Eco Internal Publication #ECO-10-430-031202

A Review of Sampling Methods Available for Environmental Workers

by

Zurab Khviengia

Research and Development, Eco & Associates, Inc.

1855 W. Katella Ave, Suite 350, Orange, CA 92867

zkhviengia@ecoinc.info

Phone: 1(714)289-0995 Fax: 1(714)289-0965

Abstract

This is a brief review of various sampling methods and associated with them errors that are most frequently used in the environmental work. Basic notions of inferential statistics are explained in layman's terms for those who are not well acquainted with statistics. The concepts of hypothesis testing and confidence intervals are discussed to some extent. Incremental sampling method as an example of the composite sampling is given special consideration, its advantages and disadvantages are discussed. A few suggestions are made as to how statistical research can be improved in the future to serve better practical needs of project managers. .

1 Inferential Statistics – Probability Based Inference

1.1 Sampling distribution

A sampling distribution is a foundation of all inferential statistics. If a variable does not have the sampling distribution, it cannot be analyzed using methods of inferential statistics. The sampling distribution is a distribution of a certain parameter (mean, proportion, etc.) under the assumption of randomness of the sample for which it is calculated.

For example, mean and proportion are normally distributed independently of the distribution of the variable itself. This is a simplified statement of the central limit theorem in statistics. The sampling distribution has nothing to do with the distribution of an actual variable and is most of the time normal even of the variable is not distributed normally.

1.2 Confidence Intervals

Once a sample is collected and the mean is calculated, the population mean is approximated as the sample mean. How confident are we that the sample mean describes the population mean? This question has no answer, but the question what is the interval that can capture population mean with certain probability, has the answer and the answer is confidence interval.

Confidence interval is equal to the sampling error (SE) times the parameter that will depend on our choice of the confidence level. The higher confidence we require, the larger the interval.

The formula for the confidence interval is

$$\mu = \overline{x} \pm t\, SE$$

The sampling error is smaller if the sample size is larger and the variability of the data is lower.

2 Sampling

2.1 A Brief Description of the Sampling Methodology

Sampling is an inherent part of all statistical research based on the methods of inferential statistics. Inferential statistics is the only branch of statistics that provides probability based answers to specific questions frequently raised in business and environmental investigations.

Sampling research is based on the mathematical notion that if random, a sample can provide a foundation for the inference about the entire population that the sample is taken from. Based on a sample, a researcher with a certain probability arrives at a conclusion about a specific population parameter, like the mean, proportion, or other. These quantities for the population can be estimated by the corresponding quantities obtained from the sampling data.

Denoting the sampling mean by $\bar{x}$, one can write the population mean as:

$$\mu = \bar{x} \pm ½\, CI,$$

meaning that with a ceratin probability the population mean will fall within the interval called confidence interval (CI). Depending on the probability chosen the interval wil be different. Typical values for the probability are 95%, 98%, 99%. We will choose 95% CI as the means to estimate our population mean.

2.2 Hypothesis Testing

Another means to obtain information about population parameters is hypothesis testing. In this case instead of trying to estimate the population parameter, one attempts to test a certain statement about the population parameter, based on the corresponding sampling parameter. Thus, if the statement is made about the population mean, the sampling mean is calculated to test that statement. Here instead of the confidence interval probability (confidence level) one sets an error, a level of significance, which one is comfortable of having when the conclusion is made about the statement's validity. Typical values of the error are, 5%, 2%, 1%. The less the error, the higher is the certainty.

In all these methodologies the key step is sampling.

3 Sampling Methods

For any result to be valid a sample should be random.

There are only four fundamental ways to collect a random sample. All other ways are some combinations of these four. The four fundamental sampling methods are: simple random sample, stratified sample, clustered sample and systematic sample.

3.1 Simple random sample

A simple random sample is collected by selecting randomly units from the entire population. Each unit will have to have an equal probability of being selected into the sample.

3.2 Stratified Sample

A stratified sample is collected by first dividing the entire population into groups, the strata, and then collecting a simple random sample from each stratum.

3.3 Clustered Sample

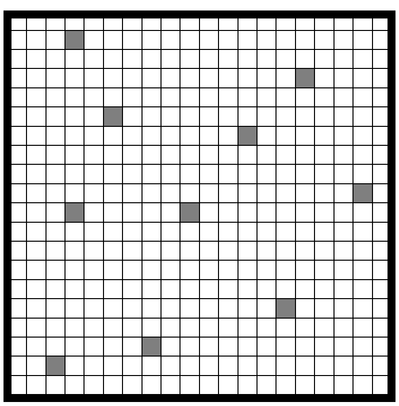
A clustered sample is collected by first dividing the population into groups, the clusters, and then selecting several clusters randomly. In general, there are no differences between the strata and clusters, but the criteria by which they are identified are different: the strata are formed by the principle of similarity, while the clusters are formed by their ability to reflect the whole on the smaller scale.

