sampling units in a population.

Sampling errors are illustrated in Figure 5.2, in which two separate, completely random samples (2A and 2B) are taken from the 400-plant population shown in Figure 5.1. In each case, ten 2m x 2m quadrats are sampled and an estimate is made of the total number of plants within the population. The sample shown in Figure 5.2A produces a population estimate of only 80 plants, whereas the sample shown in Figure 5.2B yields an estimate of 960 plants. Both estimates are poor because of sampling error (chance placement of the quadrats resulted in severe under- or overestimates of the true population total).

You can imagine the problems that can arise if you monitor the same population two years in a row and get sample information that indicates that the population shifted from 960 plants to 80 plants when it really didn't change at all. Sampling errors can lead to two kinds of mistakes: (1) missing real changes (missed-change errors), and (2) detecting apparent changes that don't really exist (false-change errors).

Sampling errors can be estimated from the sampling data. Some of the basic sampling design tools covered in Chapter 7, enable you to evaluate the effectiveness of your monitoring study by taking a closer look at the sampling data. This can be especially helpful when setting up new projects; an evaluation of pilot sampling data can point out potential sampling error problems, enabling an investigator to fix them at an early stage of the project. Good sampling designs can reduce sampling errors without increasing the cost of sampling.

2. Nonsampling errors

Nonsampling errors are errors associated with human, rather than chance, mistakes. Examples of nonsampling errors include:

- Using biased selection rules, such as selecting "representative samples" by subjectively locating sampling units, or by substituting sampling units that are "easier" to measure.

- Using vegetation measurement or counting techniques within sampling units in which attributes cannot be accurately counted or measured. For example, counts of grass stems within a quadrat with counts in the hundreds may lead to numerous counting errors.

- Inconsistent field sampling effort. Nonsampling errors can be introduced if different investigators use different levels of effort (e.g., one investigator makes counts from "eye-level," whereas another counts by kneeling next to the quadrat).

- Transcription and recording errors. Nonsampling errors can be introduced if the data recorder's "7's" look like "1's" to the person entering the data.

- Incorrect or inconsistent species identification. This category also includes biases introduced by missing certain size classes or color morphs.

A

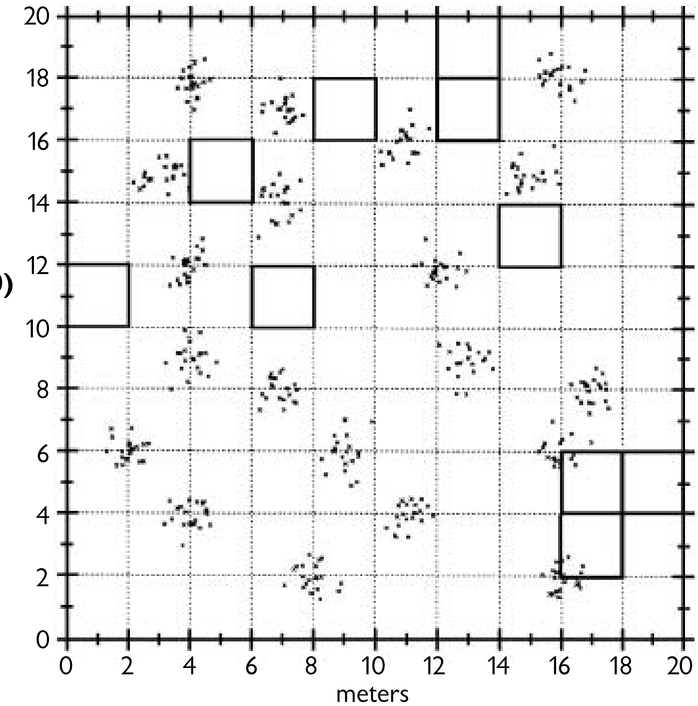
Sample information

Coordinates X	Coordinates Y	# of plants
16	2	5
16	4	3
18	4	0
0	10	0
6	10	0
14	12	0
4	14	0
8	16	0
12	16	0
12	18	0

Population parameters
Total population size:
400 plants
Mean # plants/quadrat:
$\mu = 4$
Standard deviation:
$\sigma = 5.005$

Sample statistics (n = 10)
Mean # plants/quadrat:
$\bar{x} = 0.8$
Standard deviation:
$s = 1.75$

Population estimate
Estimated population
size = 80 plants
95% confidence
interval = ± 119 plants



B

Sample information

Coordinates X	Coordinates Y	# of plants
16	0	5
10	2	11
16	4	3
14	6	5
12	8	18
6	8	10
4	16	9
2	12	6
10	16	9
14	14	20

Population parameters
Total population size:
400 plants
Mean # plants/quadrat:
$\mu = 4$
Standard deviation:
$\sigma = 5.005$

Sample statistics (n = 10)
Mean # plants/quadrat:
$\bar{x} = 9.6$
Standard deviation:
$s = 5.58$

Population estimate
Estimated population
size = 960 plants
95% confidence
interval = ± 379 plants

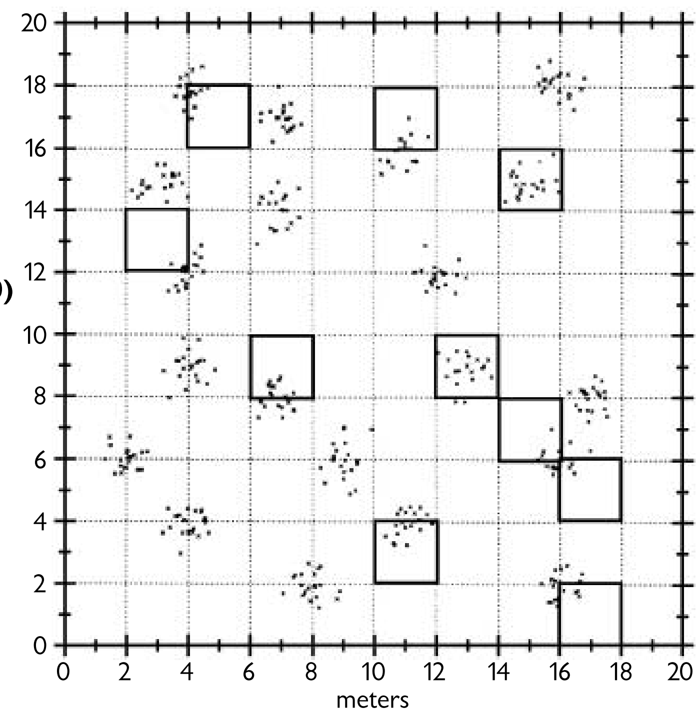


FIGURE 5.2. Examples of sampling errors from sampling the 400-plant population. The population estimates of 80 plants and 960 plants are far from the true population of 400 plants.

Because sampling designs are based on the assumption that nonsampling errors are zero, the number of nonsampling errors needs to be minimized. Ensure that your sampling unit makes sense for the type of vegetation measurement technique you have selected. When different personnel are used in the same monitoring study, conduct rigorous training and testing to ensure consistency in counts or measurements. Design field data forms (Chapter 9) that are easy to use and easy for data transcribers to interpret. Proof all data entered into computer programs to ensure that entered numbers are correct. In contrast to sampling errors, the probability of nonsampling errors occurring cannot be assessed from pilot sample data.

F. Sampling Distributions

One way of evaluating the risk of obtaining a sample value that is vastly different than the true value (such as population estimates of 80 or 960 plants when the true population is 400 plants) is to sample a population repeatedly and look at the differences among the repeated population estimates. If almost all the separate, independently derived population estimates are similar, then you know you have a good sampling design with high precision. If many of the independent population estimates are not similar, then you know your precision is low.

The 400-plant population can be resampled by erasing the 10 quadrats (as shown in either Figure 5.1 or Figure 5.2) and putting 10 more down in new random positions. We can keep repeating this procedure, each time writing down the sample mean. Plotting the results of a large number of individual sample means in a simple histogram graph yields a *sampling distribution*. A sampling distribution is a distribution of many independently gathered sample statistics (most often a distribution of sample means). Under most circumstances, this distribution of sample means fits a normal or bell-shaped curve.

A distribution of population size estimates from 10,000 separate random samples using ten 2m x 2m quadrats from the 400-plant population is shown in Figure 5.3A. The x-axis shows the range of different population estimates, and the y-axis shows the relative and actual frequency of the different population estimates. Think of this as the results of 10,000 different people sampling the same population on the same day, each one setting out 10 randomly positioned 2m x 2m quadrats (somehow without negatively impacting the population) and coming up with their own independent population estimate. The highest population estimate out of the 10,000 separate samples was 960 plants and the lowest population estimate was zero (four of the 10,000 samples yielded a population estimate of zero). The shape of this distribution indicates the magnitude of likely sampling errors. Wide distributions could yield population estimates that are "far" from the true population value. A sampling design that led to the type of sampling distribution depicted in Figure 5.3A would not be useful since few of the estimates approach the true population size of 400 plants. *One of the principal objectives in sampling design is to make the shape of sampling distributions as narrow as possible*.

Fortunately, you do not have to repeatedly sample your population and see how wide your sampling distribution is to determine if you need to change anything. There are some simple statistical tools that provide a convenient shortcut for evaluating the precision of your sampling effort from a single sample. These tools involve calculating standard errors and confidence intervals to estimate sampling precision levels.

1. Standard error

A *standard error* is the standard deviation of a large number of independent sample means. It is a measure of precision that you derive from a single sample. The formula for calculating a standard error is as follows:

Formula for standard error:

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

Where: SE = Standard error
s = Standard deviation
n = Sample size

To paraphrase the earlier statement regarding an important objective of sampling design, *one of the principal objectives in sampling design is to reduce the size of the standard error*.

This formula demonstrates that there are only two ways of minimizing standard errors, either: (1) increase sample size (n), or (2) decrease the standard deviation (s).

