Simple Random Sampling and Systematic Sampling

Simple random sampling and systematic sampling provide the foundation for almost all of the more complex sampling designs based on probability sampling. They are also usually the easiest designs to implement. These two designs highlight a trade-offs inherent in selecting a sampling design: to select sample units at random to minimize the risk of introducing biases into the sample or to select samples systematically to ensure that sample units are well-distributed throughout the population.

Both designs involve selecting n sample units from the N units available in the population and can be implemented with or without replacement.

Simple Random Sampling

When the population of interest is relatively homogeneous then simple random sampling works well, which means it provides estimates that are unbiased and have high precision. When little is known about a population in advance, such as in a pilot study, simple random sampling is a common design choice.

Advantages:

- Easy to implement
- Requires little knowledge of the population in advance

Disadvantages:

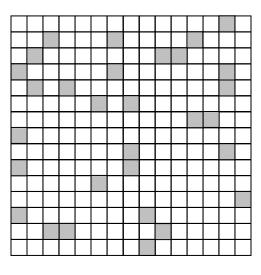
- Imprecise relative to other designs if the population is heterogeneous
- More expensive than other designs if entities are clumped and the cost to travel among units is appreciable

How it is implemented:

• Select *n* sample units at random from *N* available in the population

All units within the sampling universe must have the same probability of being selected, therefore each and every sample of size *n* drawn from the population has an equal chance of being selected.

There are many strategies available for selecting a random sample. For large populations, this often involves generating pseudorandom numbers with a computer and for small populations it might involve using a table of random numbers or even writing a unique identifier for every sample unit in the population on a scrap of paper, placing those numbers in a jar, shaking it, then selecting *n* scraps of paper from the jar blindly. The approach used for selecting the sample matters



little provided there are no constraints on how the sample units are selected and all units have an equal chance of being selected.

Estimating the Population Mean

The population mean (μ) is the true average number of entities per sample unit and is estimated with the sample mean ($\hat{\mu}$ or \bar{y}) which has an unbiased estimator:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n}$$

where y_i is the value from each unit in the sample and n is the number of units in the sample.

The population variance (σ^2) is estimated with the sample variance (s^2) which has an unbiased estimator:

$$s^{2} = \frac{\sum_{i=1}^{n} (y_{i} - \overline{y})^{2}}{n-1}$$

Variance of the estimate $\hat{\mu}$ is: $\hat{\text{var}}(\hat{\mu}) = \left(\frac{N-n}{N}\right) \frac{s^2}{n}$.

The standard error of the estimate is the square root of variance of the estimate, which as always is the standard deviation of the sampling distribution of the estimate. Standard error is a useful gauge of how precisely a parameter has been estimated.

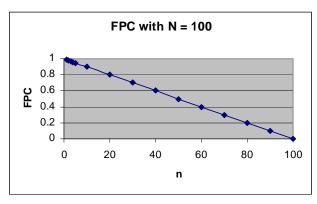
Standard error of
$$\hat{\mu}$$
 is: $SE(\hat{\mu}) = \sqrt{\left(\frac{N-n}{N}\right)\frac{s^2}{n}}$.

The quantity $\left(\frac{N-n}{N}\right)$ is the *finite population correction factor* which adjusts variance of the estimator

(not variance of the population which does not change with n) to reflect the amount of information that is known about the population through the sample. Practically, the correction factor reflects the proportion of the population that remains unknown. Therefore, as the sample size n approaches the

population size *N*, the finite population correction factor approaches zero, so the amount of variation associated with the estimate also approaches zero.

When the sample size n is large relative to the population size N, the fraction of the population being sampled n/N is small, so the correction factor has little effect on the estimate of variance (Fig. 2 - FPC.xls). If the finite population correction factor is ignored, including those cases where N is unknown,



the effect on the variance of the estimator is slight when N is large. When N is small, however, the variance of the estimator can be overestimated appreciably.

Estimating the Population Total

Like the population mean, the total number of entities in the population is another attribute estimated commonly. Unlike the population mean or proportion, estimating the population total requires that we know the number of sampling units in a population, *N*.