3.4 Systematic Sample

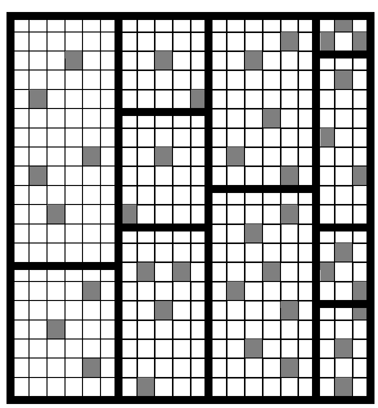
A systematic sample is collected first by dividing the population into equal size groups and then selecting the first unit of the sample randomly from the first group. All other units of the sample are collected non-randomly, and are corresponding units from all other groups.

3.5 Two-Stage Clustered Sampling

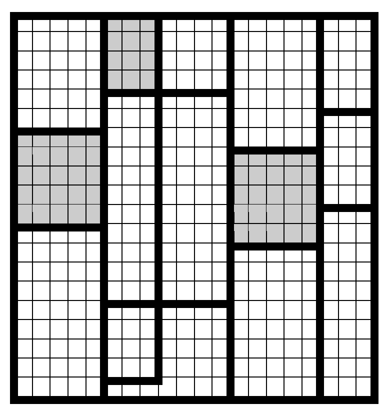
A sample in this case is collected by first dividing the population into clusters and then collecting simple random samples from randomly selected clusters.



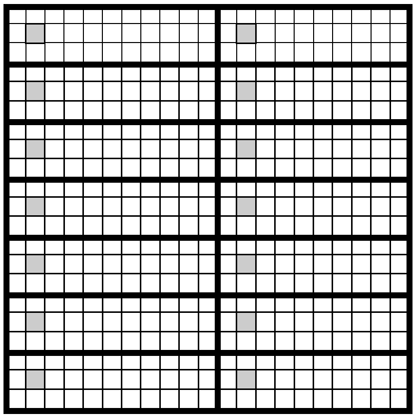
Simple Random Sample



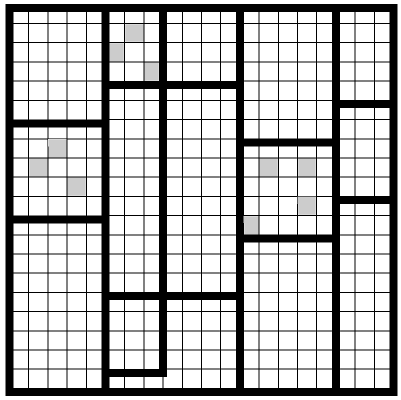
Stratified Sample



Clustered Sample



Systematic Sample



Two-Stage Clustered Sample
(stratified clustered sample)

Fig.1. Sampling methods

3.6 Other Sampling Methods

All other sampling methods are combinations of the above fundamental sampling methods.
If feasible for estimating the population parameters the best choice for sampling method is always a simple random sample.

3.7 Understanding Incremental Sampling

Incremental sampling is a composite sampling based on the systematic random sample. A composite sampling is not a sampling method but rather a procedure of combining a random sample into one composite ‘sample’, which mathematically equates to only the measuring of the mean and forgetting about the variability within the sample. This way one looses all information about the sample except its mean characteristics.

4 A Theory behind Incremental Sampling

Incremental sampling is a procedure of sampling reflecting a well known topic in mathematical statistics known as functions of random variables.

Random sampling conforms to sampling distribution: its parameters are distributed according to the well defined distribution; the mean for example is distributed normally.

Incremental sampling conforms to what we shall call the mean sampling distribution: this is a distribution of the mean values of the incremental samples, which themselves are the means of the random sample. The mean of the means that is how we can call it. Mean sampling distribution is also normal for the means.

Denoting by x_1^i, x_2^i, x_3^i , the i-th random sample for the replicates 1, 2, 3. (A replicate is a repetitive taking a systematic sample out of which an incremental sample is formed), one can write:

$$\mu \in [\,\overline{x}_1 - t_{n-1}\frac{s_1}{\sqrt{n}}, \overline{x}_1 + t_{n-1}\frac{s_1}{\sqrt{n}}]$$
$$\mu \in [\,\overline{x}_2 - t_{n-1}\frac{s_2}{\sqrt{n}}\,, \overline{x}_2 + t_{n-1}\frac{s_2}{\sqrt{n}}]$$
$$\mu \in [\,\overline{x}_3 - t_{n-1}\frac{s_3}{\sqrt{n}}\,, \overline{x}_3 + t_{n-1}\frac{s_3}{\sqrt{n}}]$$
$$\mu \in [\,\overline{x} - t_{3-1}\frac{s}{\sqrt{3}}\,, \overline{x} + t_{3-1}\frac{s}{\sqrt{3}}]$$

Where first three are the confidence interval estimates for the population mean, based on sample 1, 2, and 3, while the last equation gives the estimate for the population mean, based on the incremental sampling. where

$$s^2 = \frac{s_1^2}{n} + \frac{s_2^2}{n} + \frac{s_3^2}{n}$$

The confidence interval for incremental sample depends on the standard deviations of each replicate sample, the number of increments n, and the number of replicates r. Confidence interval decreases with an increase of r as well as with an increase of n. for low variation between the increment sample data, the rate at which the confidence interval decreases with the increase of r is considerably lower for $r \geq 6$ and the confidence interval becomes small enough for a given r for $n \geq 30$ (see Fig.1 below).