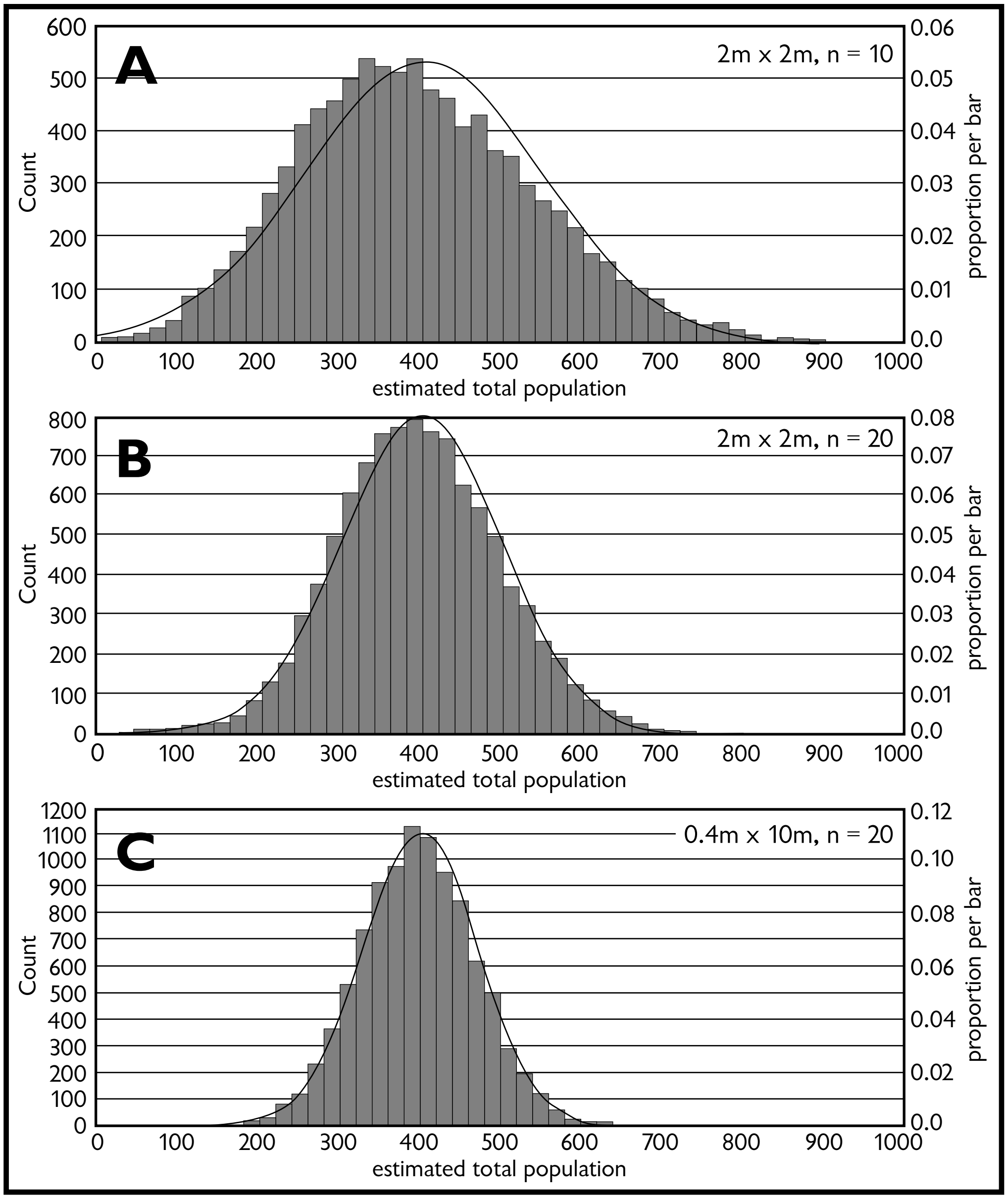


FIGURE 5.3. Sampling distributions from three separate sampling designs used on the 400-plant population. All distributions were created by sampling the population 10,000 separate times. The smooth lines show a normal bell-shaped curve fit to the data. Figure 3A shows a sampling distribution where ten 2m x 2m quadrats were used. Figure 3B shows a sampling distribution where twenty 2m x 2m quadrats were used. Figure 3C shows a sampling distribution where twenty 0.4m x 10m quadrats were used.

- **Increase sample size**. A new sampling distribution of 10,000 separate random samples drawn from our example population is shown in Figure 5.3B. This distribution came from randomly drawing samples of *twenty* 2m x 2m quadrats instead of the *ten* quadrats used to create the sampling distribution in Figure 5.3A. This increase in sample size from 10 to 20 provides a 29.3% improvement in precision (as measured by the reduced size of the standard error).

- **Decrease sample standard deviation**. Another sampling distribution of 10,000 separate random samples drawn from our 400-plant population is shown in Figure 5.3C. The sampling design used to create this distribution of population estimates is similar to the one used to create the sampling distribution in Figure 5.3B. The only difference between the two designs is in quadrat shape. The sampling distribution shown in Figure 5.3B came from using twenty 2m x 2m quadrats; the sampling distribution shown in Figure 5.3C came from using twenty 0.4m x 10m quadrats. This change in quadrat shape reduced the true population standard deviation from 5.005 plants to 3.551 plants. This change in quadrat shape led to a 29.0% improvement in precision over the 2m x 2m design shown in Figure 5.3B (as measured by the reduced size of the standard error). This 29.0% improvement in precision came without changing the sampling unit size ($4m^2$) or the number of quadrats sampled (n=20); only the quadrat shape (from square to rectangular) changed. When compared to the original sampling design of ten 2m x 2m quadrats, the twenty 0.4m x 10m quadrat design led to a 49.8% improvement in precision. Details of this method and other methods of reducing sample standard deviation are covered in Chapter 7.

How is the standard error most often used to report the precision level of sampling data? Sometimes the standard error is reported directly. You may see tables with standard errors reported or graphs that include error bars that show ± 1 standard error. Often, however, the standard error is multiplied by a coefficient that converts the number into something called a confidence interval.

2. Confidence intervals

A confidence interval provides an estimate of precision around a sample mean, a sample proportion, or an estimate of total population size that specifies the likelihood that the interval includes the true value. The vertical lines marked with the "95%" in Figure 5.4 indicate that 95% of all the samples (9,500 out of the 10,000) fit between these two lines. Five percent of the samples (2.5% in each tail of the distribution) fall outside the vertical lines. These lines

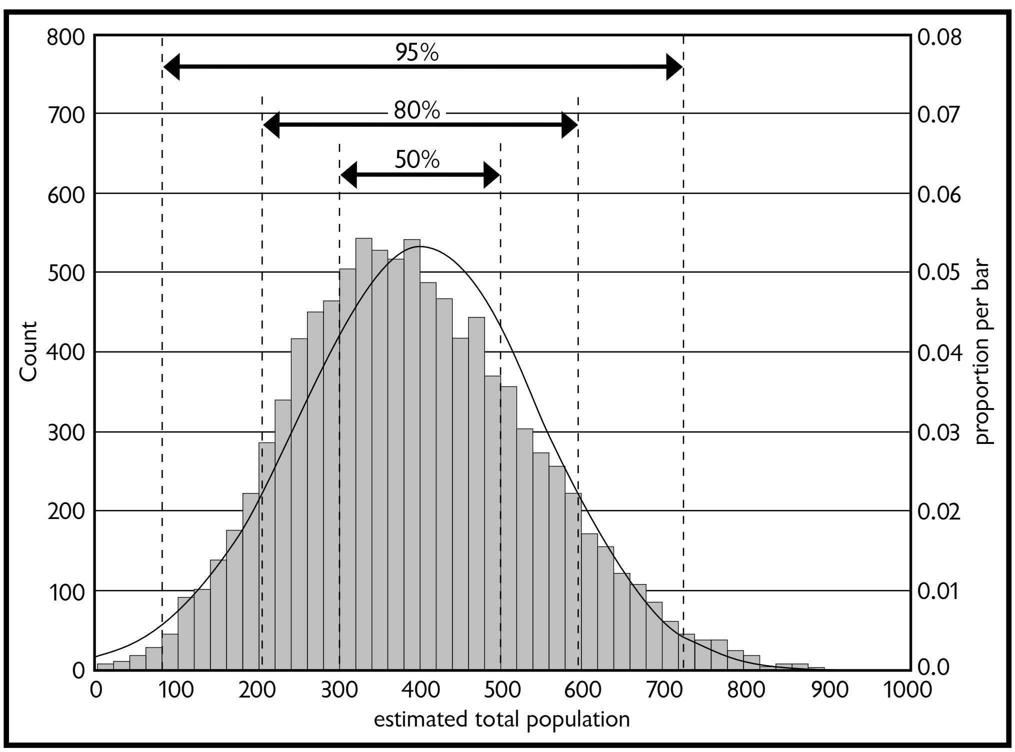


FIGURE 5.4. Distribution from sampling the 400-plant population 10,000 times using ten samples of 2m x 2m quadrats. The 95%, 80%, and 50% confidence intervals around the true population of 400 plants are shown. The smooth line shows a normal, bell-shape curve fit to the data.

are positioned equally from the center of the sampling distribution, approximately 320 plants away from the center of 400 plants. Thus, 95% of all samples are within ± 320 plants of the true population size.

A confidence interval includes two components: (1) the *confidence interval width* (e.g., ± 320 plants); and (2) the *confidence level* (e.g., 90%, 95%). The confidence level indicates the probability that the interval includes the true value. Confidence interval width decreases as the confidence level decreases. This relationship is shown in Figure 5.4 where three different confidence levels are graphed on the sampling distribution obtained by sampling the 400-plant population with ten 2m x 2m quadrats. These same three confidence intervals are shown again in Figure 5.5A, where they are graphed in a format commonly used to report confidence intervals. There is no gain in precision associated with the narrowing of confidence interval width as you go from left to right in Figure 5.5A (i.e., from 95% confidence, to 80% confidence, to 50% confidence); only the probability that the confidence interval includes the true value is altered. Another set of three confidence intervals is shown in Figure 5.5B. Like Figure 5.5A, confidence intervals get narrower as we move from left to right in the graph, but this time the confidence level is the same (95%) and the narrower widths came from using different sampling designs. There is a gain in precision associated with the narrowing of confidence interval width as you go from left to right in Figure 5.5B (i.e., from the ten 2m x 2m design to the twenty 2m x 2m design to the twenty 0.4m x 10m design) because we have reduced the uncertainty of our population estimate by tightening the confidence interval width at the same confidence level.

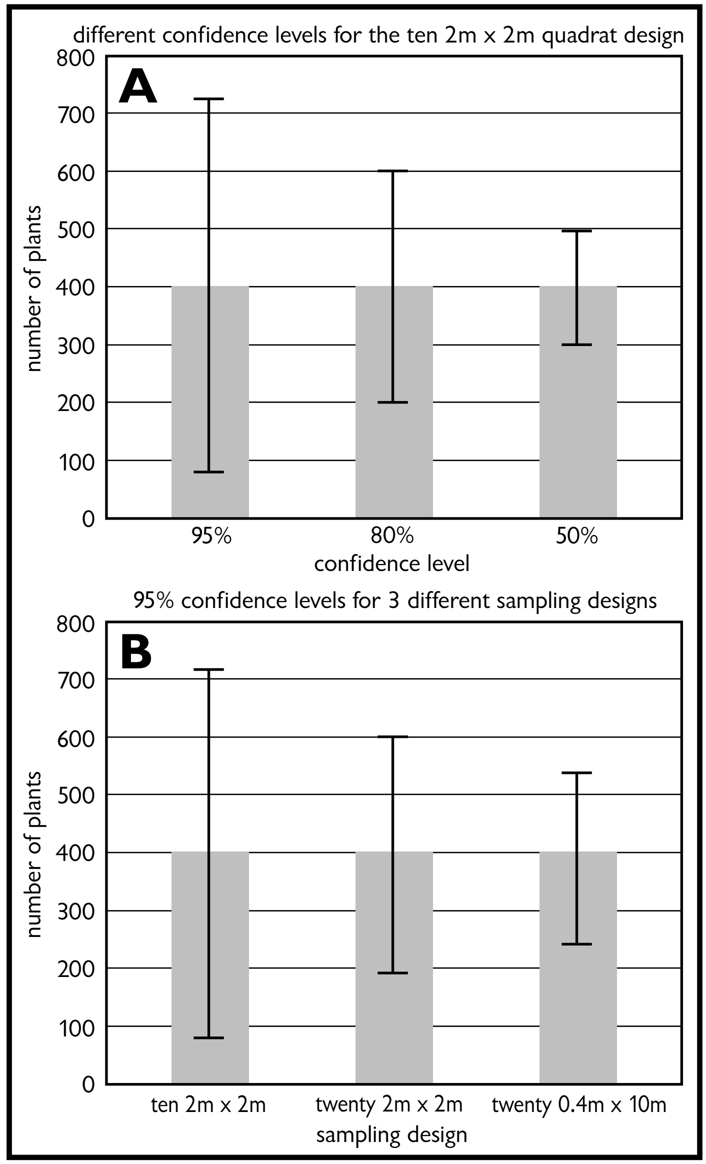


FIGURE 5.5. Comparison of confidence intervals and confidence levels for different sampling designs from the 400-plant population. Figure 5A shows three different confidence levels (95%, 80%, and 50%) for the same data set based upon sampling ten 2m x 2m quadrats. Figure 5B shows 95% confidence intervals for three different sampling designs that differ in the level of precision of the population estimates.

