United States Environmental Protection Agency | Office of Solid Waste and Emergency Response | Publication: 9355.4-04FS July 1991

EPA

A Guide: Methods for Evaluating the Attainment of Cleanup Standards For Soils and Solid Media

Office of Emergency and Remedial Response
Hazardous Site Control Division OS-220W

Quick Reference Fact Sheet

GOALS

This fact sheet highlights statistical concepts and methods used in the evaluation of the attainment of cleanup standards. It provides an example of a basic procedure for determining sample size required to obtain a given confidence level focusing on a cleanup standard specified as a mean concentration with a specified confidence. It does not provide policy on specification of cleanup levels but should be considered a technical reference guide for using some of the more common methodologies. More detailed information on these and other methodologies can be obtained from Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Solid Media, EPA 230/02-89/042. Copies of this volume are available from the National Technical Information Service, Springfield, VA 22161. Price: $28.95 (paper), $6.95 (microfiche).
[Terms in ***bold, italicized*** print are defined in the glossary on the last page of this fact sheet.]

WHY ARE STATISTICS IMPORTANT?

Statistical methods perform a powerful and useful function. They allow extrapolation from a set of samples to the entire site in a scientifically valid fashion.

Extrapolation involves uncertainty. Statistical methods enable estimation and management of the uncertainty. Ideally, uncertainty may be reduced to any desired level given complete freedom in sampling and testing. This is seldom a viable option, so statistics are used to determine a balance between sampling and certainty.

Statistical principles can be used to design sampling plans that correlate with methods of analysis tailored to evaluating attainment of cleanup standards. Correlated sampling and analysis methodologies offer higher confidence levels in decision-making.

Efficient statistical sampling plans can be developed to detect the presence of hot spots on a site. The plans allow the prediction of the uncertainty of overlooking a hot spot of a specified size. ***Sequential test procedures*** test only enough samples to accept or reject a clean or not-clean hypothesis and this can quickly indicate highly contaminated areas or areas of very low contamination.

Statistical methods can be used to compute mean concentrations over areas where information indicates that contaminant levels are substantially higher or lower than surrounding levels. This provides more accurate evaluation through limiting dilution of the mean by data from unaffected soil units.

ROLE OF STATISTICS

If a remedial cleanup goal is that each square meter of site soil surface shall have a residual concentration level no greater than (C) ppm, how can the attainment of such a goal be measured? If the site area is one hectare (2.87 acres), there are ten thousand square meters of surface area. To be absoutely sure, one must test each square meter for contamination (if one sample from each meter is known to be representative of the whole meter). Obviously, ten thousand samples is prohibitive. So, what are the alternatives?

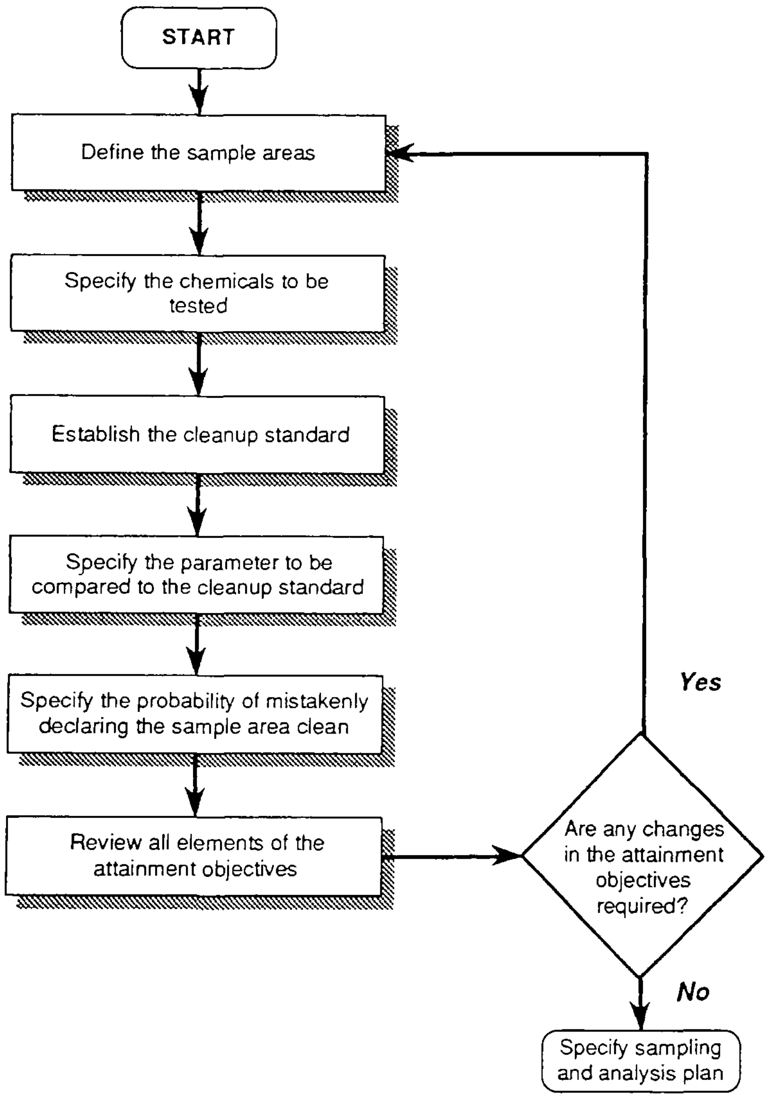
If the number of samples that can be economically and practically acquired is limited, the question immediately arises: how representative of the whole site is a small set of samples? There is a chance, for instance, that either too many samples came from relatively clean areas of the site or from the more heavily contaminated areas of the site. The possibilities present a finite probability that a ***false positive*** (α) or ***false negative*** (β) conclusion may be drawn where the actual condition of the site is misinterpreted because of uncertainties in sampling. Statistical sampling and analysis techniques allow a determination of the level of confidence for a specific set of conditions. These techniques can be used to evaluate data or determine how much data are required to confirm that a designated cleanup level has been attained.