The population total $au=\sum_{i=1}^N y_i=N\mu$ is estimated with the sample total ($\hat{ au}$) which has an unbiased $N=\frac{n}{2}$

estimator:
$$\hat{\tau} = N\hat{\mu} = \frac{N}{n}\sum_{i=1}^{n}y_{i}$$

where N is the total number of sample units in a population, n is the number of units in the sample, and y_i is the value measured from each sample unit.

In studies of wildlife populations, the total number of entities in a population is often refereed to as "abundance" and is traditionally represented with the symbol N. Consequently, there is real potential for confusing the number of entities in the population with the number of sampling units in the sampling frame. Therefore, in the context of sampling theory, we'll use $\hat{\tau}$ to represent the population total and N to represent the number of sampling units in a population. Later, when addressing wildlife populations specifically, we'll use N to represent abundance to remain consistent with the literature in that field.

Because the estimator $\hat{\tau}$ is simply the number of sample units in the population N times the mean number of entities per sample unit, $\hat{\mu}$, the variance of the estimate $\hat{\tau}$ reflects both the number of units in the sampling universe N and the variance associated with $\hat{\mu}$. An unbiased estimate for the variance of the estimate $\hat{\tau}$ is:

$$\operatorname{var}(\hat{\tau}) = N^2 \operatorname{var}(\hat{\mu}) = N^2 \left(\frac{s^2}{n}\right) \left(\frac{N-n}{N}\right)$$

where s^2 is the estimated population variance.

Example: Estimating a caribou population in Alaska.

Caribou were counted in strip transects that were 1-mile wide. A simple random sample of 15 transects (n) were chosen from the 286 transects potentially available (N). The number of caribou counted were 1, 50, 21, 98, 2, 36, 4, 29, 7, 15, 86, 10, 21, 5, 4.

The sample mean number of caribou counted per transect: = 25.93

The sample variance: $s^2 = 919.0667$

The estimated variance of the sample mean: $\hat{var}(\bar{y}) = \left(\frac{286 - 15}{286}\right) \frac{919.07}{15} = 58.0576$

The estimated standard error of the mean is: $\sqrt{58.06} = 7.62$.

An estimate of the total number of caribou in the area is: $\hat{\tau}$ = 286(25.9333) = 7417

An estimate of variance of the estimated total is: $var(\hat{\tau}) = 286^2 (58.0576) = 4,748,879$

The estimated standard error of the total is: $\sqrt{4,748,879} = 2179$

Estimating a Population Proportion

If there is interest in the composition of a population, we could use a simple random sample to estimate the proportion of the population p that is composed of elements with a particular trait, such as the proportion of plants that flower in a given year, the proportion of juvenile animals captured, the proportion of females in estrus, and so on. We will consider only classifications that follow binomial trials which means that either an element in the population has the trait of interest (flowering) or not (not flowering) although extending this idea to more complex settings is straightforward.

In the case of simple random sampling, the population proportion follows the mean exactly; that is, $p = \mu$. If this idea is new to you, convince yourself by working through an example. Say we generate a sample of 10 elements, where 4 have a value of 1 and 6 have a value of 0 (1 = presence of a trait, 0 = absence of a trait). The proportion of the sample with the trait is 4/10 or 0.40 and so is the arithmetic mean, which = 0.40 ([1+1+1+1+0+0+0+0+0+0+0]/10 = 4/10). Cosmic.

It follows that the population proportion (p) is estimated with the sample proportion (\hat{p}) which has an unbiased estimator:

$$\hat{p} = \hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

Because we are dealing with dichotomous proportions (sample unit does or does not have the trait), the population variance σ^2 is computed based on variance for a binomial which is the proportion of the population with the trait (p) times the proportion that does not have that trait (1-p) or p(1-p). The estimate of the population variance s^2 is: $\hat{p}(1-\hat{p})$.

Variance of the estimate
$$\hat{p}$$
 is: $\hat{\text{var}}(\hat{p}) = \left(\frac{N-n}{N}\right) \frac{s^2}{n-1} = \left(\frac{N-n}{N}\right) \frac{\hat{p}(1-\hat{p})}{n-1}$.