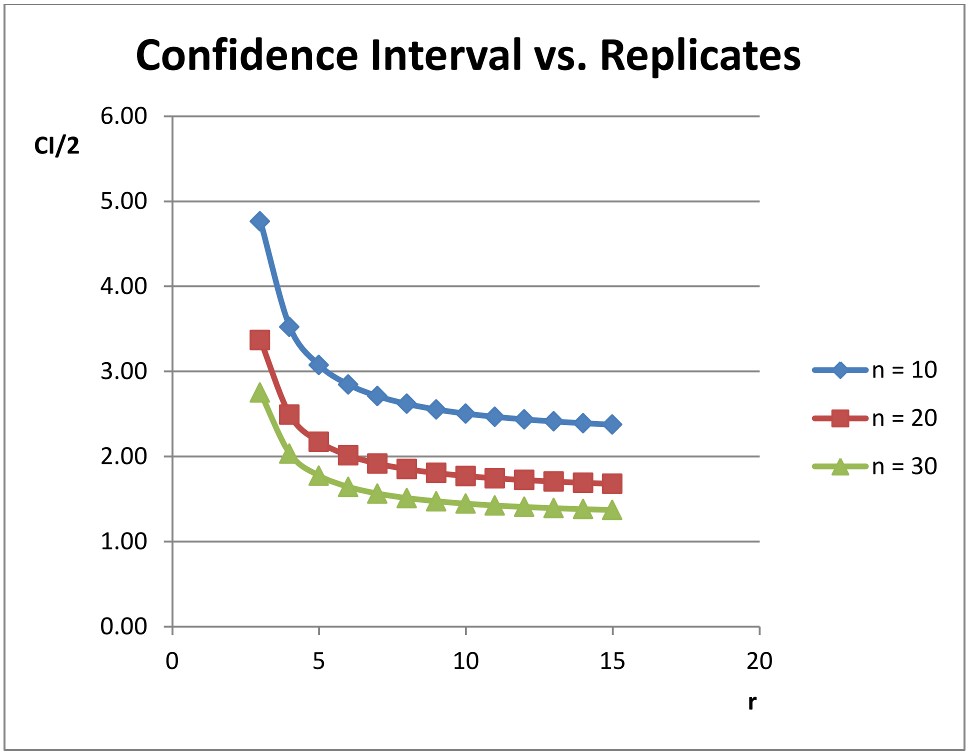


Fig.2. Incremental sample confidence interval

It is a simple mathematical analysis to see what gain we have by sacrificing the information about the samples and keeping only their means. Preliminary investigation shows that the confidence interval for the incremental sample is always larger than the confidence interval for individual samples, which also means that the mean of the means is less precise estimate of the true mean than individual means. (Being more precise, confidence interval for the incremental sample becomes approximately equal to the confidence interval for an average individual sample, - that is a sample with an average standard deviation among all individual samples, - only when number of replicates becomes equal to the number of increments (r = n), in which case there is no reason why one should use an incremental sample instead of an individual random sample). The only gain that the incremental sample gives is cost.