Statistical evaluations also provide a logical consistent approach for optimizing results from limited resources. The known properties of sample data distributions are used to design sampling plans and data analysis routines to provide predictable confidence levels for decisions. The confidence levels attainable will depend on the quantity and quality of available data.

It helps to think of cleanup standards as having four components: 1) the magnitude -- concentration deemed protective of human health and the environment; 2) a sampling plan to evaluate attainment of the specified concentration; 3) a method for comparing the data collected to the cleanup level; and 4) the probability of mistakenly declaring the sample area clean (false positive rate). All but the first step depends heavily on statistical analysis. Figure 1 indicates the steps that must be completed to define attainment objectives.

Various methods can be used to compare data to cleanup levels, e.g., 1) average condition (mean concentration ($\bar{x}$) is below cleanup level at a specified confidence level); 2) value rarely to be exceeded (specified proportion of soil is below a cleanup standard); or 3) hot spots that should be found if present. Examples of other options in which methods are combined are provided in Box 1. It is important to consider the attainment evaluation during the site investigation so that the method for evaluating attainment can be included in the remedy specification.

FIGURE 1 - Steps in Defining the Attainment Objectives



BOX 1 - Examples of Using Multiple Attainment Criteria

- Most of the soil has concentrations below the cleanup standard while concentrations are above the cleanup standard. This standard may be accomplished by testing whether the 75th percentile is below the cleanup standard and whether the mean of those concentrations above the cleanup standard is less than twice the cleanup standard.
- The mean concentration is less than the cleanup standard and the standard deviation (σ) of the data is small, thus limiting the number of extreme concentrations. This standard may be accomplished by testing if the mean is below the cleanup standard and the coefficient of variation (r) is less than a low level (.5 for example).
- The mean concentration is less than the cleanup standard and the remaining contamination is uniformly distributed across the sample area relative to the overall spread of the data. Testing these criteria may be accomplished by testing for a mean below the cleanup standard and variability between strata means that is not large compared to the variability within strata (analysis of variance).
- The mean concentration is less than the cleanup standard and no area of contaminated soil (assumed to be circular) is larger than a specified size.

STATISTICAL METHODOLOGY LIMITATIONS

When key assumptions about the site and collected data are violated, the statements of data confidence may change. Statistical assumptions include: the ***sample area*** is homogenous; the ***distribution*** of data is ***normal***, or can be transformed into near normal data (e.g., taking the log of the data tends to normalize the data thus allowing standard procedures to be used); and sampling locations were selected using a ***simple random sampling*** procedure.

PROCESS - DETERMINING WHETHER THE MEAN CONCENTRATION AT A SITE IS LESS THAN THE CLEANUP STANDARD

Power Curve

The probability of declaring a ***sample area*** clean will depend on the ***sample population*** mean concentration. The relationship between a population mean and decision outcome is shown in Figure 2. This relationship is known in statistics as a "power curve."