Standard error of
$$\hat{p}$$
 is: $SE(\hat{p}) = \sqrt{\left(\frac{N-n}{N}\right)\frac{s^2}{n-1}} = \sqrt{\left(\frac{N-n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}}$.

Systematic Sampling

Occasionally, selecting sample units at random can introduce logistical challenges that preclude collecting data efficiently. If the chance of introducing a bias is low or if ideal dispersion of sample units in the population is a higher priority that a strictly random sample, then it might be appropriate to choose samples non-randomly. Like simple random sampling, systematic sampling is a type of probability sampling where each element in the population has a known and equal probability of being selected. The probabilistic framework is maintained through selection of one or more random starting points. Although sometimes more convenient, systematic sampling provides less protection against introducing biases in the sample compared to random sampling. Estimators for systematic sampling and simple random sampling are identical; only the method of sample selected differs. Therefore, systematic sampling is used to simplify the process of selecting a sample or to ensure ideal dispersion of sample units throughout the population.

Advantages:

- Easy to implement
- Maximum dispersion of sample units throughout the population
- Requires minimum knowledge of the population

Disadvantages:

- Less protection from possible biases
- Can be imprecise and inefficient relative to other designs if the population being sampled is heterogeneous

How it is implemented:

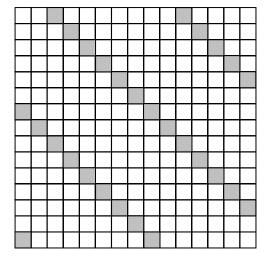
- Choose a starting point at random
- Select samples at uniform intervals thereafter

1-in-k systematic sample

Most commonly, a systematic sample is obtained by randomly selecting 1 unit from the first k units in the population and every k^{th} element thereafter. This approach is called a 1-in-k systematic sample with a random start. To choose k so than a sample of appropriate size is selected, calculate:

k = Number of units in population / Number of sample units required

For example, if we plan to choose 40 plots from a field of 400 plots, k = 400/40 = 10, so this design would be a 1-in-10 systematic sample. The example in the figure is a 1-in-8 sample drawn from a population of N = 300; this yields n = 28. Note that the sample size drawn will vary and depends on the location of the first unit drawn.



Estimating the Population Mean

The population mean (μ) is estimated with: $\hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n}$

The population variance (σ^2) is estimated with: $s^2 = \frac{\displaystyle\sum_{i=1}^n (y_i - \overline{y})^2}{n-1}$

Variance of the estimate $\hat{\mu}$ is: $\hat{\text{var}}(\hat{\mu}) = \left(\frac{N-n}{N}\right) \frac{s^2}{n}$.

Standard error of $\hat{\mu}$ is: $SE(\hat{\mu}) = \sqrt{\left(\frac{N-n}{N}\right)\frac{s^2}{n}}$.

Estimating the Population Total

The population total τ is estimated with: $\hat{\tau} = N\hat{\mu} = \frac{N}{n}\sum_{i=1}^n y_i$.

Variance of the estimate $\hat{\tau}$ is: $\hat{\text{var}}(\hat{\tau}) = N^2 \text{var}(\hat{\mu}) = N^2 \left(\frac{s^2}{n}\right) \left(\frac{N-n}{N}\right)$.

Standard error of $\hat{\tau}$ is: $var(\hat{\tau}) = \sqrt{N^2 \left(\frac{s^2}{n}\right) \left(\frac{N-n}{N}\right)}$

Estimating the Population Proportion

The population proportion (p) is estimated with the sample proportion (\hat{p}) which has an unbiased estimator:

$$\hat{p} = \hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

Because we are estimating a dichotomous proportion, the population variance σ^2 is again computed with a binomial which is the proportion of the population with the trait (p) times the proportion without that trait (1-p) or p(1-p). The estimate of the population variance s^2 is: $\hat{p}(1-\hat{p})$.