In order to calculate confidence intervals for sample means, we need two values: (1) the standard error calculated according to the above formula ($SE = s/\sqrt{n}$), and (2) the corresponding value from a table of critical values of the *t* distribution (see Appendix 8 for instructions on calculating confidence intervals around proportions). The confidence interval half-width, extending an equal distance on both sides of the mean, is the standard error × the critical *t* value (except when sampling from finite populations, see below). The appropriate critical value of t depends on the level of confidence desired and the number of sampling units (n) in the sample. A table of critical values for the t distribution (Zar 1996) is found in Appendix 5. To use this table, you must first select the appropriate confidence level column. If you want to be 95% confident that your confidence interval includes the true mean, use the column headed $\alpha(2) = 0.05$. For 90% confidence, use the column headed $\alpha(2) = 0.10$. You use $\alpha(2)$ because you are interested in a confidence interval on both sides of the mean. You then use the row indicating the number of degrees of freedom (v), which is the number of sampling units minus one $(n\text{-}1)$.

For example, if we sample 20 quadrats and come up with a mean of 4.0 plants and a standard deviation of 5.005, here are the steps for calculating a 95% confidence interval around our sample mean:

The standard error $(SE = s/\sqrt{n}) = \dfrac{5.005}{\sqrt{20}} = 5.005/4.472 = 1.119$.

The appropriate t value from Appendix 5 for 19 degrees of freedom (v) is 2.093.

One-half of our confidence interval width is then SE $\times$ t-value = $1.119 \times 2.093 = 2.342$.

Our 95% confidence interval can then be reported as 4.0 ± 2.34 plants/quadrat or we can report the entire confidence interval width from 1.66 to 6.34 plants/quadrat. This indicates a 95% chance that our interval from 1.66 plants/quadrat to 6.34 plants/quadrat includes the true value.

Another way to think of 95% confidence intervals calculated from sampling data is that the interval specifies a range that should include the true value 95% of the time. If you are calculating 95% confidence intervals and independently randomly sample a population 100 different times, you should see that approximately 95 of the intervals will include the true mean and 5 will miss it. This relationship is shown in Figure 5.6 where 100 independent population estimates are graphed with 95% confidence intervals from the 400-plant populations using samples of twenty 0.4m x 10m quadrats. You will notice that the solid dots, used to show each of the 100 population estimates, fluctuate around the true population value of 400 plants. You will also notice that 96 out of 100 confidence intervals shown in Figure 5.6 include the true value. If the confidence level had been set at 80%, then approximately 20 of the intervals would have failed to include the true value. A 99% confidence level would have led to approximately only one interval out of the 100 that did not include the true population size (in order to capture the true value more often, the individual confidence interval widths for a 99% confidence level are wider than the confidence interval widths for a 95% confidence level).

G. Finite vs. Infinite Populations

If we are sampling with quadrats, there is a finite number of quadrats that can be placed in the area to be sampled, assuming that no two quadrats overlap (this is called sampling without replacement). If the sampled area is large, then the number of quadrats placed in the area may be very large as well, but nonetheless finite. On the other hand, an infinite number of points or lines could be placed in the area to be sampled. This is because points, at least theoretically, are dimensionless, and lines are dimensionless in one direction. This means, at least for all practical purposes, that a population of points or of lines is infinite.

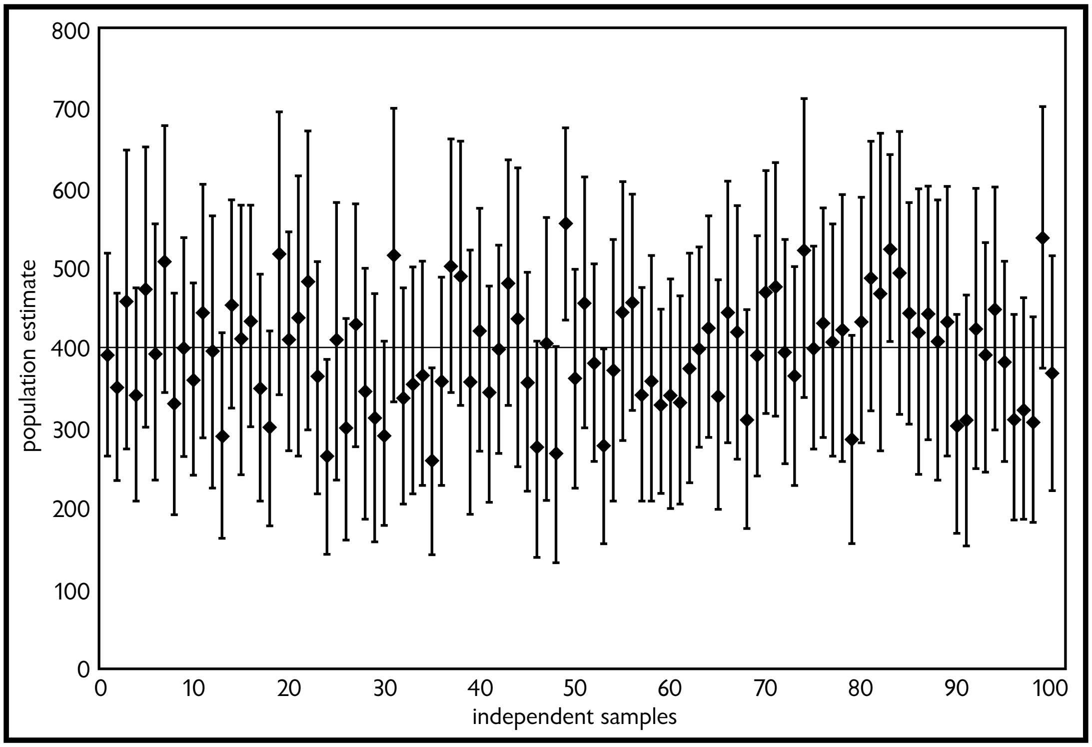


FIGURE 5.6. Population estimates from 100 separate random samples from the 400-plant population. Each sample represents the population estimate from sampling twenty 0.4m x 10m quadrats. The horizontal line through the graph indicates the true population of 400 plants. Vertical bars represent 95% confidence intervals. Four of the intervals miss the true population size.

If the area to be sampled is large relative to the area that is actually sampled, the distinction between finite and infinite is of only theoretical interest. When, however, the area sampled makes up a significant portion of the area to be sampled, we can apply the *finite population correction factor*, which reduces the size of the standard error. The most commonly used finite population correction factor is shown to the right:

Formula for the finite population correction:

$$FPC = \sqrt{\frac{N - n}{N}}$$

Where: N = total number of potential quadrat positions
n = number of quadrats sampled

When n is small relative to N, the equation is close to 1, whereas when n is large relative to N, the value approaches zero. The standard error ($s/\sqrt{n}$) is multiplied by the finite population correction factor to yield a corrected standard error for the finite population.

Consider the following example. The density of plant species X is estimated within a 20m x 50m macroplot (total area = $1000m^2$). This estimate is obtained by collecting data from randomly selected 1m x 10m quadrats ($10m^2$). Sampling without replacement, there are 100 possible quadrat positions. $\frac{\text{total area}}{\text{quadrat area}} = \frac{1000m^2}{10m^2}$

Thus, our population, N, is 100. Let's say we take a random sample, n, of 30 of these quadrats and calculate a mean of eight plants per quadrat and a standard deviation of four plants per quadrat. Our standard error is thus: $s/\sqrt{n} = 4/\sqrt{30} = 0.73$. Although our sample mean is an unbiased estimator of the true population mean and needs no correction, the standard error should be corrected by the finite population correction factor shown on the top of page 72:

Example of applying the finite population correction factor:

$$SE' = (SE)\sqrt{\frac{N - n}{N}} \qquad SE' = (0.73)\sqrt{\frac{100 - 30}{100}} = 0.61$$

Where: SE' = corrected standard error
SE = uncorrected standard error
N = total number of potential quadrat positions
n = number of quadrats sampled

Since the standard error is one of the factors used to calculate confidence intervals (the other is the appropriate value of *t* from a *t* table), correcting the standard error with the finite population correction factor makes the resulting confidence interval narrower. It does this, however, *only* if *n* is sufficiently large relative to *N*. A rule of thumb is that unless the ratio *n/N* is greater than .05 (i.e., you are sampling more than 5% of the population area), there is little to be gained by applying the finite population correction factor to your standard error.

The finite population correction factor is also important in sample size determination (Chapter 7) and in adjusting test statistics (Chapter 11). The finite population correction factor works, however, only with finite populations, which we will have when using quadrats, but will not have when using points or lines.

H. False-Change Errors and Statistical Power Considerations

These terms relate to situations where two or more sample means or proportions are being compared with some statistical test. This comparison may be between two or more places or the same place between two or more time periods. These terms are pertinent to both planning and interpretation stages of a monitoring study. Consider a simple example where you have sampled a population in two different years and now you want to determine whether a change took place between the two years. You usually start with the assumption, called the null hypothesis, that no change has taken place. There are two types of decisions that you can make when interpreting the results of a monitoring study: (1) you can decide that a change took place, or (2) you can decide that no change took place. In either case, you can either be right, or you can be wrong (Figure 5.7).

monitoring for change — possible errors

	no change has taken place	there has been a real change
monitoring system detects a change	false-change error (Type I) α	no error (Power) $1 - \beta$
monitoring system detects no change	no error $(1 - \alpha)$	missed-change error (Type II) β

FIGURE 5.7. Four possible outcomes for a statistical test of some null hypothesis, depending on the true state of nature.

1. The change decision and false-change errors

The conclusion that a change took place may lead to some kind of action. For example, if a population of a rare plant is thought to have declined, a change in management may be needed. If a change did not actually occur, this constitutes a false-change error, a sort of false alarm. Controlling this type of error is important because taking action unnecessarily can be expensive (e.g., a range permittee is not going to want to change the grazing intensity if a decline in a rare plant population really didn't take place). There will be a certain probability of concluding that a change took place even if no difference actually occurred. The probability of this occurring is usually labeled the *P*-value, and it is one of the types of information that comes out of a statistical analysis of the data. The *P*-value reports the likelihood that the observed difference was due to chance alone. For example, if a statistical test comparing two sample means yields a *P*-value of 0.24 this indicates that there is a 24% chance of obtaining the observed result even if there is no true difference between the two sample means.

Some threshold value for this false-change error rate should be set in advance so that the *P*-value from a statistical test can be evaluated relative to the threshold. *P*-values from a statistical test that are smaller than or equal to the threshold are considered statistically "significant," whereas *P*-values that are larger than the threshold are considered statistically "nonsignificant." Statistically significant differences may or may not be ecologically significant depending upon the magnitude of difference between the two values. The most commonly cited threshold for false-change errors is the 0.05 level; however, there is no reason to arbitrarily adopt the 0.05 level as the appropriate threshold. The decision of what false-change error threshold to set depends upon the relative costs of making this type of mistake and the impact of this error level on the other type of mistake, a missed-change error (see below).

