CALCULATION WORKSHEET: CONSTUCT A CONFIDENCE INTERVAL

2023 EDITION

phase
pollinator habitat aligned with solar energy

Constructing a confidence interval helps you communicate the uncertainty around parameter estimates from your data. Equations are provided below for the following strategies:

1. Simple Random Sample,
2. Two-Stage Cluster Sample, and
3. Stratified Random Sample

1. Simple Random Sample

A simple random sample is constructed when you locate plots at random within your solar site, where any given area has an equal probability of being included. A systematic sample may also use these equations.

Estimate of the mean

An unbiased estimator for the mean is simply the sum of all values divided by the number of samples.

$$\bar{x} = \frac{\sum x_i}{n}$$ ***Equation 1***

where:

$\sum x_i$ the sum of all observations, where x_i is the ith observation
n number of plots assessed

Construct a confidence interval

The estimate of the mean is very likely not exactly the same as the true mean of the population. The confidence interval provides a range of plausible values for the mean value.

First, we calculate the standard deviation (s^2):

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{(n-1)}$$ ***Equation 2***

where:

x_i the ith observation
$\bar{x}$ the sample mean
n number of plots assessed

Next, we calculate the standard error (SE):

$$SE = \sqrt{\left(1 - \frac{n}{N}\right) * \frac{s^2}{n}}$$ ***Equation 3***

where:

s^2	the sample standard deviation
n	number of plots assessed
N	total number of plots possible (divide the total acres in your monitoring plan's geographic scope by 0.0344, the size of a plot in acres)

Finally, we calculate the confidence interval:

$$CI_{min} = \bar{x} - SE * CV$$
$$CI_{max} = \bar{x} + SE * CV$$ ***Equation 4***

where:

$\bar{x}$	is the sample mean
SE	standard error of the sample
CV	Critical Value (see discussion below)

If the number of plots in the sample is greater than 30, you may use a z-score for the Critical Value, select the z-score (Z_α) from the table below corresponding to the confidence level specified in your sampling objective. If less than 30 plots were assessed, the critical value must be drawn from a t-table based on the degrees of freedom in the sample.

Table of standard normal deviates (Z_α) for various confidence levels

Confidence level	Alpha (α) level	Z_α
60%	0.4	0.84
80%	0.2	1.28
90%	0.1	1.64
95%	0.05	1.96
99%	0.01	2.58

2. Two-Stage Cluster Sample

A two-stage cluster sample includes a simple random sample of primary sampling units followed by a simple random sample of plots within the primary sampling units. Primary sampling units can be any method of dividing up your rights-of-way, for example by panel number. If the primary sampling units are not the same size, you must select primary units with the probability proportional to size.

Estimate the Mean

$$\bar{x} = \left(\frac{N}{N * M}\right) * \sum M_h * \bar{x}_h$$ ***Equation 5***

where:

N	the total number of clusters
M	total number of plots possible (divide the total acres in your monitoring plan's geographic scope by 0.0344, the size of a plot in acres)
M_h	the number of plots possible in cluster *h* (divide the total acres in the cluster by 0.0344, the size of a plot in acres)
$\bar{x}_h$	mean of cluster *h*

Note that if all clusters are approximately the same size and the number of plots assessed in each cluster is the same, you can simply average all parameter values to calculate the mean.

Construct a Confidence Interval

The confidence interval of a two-stage cluster sample must include both the variance between clusters and the variance within clusters. First, we need to calculate the variance within each cluster. Repeat this calculation for each cluster:

$$s^2{}_h = \frac{\sum(x_{ih} - \overline{x_h})^2}{(m_h - 1)}$$ ***Equation 6***

where:

x_{ih}	the ith observation from cluster *h*
$\overline{x_h}$	is the sample mean from cluster h
m_h	number of plots assessed in cluster h

We also need to estimate the total for the population:

$$t = \frac{N}{n} * \sum M_h * \overline{x_h}$$ ***Equation 7***

where:

N	the total number of clusters
n	number of clusters in the sample
M_h	the number of plots possible in cluster *h* (divide the total acres in the cluster by 0.0344, the size of a plot in acres)
$\overline{x_h}$	is the sample mean from cluster h

Next, we calculate the standard error (SE):

$$SE = \left(\frac{1}{M}\right) * \sqrt{\left(\frac{N^2}{n} * \left(1 - \frac{n}{N}\right)\right) * \frac{\sum(M_h * \overline{x_h} - \frac{t}{N})^2}{(n-1)} + \left(\frac{N}{n}\right) * \sum\left(1 - \frac{m_h}{M_h}\right) * \frac{{M_h}^2 * {s^2}_h}{m_h}} \qquad \textbf{\textit{Equation 8}}$$

where:

N	the total number of clusters
M	total number of plots possible (divide the total acres in your monitoring plan's geographic scope by 0.0344, the size of a plot in acres)
n	number of clusters in the sample
M_h	the number of plots possible in cluster *h* (divide the total acres in the cluster by 0.0344, the size of a plot in acres)
t	the estimate of the total for the population
$\overline{x_h}$	is the sample mean from cluster h
m_h	number of plots assessed in cluster h
${s^2}_h$	variance within cluster h

Finally, we calculate the confidence interval in the same way as for a simple random sample. See *Equation 4* above.

Stratified Random Sample

A stratified random sample is a random sample drawn from multiple strata.

Estimate the Mean

$$\bar{x} = \sum \frac{N}{N_h} * \overline{x_h}$$ ***Equation 9***

where:

N	total number of plots possible (divide the total acres in your monitoring plan's geographic scope by 0.0344, the size of a plot in acres)
N_h	the number of plots possible in stratum *h* (divide the total acres in the stratum by 0.0344, the size of a plot in acres)
$\bar{x}_h$	mean of stratum *h*

Construct a Confidence Interval

First, we need to calculate the variance for each stratum. Repeat the below calculation for each stratum:

$$s^2{}_h = \frac{\sum(x_{ih} - \overline{x_h})^2}{(n_h - 1)}$$ ***Equation 10***

where:

x_{ih}	the ith observation from stratum *h*
$\overline{x_h}$	is the sample mean from stratum h
n_h	number of plots assessed in stratum h

Next, we calculate the standard error (SE):

$$SE = \left(\frac{1}{N}\right) * \sqrt{\sum N_h{}^2 * (1 - \frac{n_h}{N_h}) * \frac{s^2{}_h}{n_h}}$$ ***Equation 11***

where:

N	total number of plots possible (divide the total acres in your monitoring plan's geographic scope by 0.0344, the size of a plot in acres)
N_h	the number of plots possible in stratum *h* (divide the total acres in the stratum by 0.0344, the size of a plot in acres)
$s^2{}_h$	variance of stratum h

Finally, we calculate the confidence interval in the same way as for a simple random sample. See *Equation 4* above.

Acknowledgements

This guidance has been compiled by individuals from the following organizations.

Cover photo by Lee Walston, Argonne National Lab

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number DE-EE0009371.