8

Variance of the estimate
$$\hat{p}$$
 is: $\hat{\text{var}}(\hat{p}) = \left(\frac{N-n}{N}\right) \frac{s^2}{n-1} = \left(\frac{N-n}{N}\right) \frac{\hat{p}(1-\hat{p})}{n-1}$.

How Many Samples?

Optimal allocation is an approach to maximize sampling efficiency; that is to provide high precision for a given amount of sampling effort.

A different question is how many samples should we take from the population?

First, establish the degree of precision required, B, the bound the error of estimation, which is the half-width of the confidence interval we wish to attain from sampling. Determine the sample size n required by setting $Z \times SE(\bar{y})$ equal to B and solving this expression for n.

Z is a constant that denotes the upper $\alpha/2$ point of the standard normal distribution where α is the same value used to establish the width of confidence intervals.

Population Mean

For simple random sampling, set:
$$B = Z\sqrt{\left(\frac{N-n}{N}\right)\frac{\sigma^2}{n}}$$

solve for
$$n$$
 to get: $n = \frac{1}{\frac{1}{n_0} + \frac{1}{N}}; n_0 = \frac{z^2 \sigma^2}{B^2}$ or $n = \frac{1}{\frac{1}{z^2 \sigma^2} + \frac{1}{N}}$.

Note that if n will be small relative to N, the population correction factor can be ignored, and the formula for sample size reduced to n_0 .

Example: Estimate the average body mass of male freshman μ on campus.

Assume that no prior information exists with which to estimate population variance σ^2 but we know that the mass of most male freshmen is within a range of about 100 pounds and there are N = 1000 students.

How many samples are needed to estimate μ with a bound on the error of estimation B = 3 pounds using simple random sampling?

Although it is best to have data with which to estimate σ^2 , perhaps from a small pilot study, the range is often approximately equal to 4 σ , so one-fourth of the range might be used as an approximate value of

$$\sigma$$
: $\sigma \approx \frac{range}{\Delta} = \frac{100}{\Delta} = 25$.

Substituting:
$$n = \frac{1}{\frac{1}{2^2 25^2} + \frac{1}{1000}} = \frac{1}{\frac{1}{277.78} + \frac{1}{1000}} = \frac{1}{0.0036 + 0.001} = 217.4$$

Therefore, about 218 samples are needed to estimate μ with a bound on the error of estimation B=3 pounds.

Population Total

For simple random sampling, solve for n from: $B = Z\sqrt{N(N-n)\frac{\sigma^2}{n}}$

$$n = \frac{1}{\frac{1}{n_0} + \frac{1}{N}}; n_0 = \frac{N^2 z^2 \sigma^2}{B^2} \text{ or } n = \frac{1}{\frac{1}{N^2 z^2 \sigma^2} + \frac{1}{N}}.$$

Again, if N is large relative to n, the population correction factor can be ignored, and the formula for sample size reduced to n_0 .

Example: What sample size is necessary to estimate the caribou population we examined to within B = 2000 animals of the true total with 90% confidence ($\alpha = 0.10$).

Using s^2 = 919 from earlier and Z = 1.645, which is the upper α = 0.10/2 = 0.05 point of the normal

distribution:
$$n_0 = \frac{286^2 \cdot 1.645^2 \cdot 919}{2000^2} \approx 51$$

To adjust for finite population size:
$$n = \frac{1}{\frac{1}{51} + \frac{1}{286}} \approx 44$$

Stratified Random Sampling

The way we have selected sample units thus far has required that we know little about the population of interest in advance of selecting the sample. This approach only works best when the characteristic of interest is relatively homogeneous across the population. If, however, the characteristic is heterogeneous, then estimates based on these designs will be imprecise. If we have ancillary information that is associated with the heterogeneity in the population, we can use using alternate designs to select samples which will yield increased precision for a fixed amount of effort. The first of these designs is stratified random sampling.

A stratified random sample is one obtained by dividing the population elements into mutually exclusive, non-overlapping groups (strata), then selecting a simple random sample from within each stratum (stratum is singular for strata). Every potential sample unit can be assigned to only one stratum and no units can be excluded.