2. The no-change decision, missed-change errors, and statistical power

The conclusion that no change took place usually does not lead to changes in management practices. Failing to detect a true change constitutes a *missed-change error*. Controlling this type of error is important because failing to take action when a true change actually occurred may lead to the serious decline of a rare plant population.

Statistical power is the complement of the missed-change error rate (e.g., a missed-change error rate of 0.25 gives you a power of 0.75; a missed-change error rate of 0.05 gives you a power of 0.95). High power (a value close to 1), is desirable and corresponds to a low risk of a missed-change error. Low power (a value close to 0) is not desirable because it corresponds to a high risk of a missed-change error.

Since power levels are directly related to missed-change error levels, either level can be reported and the other level can be easily calculated. Power levels are often reported instead of missed-change error levels, because it seems easier to convey this concept in terms of the certainty of detecting real changes. For example, the statement "I want to be at least 90% certain of detecting a real change of five plants/quadrat" (power is 0.90) is simpler to understand than the statement "I want the probability of missing a real change of five plants/quadrat to be 10% or less" (missed-change error rate is 0.10).

An assessment of statistical power or missed-change errors has been virtually ignored in the field of environmental monitoring. A survey of over 400 papers in fisheries and aquatic sciences found that 98% of the articles that reported nonsignificant results failed to report any power results (Peterman 1990). A separate survey, reviewing toxicology literature, found high power in only 19 out of 668 reports that failed to reject the null hypothesis (Hayes 1987). Similar surveys in other fields such as psychology or education have turned up "depressingly low" levels of power (Brewer 1972; Cohen 1988).

3. Minimum detectable change

Another sampling design concept that is directly related to statistical power and false-change error rates is the size of the change that you want to be able to detect. This will be referred to as the *minimum detectable change or MD*C. The MDC greatly influences power levels. A particular sampling design will be more likely to detect a true large change (i.e., with high power) than to detect a true small change (i.e., with low power).

Setting the MDC requires the consideration of ecological information for the species being monitored. How large of a change should be considered biologically meaningful? With a large enough sample size, statistically significant changes can be detected for changes that have no biological significance. For example, if an intensive monitoring design leads to the conclusion that the mean density of a plant population increased from 10.0 plants/m^2 to 10.1 plants/m^2, does this represent some biologically meaningful change in population density? Probably not.

Setting a reasonable MDC can be difficult when little is known about the natural history of a particular plant species. Should a 30% change in the mean density of a rare plant population be cause for alarm? What about a 20% change or a 10% change? The MDC considerations are likely to vary when assessing vegetation attributes other than density, such as cover or frequency (Chapter 8). The initial MDC, set during the design of a new monitoring study, can be modified once monitoring information demonstrates the size and rate of population fluctuations.

4. How to achieve high statistical power

Statistical power is related to four separate sampling design components by the following function equation:

Power = a function of (s, n, MDC, and α)

where: s = standard deviation
n = number of sampling units
MDC = minimum detectable change
α = false-change error rate

Power can be increased in the following four ways:

1. **Reducing standard deviation**. This means altering the sampling design to reduce the amount of variation among sampling units (see Chapter 7).

2. **Increasing the number of sampling units sampled**. This method of increasing power is straightforward, but keep in mind that increasing *n* has less of an effect than decreasing *s* since the square root of sample size is used in the standard error equation ($SE = s/\sqrt{n}$).

3. **Increasing the acceptable level of false-change errors (α)**.

4. **Increasing the MDC**.

Note that the first two ways of increasing power are related to making changes in the sampling design, whereas the other two ways are related to making changes in the sampling objective (see Chapter 6).

5. Graphical comparisons

As stated, power is driven by four different factors: standard deviation, sample size, minimum detectable change size, and false-change error rate. In this section we take a graphical look at how altering these factors changes power. The comparisons in this section are based upon sampling a fictitious plant population where we are interested in assessing plant density relative to an established threshold value of 25 plants/m^2. Any true population densities less than 25 plants/m^2 will trigger management action. We are only concerned with the question of whether the density is lower than 25 plants/m^2 and not whether the density is higher. In this example, our null hypothesis (H_O) is that the population density equals 25 plants/m^2 and our alternative hypothesis is that density is less than 25 plants/m^2. The density value of 25 plants/m^2 is the most critical single density value since it defines the lower limit of acceptable plant density.

The figures in this section are all based upon sampling distributions where we happen to know the true plant density. Recall that a sampling distribution is a bell-shaped curve that depicts the distribution of a large number of independently gathered sample statistics. A sampling distribution defines the range and relative probability of any possible sample mean. You are more likely to obtain sample means near the middle of the distribution than you are to obtain sample means near either tail of the distribution.

A sampling distribution based on sampling our fictitious population with a true mean density of 25 plants/m^2 is shown in Figure 5.8A. This distribution is based on a sampling design using thirty 1m x 1m quadrats where the true standard deviation is ± 20 plants/quadrat. If 1,000 different people randomly sample and calculate a sample mean based upon their 30 quadrat values, approximately half the individually drawn sample means will be less than 25 plants/m^2 and half will be greater than 25 plants/m^2. Approximately 40% of the samples will yield sample means less than or equal to 24 plants/m^2. A few of our 1,000 individuals will obtain estimates of the mean density that deviate from the true value by a large margin. One of the individuals will likely stand up and say, "my estimate of the mean density is 13 plants/m^2", even though the true density is actually 25 plants/m^2. As interpreters of the monitoring information, we would conclude that since 999 of the 1,000 people obtained estimates of the density that were greater than 13, the true density is probably not 13. Our best estimate of the true mean density will be the average of the 1,000 separate estimates (this average is likely to be extremely close to the actual true value).

Now that we have the benefit of 1,000 independent estimates of the true mean density, we can return to the population at a later time, take a single random sample of thirty 1m x 1m quadrats, calculate the sample mean, and then ask the question, "what is the probability of obtaining our sample mean value if the true population is still 25 plants/m^2?" If our sample mean density turns out to be 24 plants/m^2, would this lead to the conclusion that the population has crossed our threshold value? Seeing that our sample mean is lower than our target value might raise some concerns, but we have no objective basis to conclude that the true population is not, in fact, still actually 25 plants/m^2. We learned in the previous paragraph that a full 40% of possible samples are likely to yield mean densities of 24 plants/m^2 or less if the true mean is 25 plants/m^2. Thus, the probability of obtaining a single sample mean of 24 plants/m^2 or less when the true density is actually 25 plants/m^2 is approximately 0.40. Obtaining a sample mean of 24 plants/m^2 is consistent with the hypothesis that the true population density is actually 25 plants/m^2.

How small a sample mean do we need to obtain to feel confident that the population has indeed dropped below 25 plants/m^2? What will our interpretation be if we obtained a sample mean of 22 plants/m^2? Based upon our sampling distribution from the 1,000 people, the probability of obtaining an estimate of 22 plants/m^2 or less is around 20%, which represents a 1-in-5 chance that the true mean is still actually 25 plants/m^2. Based upon the sampling distribution from our 1,000 separate samplers, we can look at the likelihood of obtaining other different sample means. The probability of obtaining a sample of 20 plants/m^2 is 8.5%,

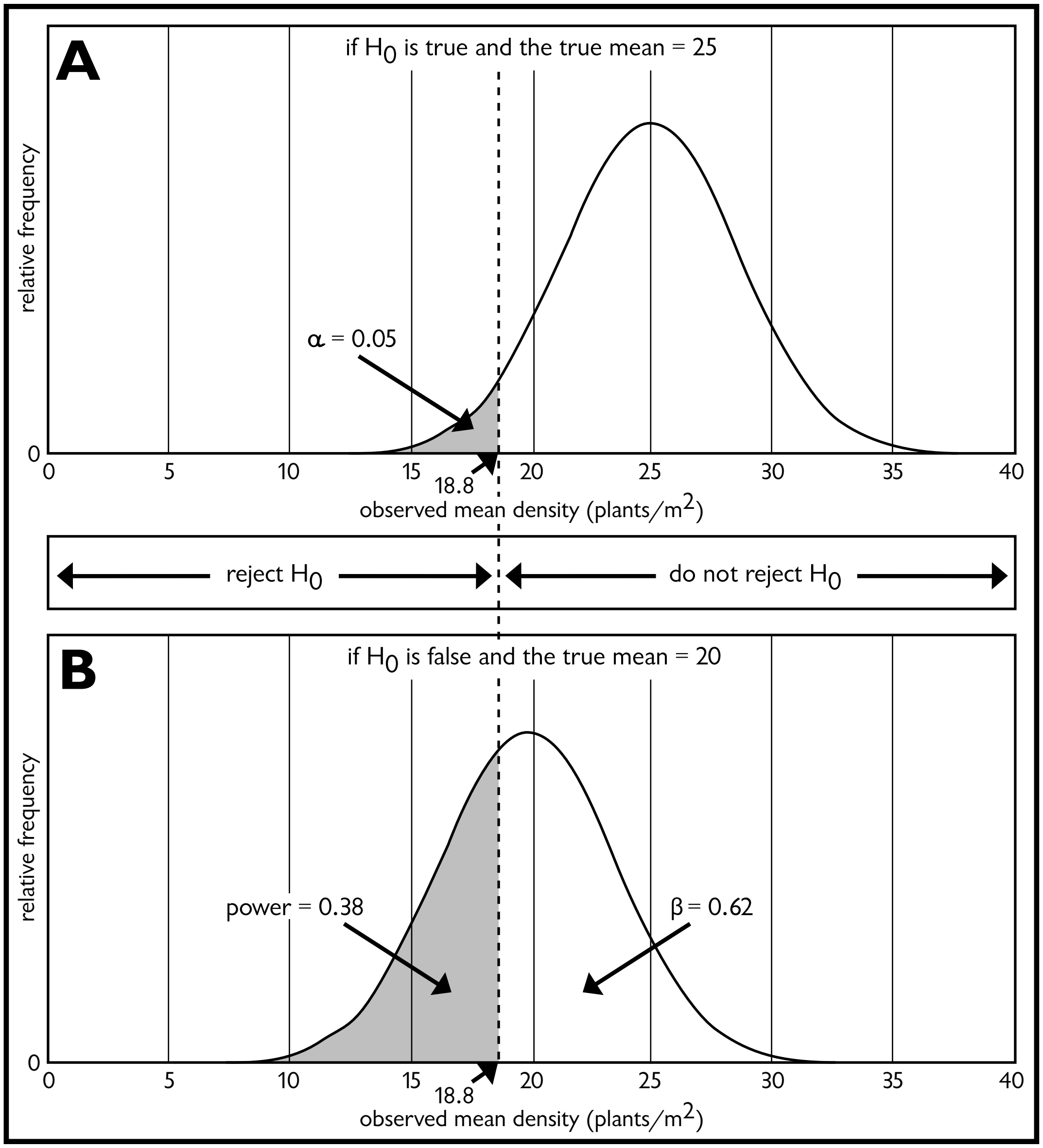


FIGURE 5.8. Example of sampling distributions for mean plant density in samples of 30 permanent quadrats where the among-quadrat standard deviation is 20 plants/m^2. Part A is the sampling distribution for the case in which the null hypothesis, H_0, is true and the true population mean density is 25 plants/m^2. The shaded area in part A is the critical region for $\alpha = 0.05$ and the vertical dashed line is at the critical sample mean value, 18.8. Part B is the sampling distribution for the case in which the H_0 is false and the true mean is 20 plants/m^2. In both distributions, a sample mean to the left of the vertical dashed line would reject H_0, and to the right of it, would not reject H_0. Power and β values in part B, in which H_0 is false and the true mean = 20, are the proportion of sample means that would occur in the region in which H_0 was rejected or not rejected, respectively (adapted from Peterman 1990).

and the probability of obtaining a sample of 18 plants/m^2 is 2.9% if the true mean density is 25 plants/m^2.