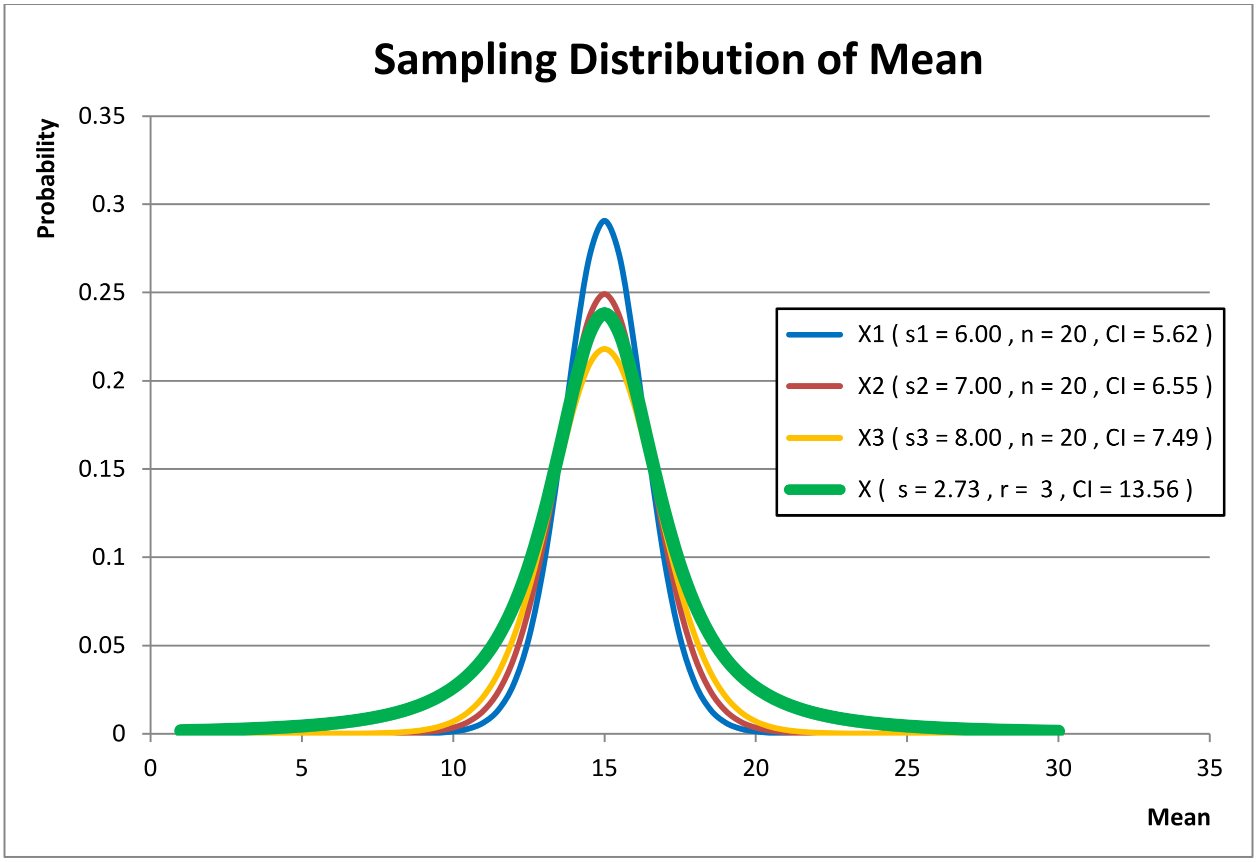


Fig.2. Sampling distribution of incremental sample

This procedure is frequently used in the environmental research due to its cost effectiveness.

Note: In mathematical notations one can introduce a function f(x) of an outcome of random events and define an event as a sample of size n for a random variable x, for example. Then f(x) = f(x1, x2, . . . , x10) where xi is the i-th outcome of an event. For function f one can derive various characteristics. The simplest ones would be the mean, standard deviation, variance, and so on. If

$$f(x) = \sum x_i/\text{n} \text{ , i= 1,2, . . . , n}$$

then for N events, the mean f(x) will be

$$\overline{f} = \sum_{i=1}^{N} f^i /N$$

where fi is the value of f(x) for the event I, and the variance will be

$$s^2(f) = \sum_{i=1}^{N} s^2\,(i)/n$$

This simple mathematical exercise illustrates how the properties of the functions of the outcomes of random events can be studied.

5 How random is random

A discrete random variable is a mathematical notion and denotes a variable that can take multiple values, one in each event, and the ratio of the number of times a given value is taken to the total number events is a well defined fraction when the number of events is infinite. This is a theory, but in practice infinity is unreachable abstraction. No one can claim what the probability is unless they have tried an infinity number of times, and no one has. This makes a notion of randomness an unreachable abstraction.

However, without using this abstract notion, the entire branch of statistics called inferential statistics will lose its significance. We shall always assume that random number generators provide a valid random sample and will never question the randomness of these numbers.

Random numbers are random not because they have been picked randomly but because if we keep picking them according to the same rule infinity times, they will display a random nature. “Picking randomly” even in mathematics is an unexplainable abstraction.

6 Confidence Intervals for Various Sampling Methods

Simple Random Sample

$$\mu = \overline{x} \pm t_{n-1}\frac{s}{\sqrt{n}}\sqrt{(N-n)/(N)}$$

Stratified Sample

$$\mu = \overline{x} \pm t_{df}\frac{s}{\sqrt{n}}$$

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{L} N_i\overline{x}_i$$

$$s^2 = \frac{1}{N^2}\sum_{i=1}^{L}\frac{N_i(N_i - n_i)}{n_i}(s_i)^2$$

$$df = \frac{\left(\sum_{i=1}^{L} a_i s_i^2\right)^2}{\sum_{i=1}^{L}\{(a_i s_i^2)^2 / (n_i - 1)\}}$$

$$a_i = \frac{N_i(N_i - n_i)}{n_i}$$

$$s_i = \sum_{j=1}^{n_i}\frac{(x_{ij} - \overline{x}_i)^2}{n_i - 1}$$

Clustered Sample

$$\mu = \overline{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$

$$\overline{x} = \frac{1}{\sum_{i=1}^{n} M_i}\sum_{i=1}^{n} M_i \overline{x}_i$$

$$s^2 = \frac{(N - n)}{N\sum_{i=1}^{n} {M_i}^2}\sum_{i=1}^{n}(s_i)^2$$

Systematic Sample

Type 1. When only the first unit is chosen randomly
(Is equivalent to the simple random sample)

$$\mu = \overline{x} \pm t_{n-1}\frac{s}{\sqrt{n}}\sqrt{(N - n)/(N)}$$

Type 2. When first several units are chosen randomly and the pattern is repeated.
(Can be treated as a type of a two-stage clustered sample with equal clusters)

$$\mu = \overline{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}\overline{x}_i$$

$$s^2 = \frac{(N-n)}{NnM^2} s_u^2$$

$$s_u^2 = \sum_{i=1}^{n} \frac{(\overline{x}_i - \overline{x})^2}{n-1}$$

Two-stage Clustered Sample

$$s^2 = \frac{(N-n)}{N \sum_{i=1}^{n} {M_i}^2} s_u^2 + \frac{1}{N \sum_{i=1}^{n} {M_i}^2} \sum_{i=1}^{n} \frac{M_i(M_i - n_i)}{n_i} (s_i)^2$$

7 Non-Probability Sampling

Highly discouraged by all manuals in statistics, non-probability sampling can be a viable alternative to probability sampling if the sample is chosen with the knowledge of a particular application, and instead of attempting to make inference one makes assumptions and attempts to calculate total (population) parameters based on the sampling results.