Stratifying involves classifying sampling units of the population into relatively homogeneous groups, usually before selecting sample units. Strata are based on information other than the characteristic being measured that is known to or thought to vary with the characteristic of interest in such a way that the characteristic is more homogeneous within strata than among strata. Therefore, any feature that explains variation in the characteristic of interest can be used as a basis for stratifying. For example, if our goal is to estimate the total number of agaves in an area and we know from previous work that agave abundance varies with soil type, we might choose to stratify the population by soil type. Because samples within strata are likely to be more similar than those among strata, sampling error will be lower and estimates generated within strata will have higher precision than simple random samples drawn from the same population.

As most ecological systems are heterogeneous, stratifying is a common approach for increasing precision in ecological studies. Common strata in ecological studies include elevation, aspect, or other geographic features for studying plant communities and vegetation communities for studying animal communities. When choosing among potential strata, you should seek to minimize variation within strata and maximize variation among strata.

Stratified random sampling is appropriate whenever there is heterogeneity in a population that can be classified with ancillary information; the more distinct the strata, the higher the gains in precision. The same population can be stratified multiple times simultaneously.

Advantages:

- Higher precision of estimates
- More efficient
- Separate estimates for each stratum

Disadvantages:

- Requires ancillary information
- Can be more time consuming to plan and implement

How it is implemented:

- Divide the entire population into non-overlapping strata
- Selected a simple random sample from within each strata

L = number of strata

 N_i = number of sample units within stratum i

N = number of sample units in the population

Estimating the Population Mean

Estimates from stratified random samples are simply the weighted sum of estimates from a series of simple random samples, each generated within a unique stratum. This should be apparent in the estimators below, such as that for the population mean, which is an average of the means from each stratum weighted by the number of sample units measured within each stratum. With only one stratum, stratified random sampling reduces to simple random sampling.

The population mean (μ) is estimated with:

$$\hat{\mu} = \frac{1}{N} (N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2 + \dots + N_L \hat{\mu}_L) = \frac{1}{N} \sum_{i=1}^L N_i \hat{\mu}_i$$

Variance of the estimate $\hat{\mu}$ is again just a weighted average of estimates from a series of random samples, although it looks a bit cumbersome:

$$\hat{\text{var}}(\hat{\mu}) = \frac{1}{N^2} \left[N_1^2 \left(\frac{N_1 - n_1}{N_1} \right) \left(\frac{s_1^2}{n_1} \right) + \dots + N_L^2 \left(\frac{N_L - n_L}{N_L} \right) \left(\frac{s_L^2}{n_L} \right) \right] = \frac{1}{N^2} \sum_{i=1}^L N_1^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{s_i^2}{n_i} \right)$$

$$\text{Standard error of } \hat{\mu} \text{ is: } SE(\hat{\mu}) = \sqrt{\frac{1}{N^2} \sum_{i=1}^L N_1^2 \! \left(\frac{N_i - n_i}{N_i} \right) \! \left(\frac{s_i^2}{n_i} \right) } \, .$$

Estimating the Population Total

Like the mean, estimating a total for a stratified random sample is a matter of summing individual estimates of the total from each stratum, $N_i \hat{\mu}_i$.

The population total
$$\tau$$
 is estimated with: $\hat{\tau} = N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2 + \dots + N_L \hat{\mu}_L = \sum_{i=1}^L N_i \hat{\mu}_i$

Variance of the estimated total
$$\hat{\tau}$$
 is: $\hat{\text{var}}(\hat{\tau}) = N^2 \hat{\text{var}}(\hat{\mu}) = \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i}\right) \left(\frac{s_i^2}{n_i}\right)$

Standard error of $\hat{\tau}$ is the square root of $\hat{var}(\hat{\tau})$.

Estimating the Population Proportion

Estimating the proportion of the population with a particular trait (*p*) using stratified random sampling involves combining estimates from multiple simple random samples, each generated within a stratum. The population proportion is estimated with the sample proportion:

$$\hat{p} = N_1 \hat{p}_1 + N_2 \hat{p}_2 + \dots + N_L \hat{p}_L = \sum_{i=1}^{L} N_i \hat{p}_i$$

Variance of the estimate \hat{p} is:

$$\hat{\text{var}}(\hat{p}) = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \hat{\text{var}}(\hat{p}_i) = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{\hat{p}_i (1 - \hat{p}_i)}{n_i - 1} \right)$$

Standard error of \hat{p} is the square root of $var(\hat{p})$.