Since in most circumstances we will only have the results from a single sample (and not the benefit of 1,000 independently gathered sample means), another technique must be used to determine whether the population density has dropped below 25 plants/m^2. One method is to run a statistical test that compares our sample mean to our density threshold value (25 plants/m^2). The statistical test will yield a *P*-value that defines the probability of obtaining our sample mean if the true population density is actually 25 plants/m^2. As interpreters of our monitoring information, we will need to set some probability threshold *P*-value to guide our interpretation of the results from the statistical test. This *P*-value threshold defines our acceptable false-change error rate. If we run a statistical test that compares our sample mean to our density threshold value (25 plants/m^2) and the *P*-value from the test is lower than our threshold value, then we conclude that the population density has, in fact, declined below 25 plants/m^2. Thus, if we set our *P*-value threshold to 0.05 and the statistical test yields a *P*-value of 0.40, then we *fail to reject* the null hypothesis that the true population density is 25 plants/m^2. If, however, the statistical test yields a *P*-value of 0.022, this is lower than our threshold *P*-value of 0.05, and we would reject the null hypothesis that the population is 25 plants/m^2 in favor of our alternative hypothesis that the density is lower than 25 plants/m^2.

The relationship between the *P*-value threshold of 0.05 and our sampling distribution based upon sampling thirty 1m x 1m quadrats is shown in Figure 5.8A. The threshold density value corresponding to our *P*-value threshold of 0.05 is 18.8 plants/m^2, which is indicated on the sampling distribution by the dashed vertical line. Thus, if we obtain a mean density of 18 plants/m^2, which is to the left of the vertical line, we reject the null hypothesis that the population density is 25 plants/m^2 in favor of an alternative hypothesis that density is lower than 25 plants/m^2. If we obtain a mean density of 21 plants/m^2, which is to the right of the vertical line, then we fail to reject the null hypothesis that the population density is really 25 plants/m^2.

So far we have been discussing the situation where the true population density is right at the threshold density of 25 plants/m^2. Let's look now at a situation where we know the true density has declined to 20 plants/m^2. What is the likelihood of our detecting this true density difference of 5 plants/m^2? Figure 5.8B shows a new sampling distribution based upon the true density of 20 plants/m^2 (standard deviation is still ± 20 plants/m^2). We know from our previous discussion that sample means to the right of the vertical line in Figure 5.8A lead to the conclusion that we can't reject the null hypothesis that our density is 25 plants/m^2. If our new sample mean turns out to exactly match the new true population mean (i.e., 20 plants/m^2), will we reject the idea that the sample actually came from a population with a true mean of 25 plants/m^2? No, at least not at our stated *P*-value (false-change error) threshold of 0.05. A sample mean value of 20 plants/m^2 falls to the right of our dashed threshold line in the "do not reject H_O" portion of the graph and we would have failed to detect the true difference that actually occurred. Thus, we would have committed a missed-change error.

What is the probability of missing the true difference of five plants/m^2 show in Figure 5.8B? This probability represents the missed-change error rate (β) and it is defined by the nonshaded area under the sampling distribution in Figure 5.8B, which represents 62% of the possible sample mean values. Recall that the area under the whole curve defines the entire range of possible values that you could obtain by sampling the population with the true mean = 20 plants/m^2. If we bring back our 1,000 sampling people and have each of them sample thirty 1m x 1m quadrats in our new population, we will find that approximately 620 of them will obtain estimates of the mean density that are greater than the 18.8 plants/m^2 threshold value that is shown by the vertical dashed line. What about the other 380 people? They will obtain population estimates fewer than the critical threshold of 18.8 plants/m^2 and they will reject the null hypothesis that the population equals 25 plants per quadrat. This proportion of 0.38 (380 people out of 1,000 people sampling) represents the *statistical power* of our sampling design and it is represented by the shaded area under the curve in Figure 5.8B. If the true population mean is indeed 20 plants/m^2 instead of 25 plants/m^2, then we can be 38%

(power = 0.38) sure that we will detect this true difference of five plants/m^2. With this particular sampling design (thirty 1m x 1m quadrats) and a false-change error rate of α=0.05, we run a 62% chance (β=0.62) that we will commit a missed-change error (i.e., fail to detect the true difference of five plants/m^2). If the difference of five plants/m^2 is biologically important, a power of only 0.38 would not be satisfactory.

We can improve the low-power situation in four different ways: (1) increase the acceptable false-change error rate; (2) increase the acceptable MDC; (3) increase sample size; or (4) decrease the standard deviation. New paired sampling distributions illustrate the influence of making each of these changes.

a. Increasing the acceptable false-change error rate

In Figure 5.8B, a false-change error rate of α=0.05 resulted in a missed-change error rate of β=0.62 to detect a difference of five plants/m^2. Given these error rates, we are more than 12 times more likely to commit a missed-change error than we are to commit a false-change error. What happens to our missed-change error rate if we specify a new, higher false-change error rate? Shifting our false-change error rate from α=0.05 to α=0.10 is illustrated in Figure 5.9

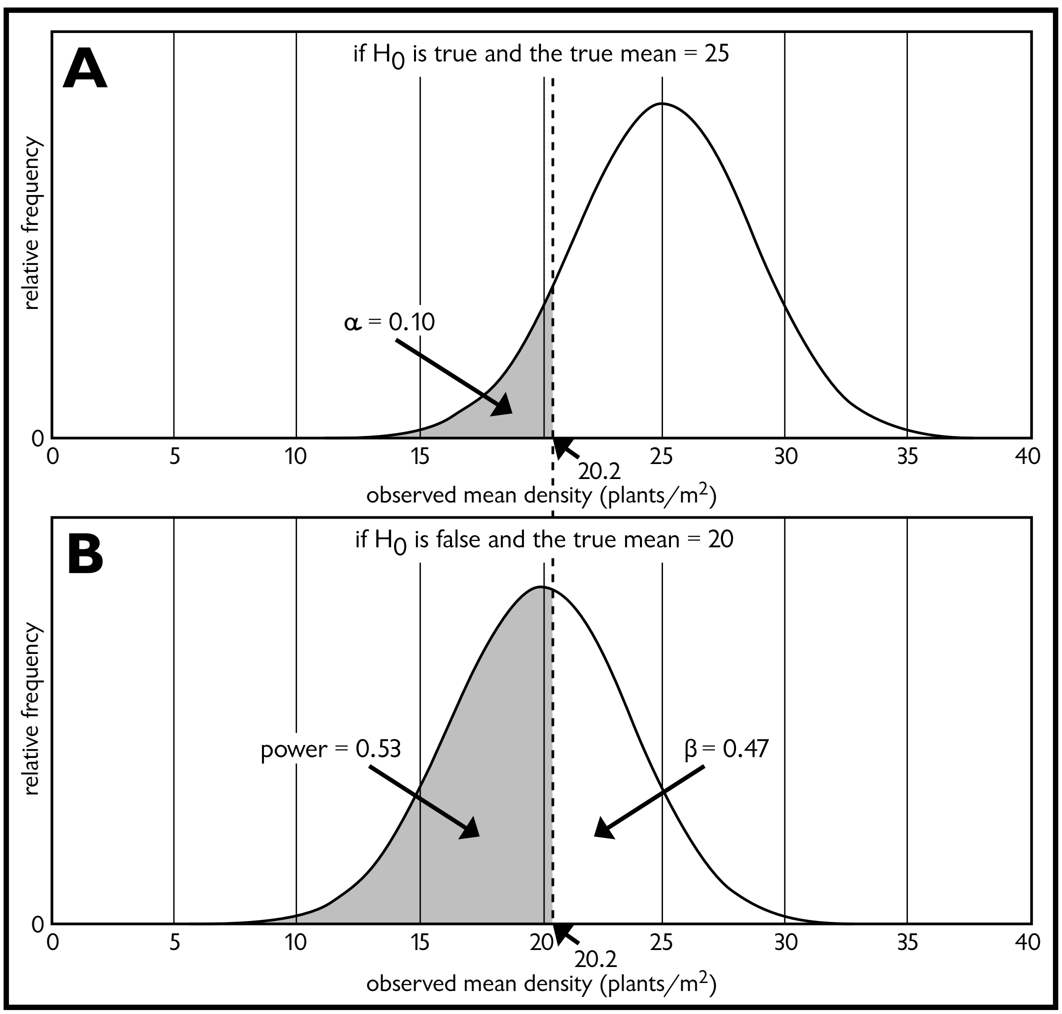


FIGURE 5.9. The critical region for the false-change error in the sampling distributions from Figure 5.8 has been increased from α = 0.05 to α = 0.10. Part B, in which the H_0 is false and the true mean = 20, shows that power is larger for α = 0.10 than for Figure 5.8 where α = 0.05 (adapted from Peterman 1990).

for the same sampling distributions shown in Figure 5.8. Our critical density threshold at the *P*=0.10 level is now 20.21 plants/m^2, and our missed-change error rate has dropped from β=0.62 down to β=0.47 (i.e., the power to detect a true five plant/m^2 difference went from 0.38 to 0.53). A sample mean of 20 plants/m^2 will now lead to the correct conclusion that a difference of five plants/m^2 between the populations does exist. Of course, the penalty we pay for increasing our false-change error rate is that we are now twice as likely to conclude that a difference exists in situations when there is no true difference and our population mean is actually 25 plants/m^2. Changing the false-change error rate even more, to α=0.20, (Figure 5.10) reduces the probability of making a missed-change error down to β=0.29 (i.e., giving us a power of 0.71 to detect a true difference of five plants/m^2).