In many cases, non-probability sampling will give better results than methodologies of inferential statistics. This alternative should seriously be considered by all result oriented managers in their decision-making process.

The best type of non-probability sampling is selective sampling.

8 Practical Application of Sampling

8.1 Consecutive Incremental Sampling

Incremental sampling was proposed as a low cost alternative to more conventional procedures of sampling. Its main application lies in the estimating in statistically acceptable terms the mean concentration over a certain area. The estimate of the mean concentration is used to make a conclusion about compliancy to the acceptable level of contamination. Toward this goal this brief commentary is written.

Incremental sampling makes sense only when the number of increments is greater than 2. We recommend that in all practical applications first three incremental samples be taken. If the mean contamination is above the compliance threshold, there is no need to continue sampling. If, on the other hand, the mean contamination of incremental sample is below the compliance threshold, one has to calculate confidence interval to ascertain the statistical significance of this result. If the UCL (upper confidence level) is below the threshold, we can conclude that at a given confidence level, the site mean contamination is less than the threshold value. Otherwise, more sampling needs to be done to decrease the confidence interval by increasing the number replicates and hopefully decreasing the variability of the data. We recommend that two more replicates be taken and the results looked at again. If the UCL computed is below the threshold level, the sampling process will stop and the decision will be made. If the UCL is still above the threshold level, then the decision has to be made as to whether the additional sampling needs to be done in the attempt to reduce the UCL or the existing result assumed to be true and the site considered to be contaminated.

The above described consecutive incremental sampling is recommended every time the contamination of the site is studied for the compliance to the threshold levels.

8.2 On The Sample Size

The question that statistics can never answer in the beginning of investigation when there is no data available yet, is the question that every practical worker asks in the beginning of the investigation: how many samples do I need to take?

If one forgets about statistics, people say: take as many as you can to get as good a result as you can. Not a very practical advice, although conceptually true.

However, we propose a two-tier sampling plan to address the issue of the most appropriate sample size for a specific problem at hand. In the beginning, a preliminary simple sampling should be preformed mostly to collect the information about variability of the data. Here we propose selective sampling instead of the more common random sampling. During this selective sampling procedure, a judgment call has to be made about the highest and the lowest contaminated areas on the site, and a small sample, of size 3 to 5, has to be taken from those areas. The data thus collected will provide the basis for estimate of variability (upper limit) which can be used to predict the actual sample size for the main sampling event. Assuming that the standard deviation from the first tier selective sampling data of size n is s, one can estimate the sample size by requiring the width of the confidence interval, or a t times the standard error, to be of a certain magnitude tSE. The most reasonable requirement would be for a relative size of the tSE, relative to the mean that is, to be estimated. We shall call it tRSE=SE/mean.

$$n = \frac{1}{\frac{1}{t_{n-1}^2}\left(\frac{tRSE}{RSD}\right)^2}$$

The above formula should be the basis in determining the sample size for simple random sample. If tRSE is chosen to be equal to RSD, the sample size is approximately equal to 6. The less the tRSE is than RSD, the higher the sample size. Below is given a chart that can be used to find a sample size for various values of tRSE/RSD. One has to select a corresponding graph and find its intersection with the red line of sample size. The point of intersection will provide the value for n. For example if we choose the half of the confidence interval width (RSE) to be half of the relative standard deviation of the selective sample ($(tRSE/RSD)^2 = 1/4$), the sample size will be 18.