Example: Simple example of 12 samples taken from a population of 41 entities.

| Stratum (i) | N _i | n _i | \overline{y} | S _i ² |
|-------------|----------------|----------------|----------------|-----------------------------|
| 1 | 20 | 5 | 1.6 | 3.3 |
| 2 | 9 | 3 | 2.8 | 4.0 |
| 3 | 12 | 4 | 0.6 | 2.2 |

Estimate of the population mean: $\overline{y} = \frac{1}{41} [20(1.6) + 9(2.8) + 12(0.6)] = \frac{1}{41} [64.4] = 1.57$ Estimate of the population total = 41 × 1.57 = 64.4.

Estimated variance of the estimated population mean is:

$$\hat{\text{var}}(\overline{y}) = \frac{1}{41^2} \left[20(20 - 5) \frac{3.3}{5} + 9(9 - 3) \frac{4.0}{3} + 12(12 - 4) \frac{2.2}{4} \right] = \frac{322.8}{41^2} = 0.192$$

Estimated variance of the estimated population total = $412 \times 0.192 = 322.8$.

Allocating Sampling Effort among Strata

After deciding to use stratify random sampling, we need to decide how to divide sampling effort among different strata; that process is called **allocation**. When deciding where to expend effort, the question becomes how best to allocate sampling effort among strata so that the sampling process will be the most efficient balance of effort, cost, and precision. Should we allocate the same sampling effort to

each stratum? If strata are of different sizes, as is usually the case, should we allocate more effort to larger stratum?

There are many strategies for allocating sampling effort, and the more information available about the population of interest, the more efficient the allocation strategy can be. Information on the variability of samples within each stratum, the relative cost of obtaining a sample from each stratum, and the number of sample units in each stratum can all help to increase sampling efficiency. Some of the most common allocations strategies are uniform, proportional to size, variation, and cost, and optimal, which simultaneously considers size, variation, and cost or whichever combination of those is available. All strategies function by create a simple proportional multiplier by which a fixed number of samples can be allocated among strata.

Uniform Allocation

The simplest allocation strategy is to select the same number of samples from each stratum, which is an ideal approach if there is no information available about variability of units within strata, the cost of sampling is similar for all strata, and strata are of similar size.

Allocation Proportional to Size or Variation

The number of sample units to select from each stratum can be made proportional to the number of sample units (or size) within each stratum. Variation in a stratum often increases with a the size of a stratum, so in some cases this approach can be considered as a rough approach for allocating more effort to strata that are likely to be more variable strata. To allocation proportional to stratum size:

$$n_{i} = n \left(\frac{N_{i}}{\sum_{i=1}^{L} N_{i}} \right) = n \left(\frac{N_{i}}{N} \right)$$

To allocation proportional to the amount of variation among elements within each stratum, as measured by the estimated standard deviation within each stratum:

$$n_i = n \left(\frac{s_i}{\sum_{i=1}^{L} s_i} \right)$$

This approach relies on estimates generated from a previous study or alternatively by the ability to gauge relative differences in variation among strata, such as expecting one stratum to have 1.5 times the variation as another stratum.

Optimal Allocation

Both allocation approaches above are special cases of the optimal allocation strategy which estimates the population mean or total with the lowest variance for a given sample size in stratified random

sampling. The number of samples selected from each stratum is proportional to the size, variation, as well as the cost (c_i) of sampling in each stratum. More sampling effort is allocated to larger and more variable strata, and less to strata that are more costly to sample.

$$n_{i} = n \left(\frac{\frac{N_{i} s_{i}}{\sqrt{c_{i}}}}{\frac{\sum_{k=1}^{L} N_{k} s_{k}}{\sqrt{c_{k}}}} \right)$$