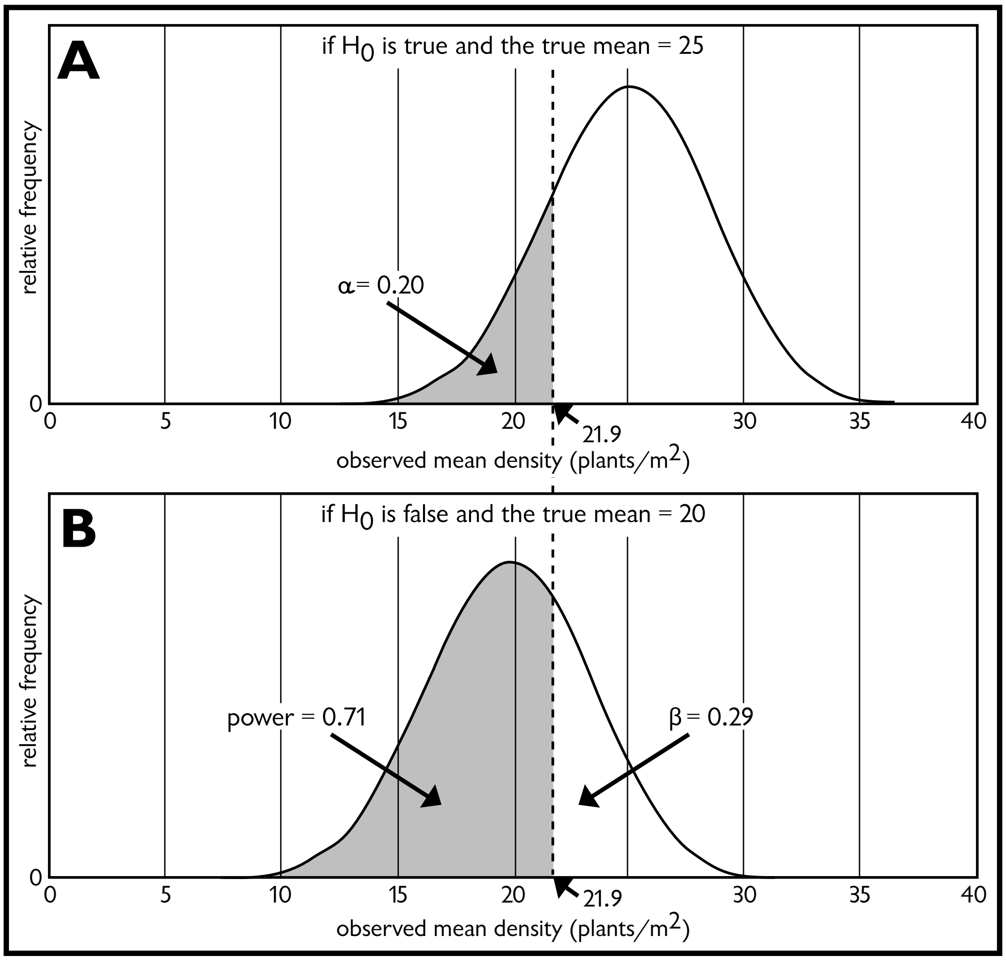


FIGURE 5.10. The critical region for the false-change error in the sampling distributions from Figure 5.8 has been increased from α = 0.05 to α = 0.20. Part B, in which the H_0 is false and the true mean = 20, shows that power is larger for α = 0.20 than for Figure 5.8 where α = 0.05 or Figure 5.9 where α = 0.10. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

b. Increasing the acceptable MDC

Any sampling design is more likely to detect a true large difference than a true small difference. As the magnitude of the difference increases, we will see an increase in the power to detect the difference. This relationship is shown in Figure 5.11B, where we see a sampling distribution with a true mean density of 15 plants/m^2, which is 10 plants/m^2 below our threshold density of 25 plants/m^2. The false-change error rate is set at α=0.05 in this example.

This figure shows that the statistical power to detect this larger difference of 10 plants/m² (25 plants/m² to 15 plants/m²) is 0.85 compared with the original power value of 0.38 to detect the difference of five plants/m² (25 plants/m² to 20 plants/m²). Thus, with a false-change error rate of 0.05, we can be 85% certain of detecting a difference from our 25 plants/m² threshold of 10 plants/m² or greater. If we raised our false-change error from α=0.05 to α=0.10 (not shown in Figure 5.11), our power value would rise to 0.92, which creates a sampling situation where our two error rates are nearly equal (α=0.10, β=0.08).

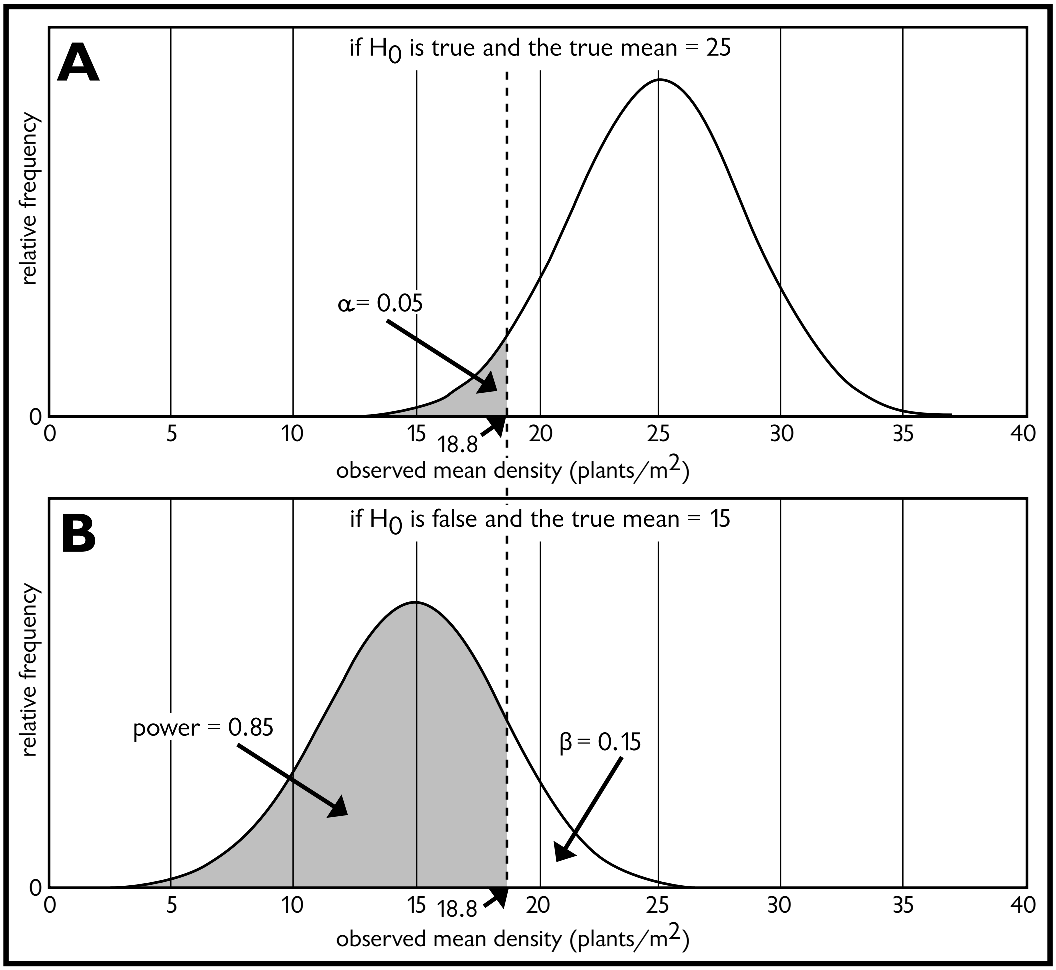


FIGURE 5.11. Part A is the same as Figure 5.8; in part B, the true population mean is 15 plants/m² instead of the 20 plants/m² shown in Figure 5.8. Note that power increases (and β decreases) when the new true population mean gets further from the original true mean of 25 plants/m². Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

c. Increasing the sample size

The sampling distributions shown in Figures 5.8 to 5.11 were all created by sampling the populations with n=30 1m x 1m quadrats. Any increase in sample size will lead to a subsequent increase in power to detect some specified minimum detectable difference. This increase in power results from the sampling distributions becoming narrower. Sampling distributions based on samples of n=50 are shown in Figure 5.12 where the true difference between the two populations is once again five plants/m² with a false-change error rate threshold of α=0.05. The increase in sample size led to an increase in power from power=0.38 with n=30 to power=0.54 with n=50. Note that the critical threshold density associated with an α = 0.05 is now 20.3 plants/m^2 as compared to the 18.8 plants/m^2 threshold when n = 30.

d. Decreasing the standard deviation

The sampling distributions shown in Figures 5.8 to 5.12 all are based on sampling distributions with a standard deviation of ±20 plants/m^2. The quadrat size used in the sampling was a square 1m x 1m quadrat. If individuals in the plant population are clumped in distribution, then it is likely that a rectangular shaped quadrat will result in a lower standard deviation (See Chapter 7 for a detailed description of the relationship between standard deviation and sampling unit size and shape). Figure 5.13 shows sampling distributions where the standard deviation was reduced from ±20 plants/m^2 to ±10 plants/m^2. Note that the critical threshold density associated with an α = 0.05 is now 21.9 plants/m^2 as compared to the 18.8 plants/m^2 threshold when the standard deviation was ± 20 plants/m^2. This reduction in the true standard deviation came from a change in quadrat shape from the 1m x 1m square shape to a 0.2m x 5m rectangular shape. Note that quadrat area (1m^2) stayed the same so that the mean densities are consistent with the previous sampling distributions shown in Figures 5.8

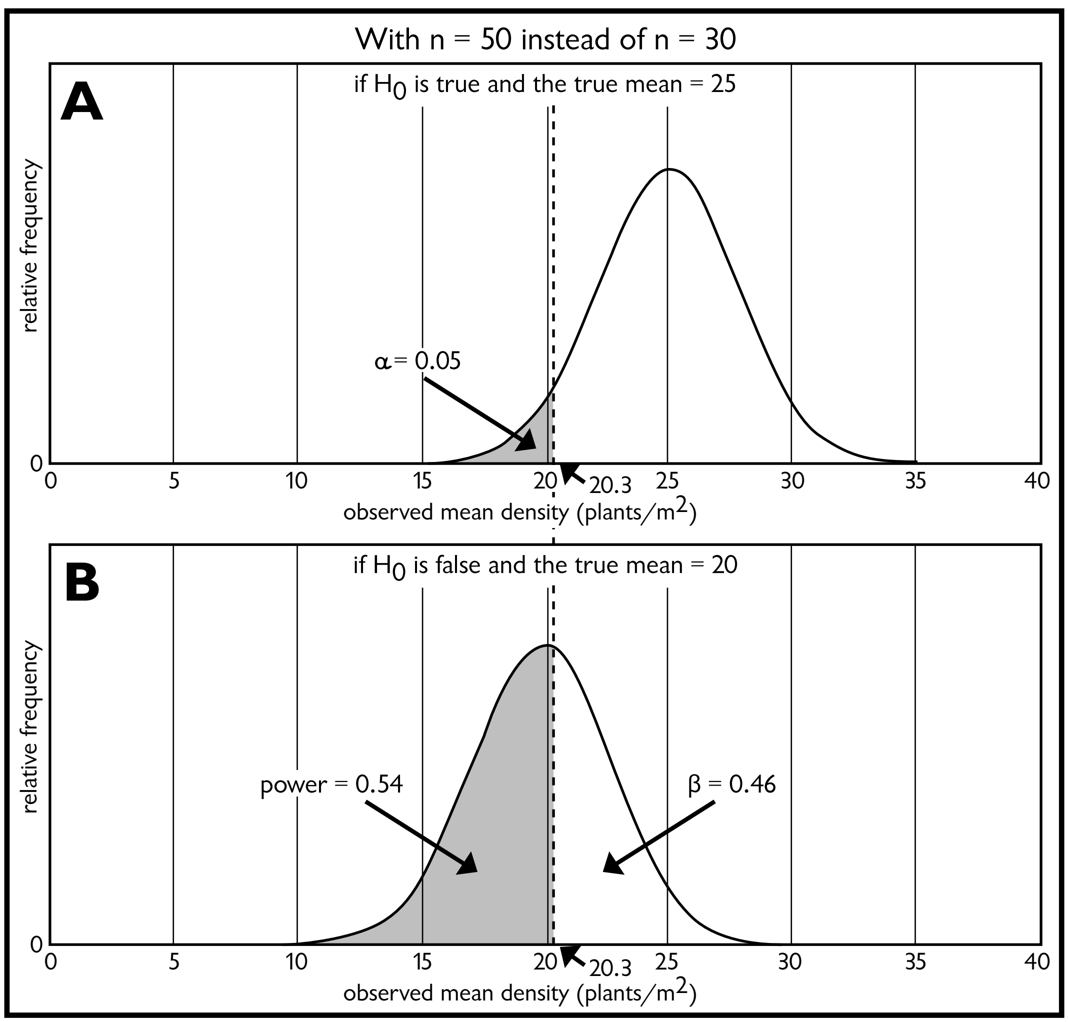


FIGURE 5.12. The sample size was increased to n = 50 quadrats from the n = 30 quadrats shown in Figure 5.8. Note that power increases (and β decreases) at larger sample sizes. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