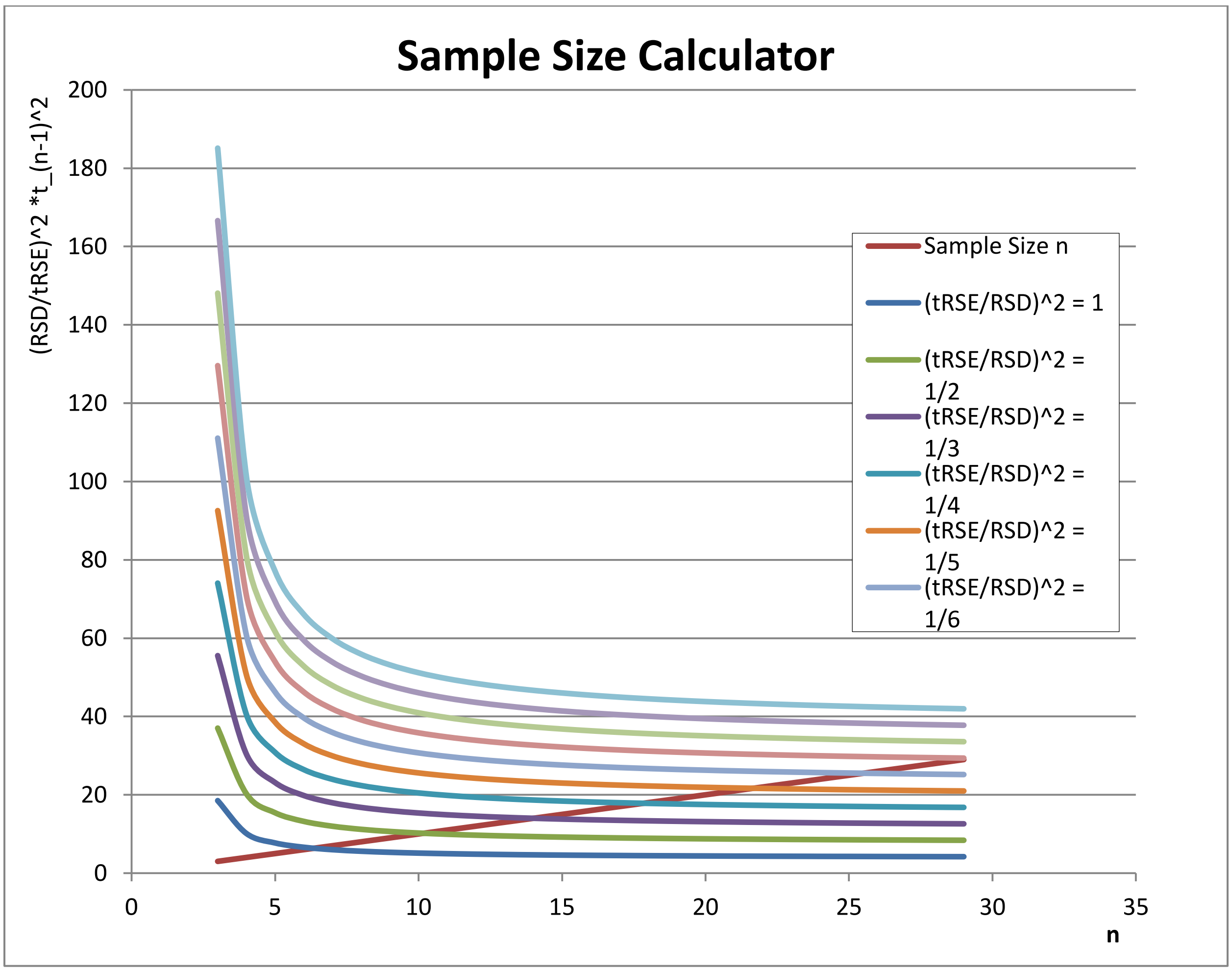


Fig.4. Sample Size Calculator

In case when the data from the site is available, the sample size for the next round should be determined using the actual data. The calculation and the way pf finding the sample size will be identical to the above described. If both the selective sample data and the actual data are available, RSD for actual data should be used.

The above procedure can be applied to the determination of the sample size within each strata or cluster for stratified and two-stage clustered sampling methods. In this case a sample size found will be the sample size for that individual stratum or a cluster for which the selective sampling was conducted.

If incremental sampling is used, then the sample size and the number of replicates can be found using the following formula:

$$n = t_{r-1}\left(\frac{RSD}{tRSE}\right)^2$$

where r is the number of replicates and n is the increment size. For desired n the intersect should be found with a corresponding graph and the x-value of the intercept will give the value of r.

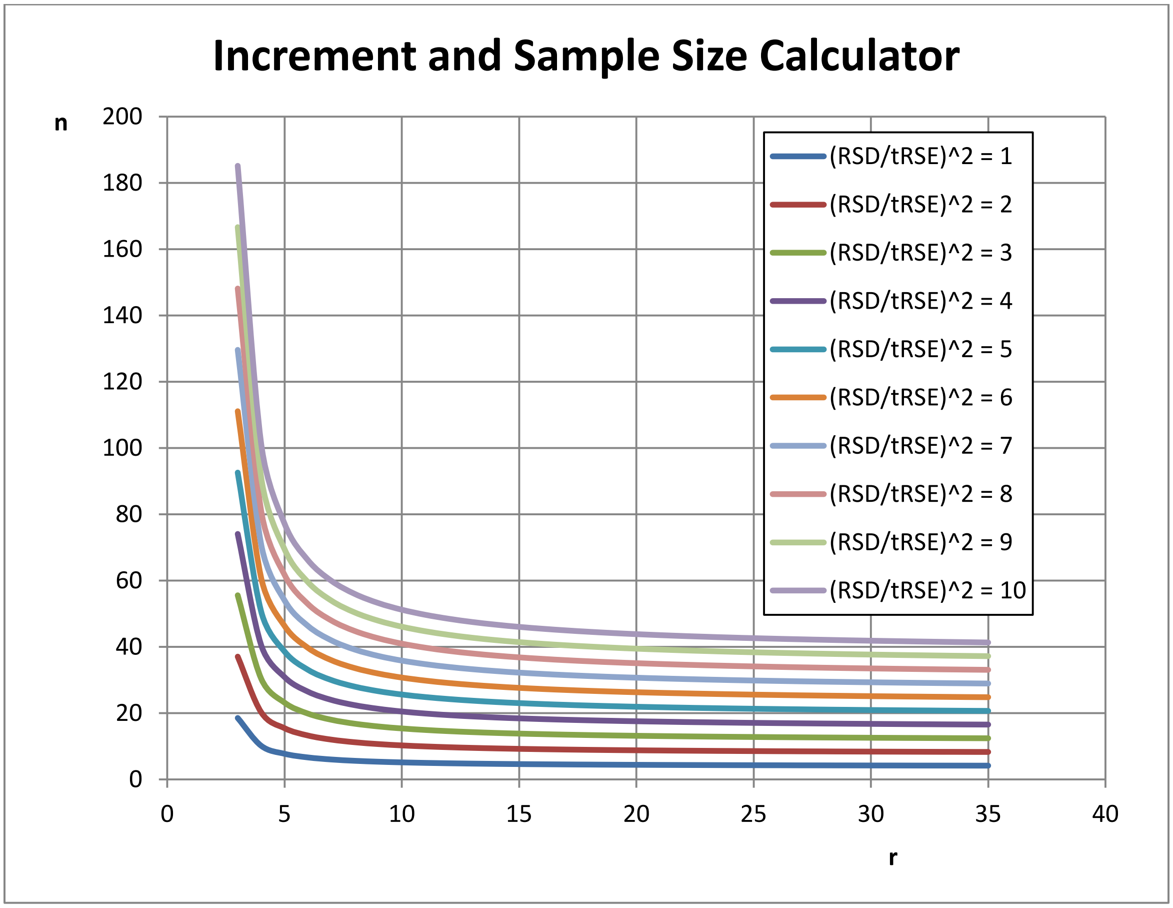


Fig.5. Sample size calculator for incremental sampling.

As clear from the graph above, given confidence interval can be obtained by different combinations of r and n. The table below gives some of these values.

r	(RSE/RSD)^2	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	1/10
		n									
3		19	37	56	74	93	111	130	148	167	185
4		10	20	30	41	51	61	71	81	91	101
5		8	15	23	31	39	46	54	62	69	77
6		7	13	20	26	33	40	46	53	59	66
7		6	12	18	24	30	36	42	48	54	60
8		6	11	17	22	28	34	39	45	50	56
9		5	11	16	21	27	32	37	43	48	53
10		5	10	15	20	26	31	36	41	46	51
11		5	10	15	20	25	30	35	40	45	50
12		5	10	15	19	24	29	34	39	44	48
13		5	9	14	19	24	28	33	38	43	47
14		5	9	14	19	23	28	33	37	42	47
15		5	9	14	18	23	28	32	37	41	46
16		5	9	14	18	23	27	32	36	41	45
17		4	9	13	18	22	27	31	36	40	45
18		4	9	13	18	22	27	31	36	40	45
19		4	9	13	18	22	26	31	35	40	44
20		4	9	13	18	22	26	31	35	39	44
21		4	9	13	17	22	26	30	35	39	44
22		4	9	13	17	22	26	30	35	39	43
23		4	9	13	17	22	26	30	34	39	43
24		4	9	13	17	21	26	30	34	39	43
25		4	9	13	17	21	26	30	34	38	43
26		4	8	13	17	21	25	30	34	38	42
27		4	8	13	17	21	25	30	34	38	42
28		4	8	13	17	21	25	29	34	38	42
29		4	8	13	17	21	25	29	34	38	42
30		4	8	13	17	21	25	29	33	38	42
31		4	8	13	17	21	25	29	33	38	42
32		4	8	12	17	21	25	29	33	37	42
33		4	8	12	17	21	25	29	33	37	41
34		4	8	12	17	21	25	29	33	37	41
35		4	8	12	17	21	25	29	33	37	41
1000		4	8	12	15	19	23	27	31	35	39
2000		4	8	12	15	19	23	27	31	35	38

As can be seen from the above table, the increase in replication does not decrease the confidence interval after r becomes greater than 6. Another observation is that the confidence in terval keeps decreasing with the increase of the number of increments.

In general, instead of taking incremental sample with, for example, r=10 replicates, we would recommend taking a random sample of size 10. Both lead to the same confidence interval while the latter is much easier to achieve practically.

8.3 About Hypothesis Testing vs. Confidence Interval

Using UCL at 95% is equivalent of t-testing at a=2.5% level of significance the hypothesis: Ho: u>u threshold. A left tale test when rejected provides a confirmation to the alternative hypothesis Ha: u<u threshold. These two approaches are equivalent and identical. However, in some textbooks and instructional manuals employing 95% UCL method it is overlooked that the error is not 5 % as some of them claim but actually is smaller and equals to 2.5 % which in real life applications may make a big difference.

8.4 What should be the Sample Size relative to the population Size?

Increasing a sample size produces an additional factor in reducing the confidence interval when the sample size is comparable to the population size. Table below can be used to estimate sample size on this basis.

n/N	Additional factor
1%	0.99
5%	0.97
10%	0.95
15%	0.92
20%	0.89
25%	0.87

Table 2. Reduction of confidence interval by an additional factor due to the relative increase in sample size.

8.5 About Regression Analysis

In environmental applications it is frequently needed to know whether there is a trend present in the sampling data. A trend line does not provide this information. It only indicates that there might be a trend. The actual proof that there is a trend is provided by the t-test of significance of the trend exhibited by the sampling data.