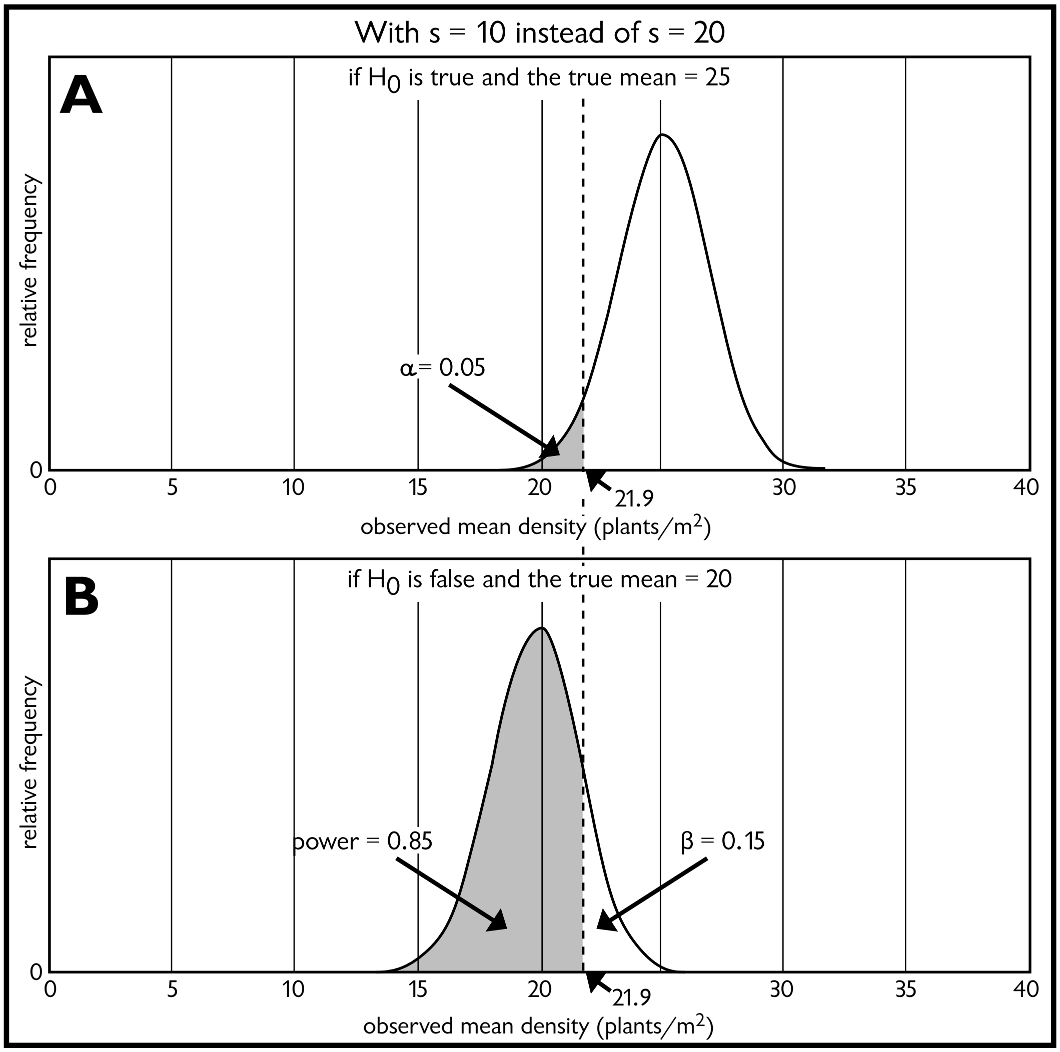


FIGURE 5.13. The standard deviation (s) of 20 plants/m^2 shown in Figure 5.8 is reduced to ten plants/m^2. Note that power increases (and β decreases), as the standard deviation decreases. Again, a sample mean to the left of the vertical dashed line would reject H_0, while one to the right of it would not reject H_0 (adapted from Peterman 1990).

through 5.12. This reduction in standard deviation led to a dramatic improvement in power, from 0.38 (with sd = 20 plants/m^2) to 0.85 (with sd = 10 plants/m^2). Reducing the standard deviation has a more direct impact on increasing power than increasing sample size, because the sample size is reduced by taking its square root in the standard error equation ($SE = s/\sqrt{n}$). Recall that the standard error provides an estimate of sampling precision from a single sample without having to enlist the support of 1,000 people who gather 1,000 independent sample means.

e. Power curves

The relationship between power and the different sampling design components that influence power can also be displayed in power curve graphs. These graphs typically show power values on the y-axis and either sample size, MDC, or standard deviation values on the x-axis. Figure 5.14A shows statistical power graphed against different magnitudes of change for the same hypothetical data set described above and shown in Figures 5.8 to 5.11. Four different power curve lines are shown, one for each of the following four different false-change (α) error rates: 0.01, 0.05, 0.10, and 0.20. The power curves are based on sampling with a sample size of 30 quadrats and a standard deviation of 20 plants/m^2. For any particular false-change error rate, power increases as the magnitude of the minimum detectable change increases. When α=0.05, the power to detect small changes is very low (Figure 5.14A). For example, we have only a 13% chance of detecting a difference of 2 plants/m^2 (i.e., a density of 23 plants/m^2 which is 2 plants/m^2 below our threshold value of 25 plants/m^2). In contrast, we can be 90% sure of detecting a minimum difference of 11 plants/m^2. We can also attain higher power by increasing

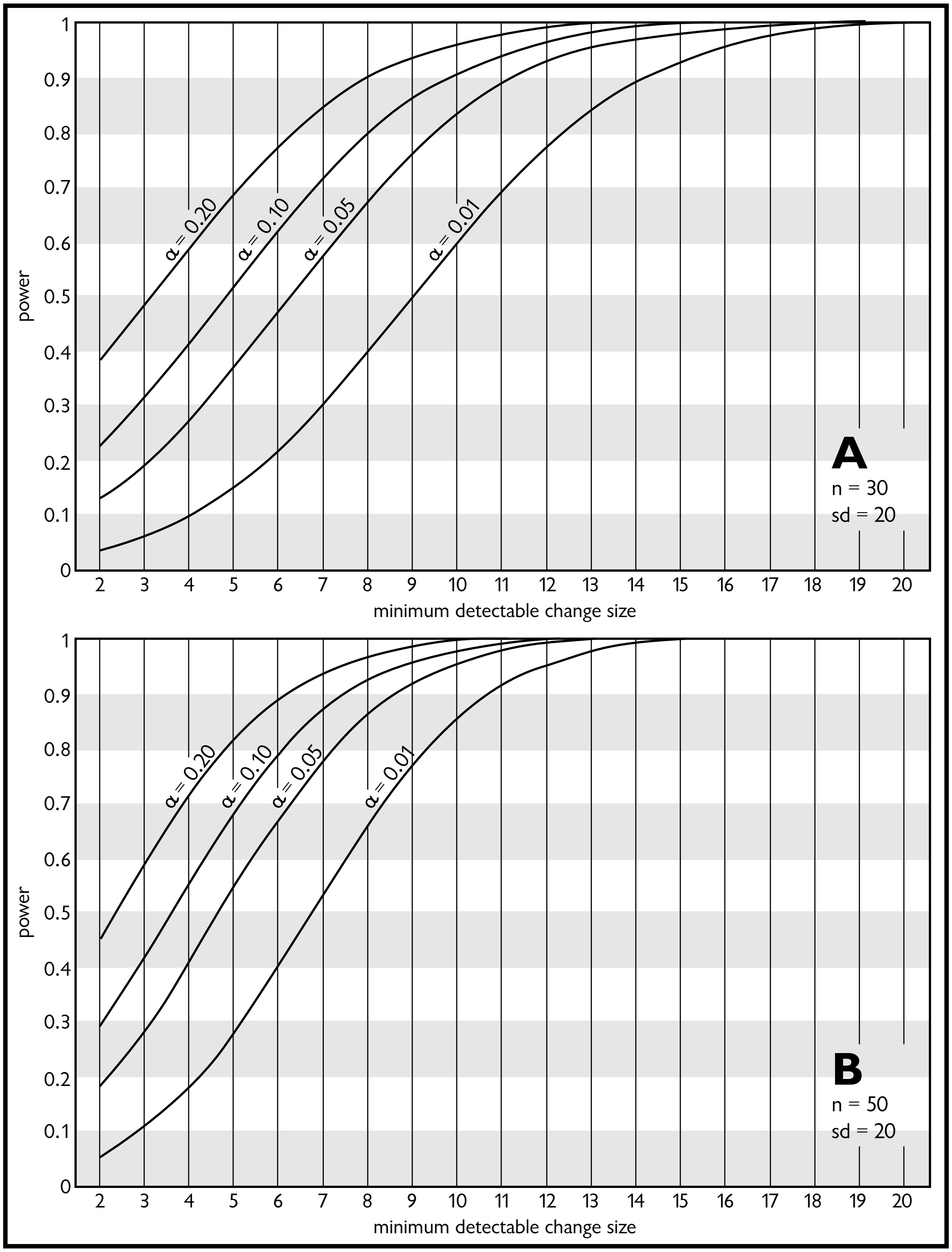


FIGURE 5.14. Power curves showing power values for various magnitudes of minimum detectable change and false-change error rates when the standard deviation is 20. Part A shows power curves with a sample size of 30. Part B shows power curves with a sample size of 50.

the false-change error rate. The power to detect a change of eight plants/m^2 is only 0.41 when α=0.01, but it increases to 0.69 at α=0.05, to 0.81 at α=0.10, and to 0.91 at α=0.20.

A different set of power curves are shown in Figure 5.14B where the sample size is n=50 instead of the n=30 shown in Figure 5.14A. This larger smaller sample size shifts all of the power curves to the left, making it more likely that smaller changes will be detected. For example, with a false-change error rate of α=0.10, the power to detect a seven plant/m^2 difference is 0.88 with a sample size of n=50 quadrats compared to the power of 0.73 with a sample size of n=30 quadrats.

Power curves that show the effect of reducing the standard deviation are shown in Figure 5.15. Figure 5.15A is the same as Figure 5.14A where the standard deviation is 20 plants/m^2. Figure 5.15B shows the same power curves except they are based on a standard deviation of 10 plants/m^2. The smaller standard deviation shifts all of the power curves to the left and results in much steeper slopes. The smaller standard deviation leads to substantially higher power levels for any particular MDC value. For example, the power to detect a change of five plants/m^2 with a false change error rate of α=0.10 is only 0.53 in Figure 5.15A as compared to the power of 0.92 in Figure 5.15B.

6. Setting false-change and missed-change error rates

Both false-change and missed-change error rates can be reduced by sampling design changes that increase sample size or decrease sample standard deviations. Missed-change and false-change error rates are inversely related, which means that reducing one will increase the other (but not proportionately) if no other changes are made. The decision of which type of error is more important should be based on the nature of the changes you are trying to determine, and the consequences of making either kind of mistake.

Because false-change and missed-change error rates are inversely related to each other, and because these errors have different consequences to different interest groups, there are different opinions as to what the "acceptable" error rates should be. The following examples demonstrate the conflict between false-change and missed-change errors.

- **Testing for a lethal disease**. When screening a patient for some disease that is lethal without treatment, a physician is less concerned about making a false diagnosis error (analogous to a false-change error) of concluding that the person has the disease when they do not than failing to detect the disease (analogous to a missed-change error) and concluding that the person does not have the disease when in fact they do.

- **Testing for guilt in our judicial system**. In the United States, the null hypothesis is that the accused person is innocent. Different standards for making judgement errors are used depending upon whether the case is a criminal or a civil case. In criminal cases, proof must be "beyond a reasonable doubt." In these situations it is less likely that an innocent person will be convicted (analogous to a false-change error), but it is more likely that a guilty person will go free (analogous to a missed-change error). In civil cases, proof only needs to be "on the balance of probabilities." In these situations, there is a greater likelihood of making a false conviction (analogous to a false-change error), but a lower likelihood of making a missed conviction (analogous to a missed-change) error.

- **Testing for pollution problems**. In pollution monitoring situations, the industry has an interest in minimizing false-change errors and may desire a very low false-change error

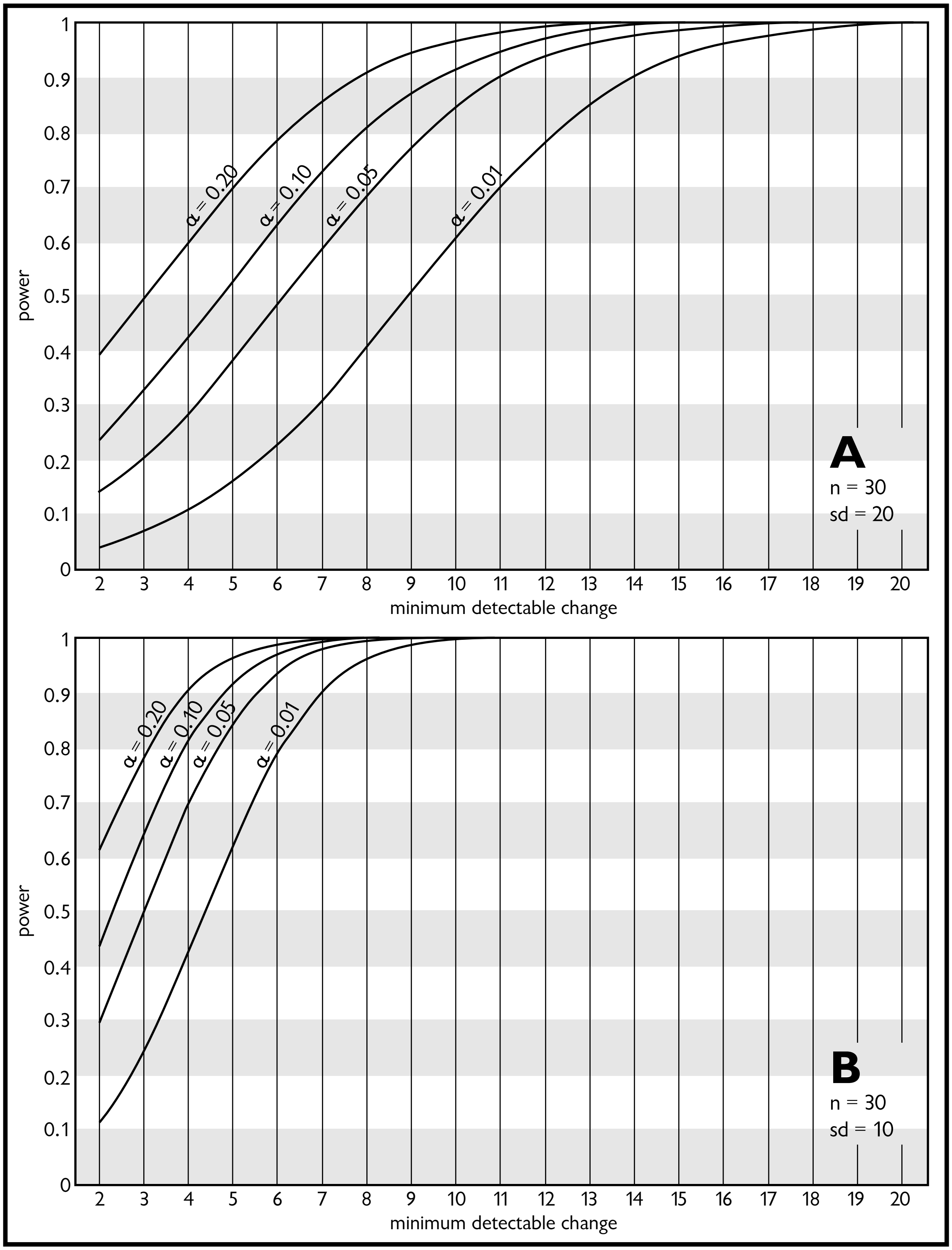


FIGURE 5.15. Power curves showing power values for various magnitudes of minimum detectable change and false-change error rates when the sample size = 30. Part A shows power curves with a standard deviation of 20 plants/m². Part B shows power curves with a standard deviation of 10 plants/m².

rate (e.g., $\alpha = 0.01$ or 0.001). Companies do not want to be shut down or implement expensive pollution control procedures if a real impact has not occurred. In contrast, an organization concerned with the environmental impacts of some pollution activity will likely want to have high power (low missed-change error rate) so that they do not miss any real changes that take place. They may not be as concerned about occasional false-change errors (which would result in additional pollution control efforts even though real changes did not take place).

Missed-change errors may be as costly or more costly than false-change errors in environmental monitoring studies (Toft and Shea 1983; Peterman 1990; Fairweather 1991). A false-change error may lead to the commitment of more time, energy, and people, but probably only for the short period of time until the mistake is discovered (Simberloff 1990). In contrast, a missed-change error, as a result of a poor study design, may lead to a false sense of security until the extent of the damages are so extreme that they show up in spite of a poor study design (Fairweather 1991). In this case, rectifying the situation and returning the system to its preimpact condition could be costly. For this reason, you may want to set equal false-change and missed-change error rates or even consider setting the missed-change error rate lower than the false-change error rate (Peterman 1990; Fairweather 1991).

There are many historical examples of costly missed-change errors in environmental monitoring. For example, many fish population monitoring studies have had low power to detect biologically meaningful declines so that declines were not detected until it was too late and entire populations crashed (Peterman 1990). Some authors advocate the use of something they call the "precautionary principle" (Peterman and M'Gonigle 1992). They argue that, in situations where there is low power to detect biologically meaningful declines in some environmental parameter, management actions should be prescribed as if the parameter had actually declined. Similarly, some authors prefer to shift the burden of proof in situations where there might be an environmental impact from environmental protection interests to industry/development interests (Peterman 1990; Fairweather 1991). They argue that a conservative management strategy of "assume the worst until proven otherwise" should be adopted. Under this strategy, developments that may negatively impact the environment should not proceed until the proponents can demonstrate, with high power, a lack of impact on the environment.

7. Why has statistical power been ignored for so long?

It is not clear why missed-change errors, power, and minimum detectable change size have traditionally been ignored. Perhaps researchers have not been sufficiently exposed to the idea of missed-change errors. Most introductory texts and statistics courses deal with the material only briefly. Computer packages for power analysis have only recently become available. Perhaps people have not realized the potentially high costs associated with making missed-change errors. Perhaps researchers have not understood how understanding power can improve their work.

The issue of power and missed-change errors has gained a lot of attention in recent years. A literature review in the 1980's would not have turned up many articles dealing with statistical power issues. A literature review today would turn up dozens of articles in many disciplines from journals all over the world (see Peterman 1990 and Fairweather 1991 for good review papers on statistical power). Journal editors may soon start requiring that power analysis information be reported for all nonsignificant results (Peterman 1990). There may also be some departure from the strict adherence to the 0.05 significance level (Peterman 1990; Fairweather 1991).

8. Use of prior power analysis during study design

Power analysis can be useful during both the design of monitoring studies and in the interpretation of monitoring results. The former is sometimes called "prior power analysis," whereas the latter is sometimes called "post-hoc power analysis" (Fairweather 1991). Post-hoc power analysis is covered in Chapter 11.

The use of power analysis during the design and planning of monitoring studies provides valuable information that can help avoid monitoring failures. Once some preliminary or pilot data have been gathered, or if some previous years' monitoring data are available, power analysis can be used to evaluate the adequacy of the sampling design. Prior power analysis can be done in several different ways. All are based upon the power function described earlier:

Power = a function of (s, n, MDC, and α)

The power of a particular sampling design can be evaluated by plugging sample standard deviation, sample size, the desired MDC, and an acceptable false-change error rate, into equations or computer programs (Appendix 16) and solving for power. If the power to detect a biologically important change turns out to be quite low (high probability of a missed-change error), then the sampling design can be modified to try to achieve higher power.

Alternatively, a desired power level can be specified and the terms in the power function can be rearranged to solve for sample size. This will give you assurance that your study design will succeed in being able to detect a certain magnitude of change at the specified power and false-change error rate. This is the format for the sample size equations that are discussed in Chapter 7 and presented in Appendix 7.

Still another way to do prior power analysis is to specify a desired power level and a particular sample size and then rearrange the terms in the power function to solve for the MDC (Rotenberry and Wiens 1985; Cohen 1988). If the MDC is unacceptably large, then attempts should be made to improve the sampling design. If these efforts fail, then the decision must be made to either live with the large MDC or to reject the sampling design and perhaps consider an alternative monitoring approach.

The main advantage of prior power analysis is that it allows the adequacy of the sampling design to be evaluated at an early stage in the monitoring process. It is much better to learn that a particular design has a low power at a time when modifications can easily be made than it is to learn of low power after many years of data have already been gathered. The importance of specifying acceptable levels of false-change and missed-change errors along with the magnitude of change that you want to be able to detect is covered in Chapter 6 the next chapter, which introduces sampling objectives.

Literature Cited

Brewer, J. K. 1972. On the power of statistical tests in the American Educational Research Journal. American Educational Research Journal 9: 391-401.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, N. J. Lawrence Erlbaum Associates.

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring. Australian Journal of Marine and Freshwater Research 42: 555-567.

Hayes, J. P. 1987. The positive approach to negative results in toxicology studies. Ecotoxicology and Environmental Safety 14: 73-77.

McCall, C. H. 1982. Sampling and statistics handbook for research. Ames, IA: The Iowa State University Press.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences. 47: 2-15.

Peterman, R. M.; M'Gognigle. M. 1992. Statistical power analysis and the precautionary principle. Marine Pollution Bulletin. 24(5): 231-234.

Rotenberry, J. T.; Wiens J. A. 1985. Statistical power analysis and community-wide patterns. American Naturalist. 125: 164-168.

Simberloff, D. 1990. Hypotheses, errors, and statistical assumptions. Herpetelogica. 46: 351-357.

Toft, C. A.; P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist. 122: 618-625.

Zar, J. H. 1996. Biostatistical Analysis. Englewood Cliffs, N.J. Jersey: Prentice-Hall, Inc.