The prediction values based on the trend line are actually the center points of the prediction intervals that actually estimate the prediction value. Without prediction intervals a predicted value becomes just a point estimate without any confidence associated with it.

One of the conditions for the regression analysis to be true is the assumption of the normality of the population distribution. This assumption is very strict and may not hold true in general.

8.6 Seasonality

Seasonality cannot be analyzed unless seasonal data is taken.

8.7 Inferential vs. Descriptive Statistics

Inferential statistics should not be considered as a primary source of information because even in the ideal research the results may still be erroneous due to the associated with the method error. Which means that there is a certain percent probability that the result is wrong. To remind the reader what this means, something that has 1% probability of occurring may be the outcome of the event the first million times, although true, the next 99 million times it may never occur! More stress should be placed on descriptive statistics. Describing the data gives a manager more information than predicting its behavior. This professionals can do without us, researchers.

9 The Best Way to Conduct Statistical Research

This is written for the environmental work only, but could also be applied to many other areas of business research. Statistics as a tool to investigate existing conditions and make conclusions based on the past behavior can and should be applied in environmental work. The question is how.

This may be a revolution in statistical research., and as all revolutions may encounter resistance from those who are at the power today – statisticians. No one will be willing to give up what has been accepted to be the best methodology for years. The companies do not accept their bankruptcy so readily.

Anticipating a steady and scientific denial of this methodology , the author addresses the people who use the results of statistical research in their decision-making process – managers and executives: I will give you two reasons why you should use my methodology over theirs:

1. If you ever care about the cost of statistical research, it will be cut by half or even more.
2. If you ever wonder what those numbers mean, you will have a clear understanding of each number given to you through proposed methodology.

I will also give you one biggest reason why you should reject this methodology:

- Simply because it has not been accepted yet by the top research companies that you are most likely to hire for your project. Until such time that they do, they will always reject it in favor of their own methodologies.

Any practical manager should ask at this point: what is your methodology? I will describe it to you in one small subsection below.

10 New Methodology of Research based on Statistics

There are no new methods in this methodology. But its philosophy is completely different from the one that exists today. I will recap what exists today first.

In environmental work, research is based on collecting a sample and analyzing it. When statistician today sees your sample and the results of lab analysis, he uses the methods of inferential statistics to infer or generalize the results to the entire area of interest. He gives you errors (levels of confidence, etc.) and numbers (estimates). You look at numbers and almost always ignore errors if they are acceptably low. I will tell you: one tenth percent of the error may destroy your entire project. A probability of destroying your project will be one tenth percent. So if there are one thousand projects based on this methodology, one of them will be destroyed and that one may be yours. And think how many projects are carried on in the United States - tens of thousands, if not more. Of course, you can risk it, but then you should not blame your statisticians for the failure of your endeavor. They always tell you what the error is. Of course, they never tell you that that error may lead to failure of your undertaking – for very understandable reasons. So here is your old, acclaimed methodology: You risk you success if you base your decision on its conclusions.

The above scenario is given for a perfectly well conducted research. But there is no such thing as ideal research. The sample is not completely random, the data is not completely reliable, or there can be many other reasons that make your research normal, not so perfect, in reality. For normal research there are even more reasons why you should worry about success. But without having better methodology, we use what we have.

However, what if one of the experts came up with an idea: let us not try to infer anything from the results of sampling. That will kill all errors at one go. You will no longer need to worry about the randomness of your sample or meeting the requirements of this or that inferential method you use. Simply do not use it. What can one do with the sampling results? Analyze them using descriptive statistics. (for those who are not statisticians, I will remind that the line fitting and prediction is one of the methods of descriptive statistics!). But, and here comes a new challenge, a sample should be collected based on the expert knowledge of environmental workers of the area of interest. Environmentalists should tell the statisticians where to sample and what is the expected picture of contamination in the area. It is a team work of an environmentalist and a statistician to choose a sample. It will not be random. But how much more representative this sample will be than a random sample of a statistician! All statisticians working on the environmental projects should become practical environmentalists. Hopefully, you are beginning to see advantages of this methodology. You will no longer be faced with bare numbers that no statistician will know how relate to your environmental problems. Because they will not be able to analyze the data if they do not understand the problem. And if you are worried about their results, they will be identical to what you are getting now, except they will have no errors. They will show you the trend, they will give you the predictions, they will tell you if there are any interesting abnormalities in the data, etc. You will not be bombarded by confidence levels, significance levels, statistical powers , etc. Believe me, even we, professional statisticians, sometimes do not understand these complicated notions fully ourselves, that is why we keep doing research in this area.

So here is a brief synopsis of my methodology:

1. Sit down and try to understand the problem, become an environmentalist and get involved in environmental part of the project.
2. Together with environmental workers designate your sample. Use your statistical knowledge to advise them, but let them have the last word.
3. Do your sampling and lab analysis.
4. Use all methodologies known in statistics to provide clean and understandable statistically sound picture of contamination in the area. Do not use inferential statistics.
5. Let them make conclusions, they will understand every single number you give them this way.
6. Get your paycheck and be happy.

You may disagree with the first five points but the last one all statisticians will accept gladly . . .

11 Conclusion

Never liked writing this section in the reports . . . let us simply skip it . . .