Power curves can facilitate understanding the relationship between mean concentration and confidence level. Power curves also can help determine an appropriate sample size.

Sampling Plan

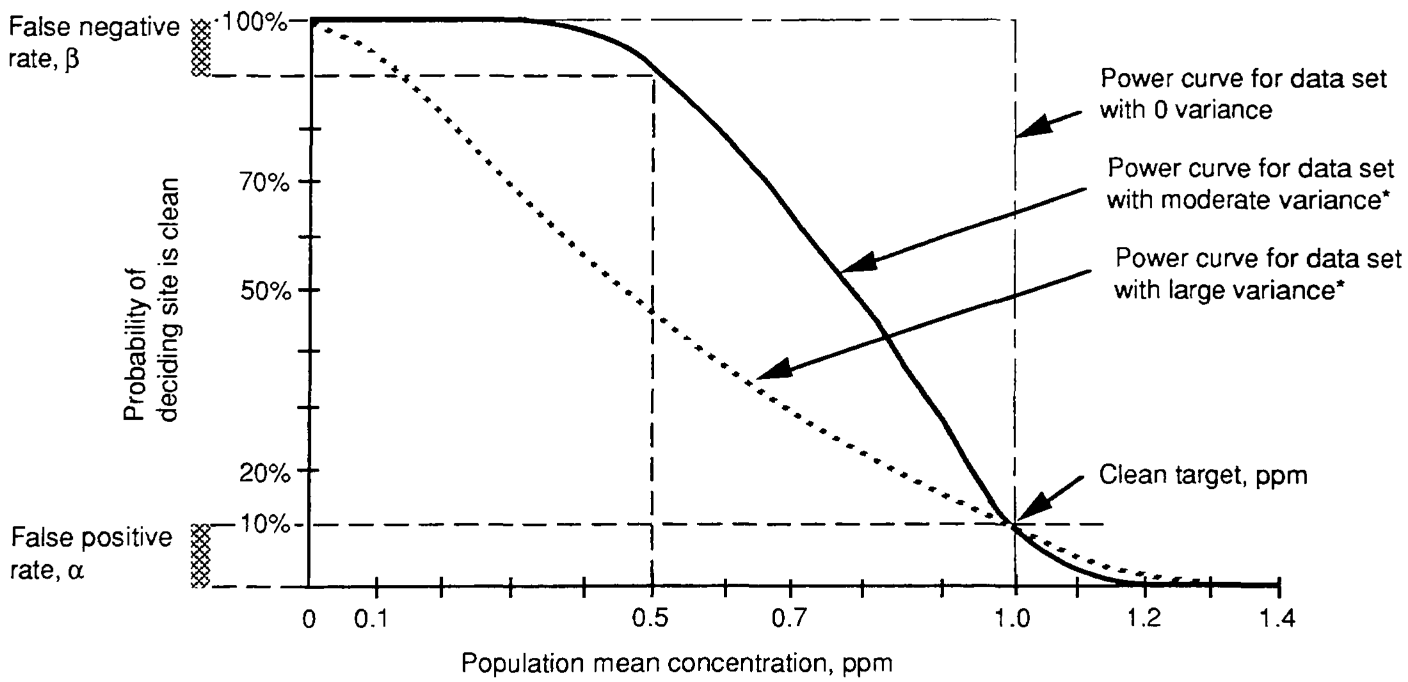
Once the cleanup concentration and statistical method (i.e., for this discussion, the mean concentration) has been specified, the sampling and analysis plans should be developed. There are two basic types of sampling plans: systematic and random. These are illustrated in Figure 3.

Pros and Cons - Systematic or Random Sampling

Systematic sampling is generally easier to carry out. Such sampling almost always results in both lower costs and in higher data reliability than ***simple random sampling***. Systematic sampling also protects against having large contiguous areas of high

FIGURE 2 - Power Curve

This curve represents a condition where, when both the ***false negative*** (β) and ***false positive*** (α) risks are set at 10%, the population mean concentration must be 0.5 ppm (or less) in order to be 90% certain the site is clean at the 1 ppm level. Power curves have been developed for several values of α and can be found in Appendix A of <u>Methods for Evaluating the Attainment of Cleanup Standards</u>. They are defined by the cleanup level, the false negative rate, and the ***variance*** and can be used to determine the mean concentration required to achieve a particular false positive rate. (See example calculation at end of fact sheet.)

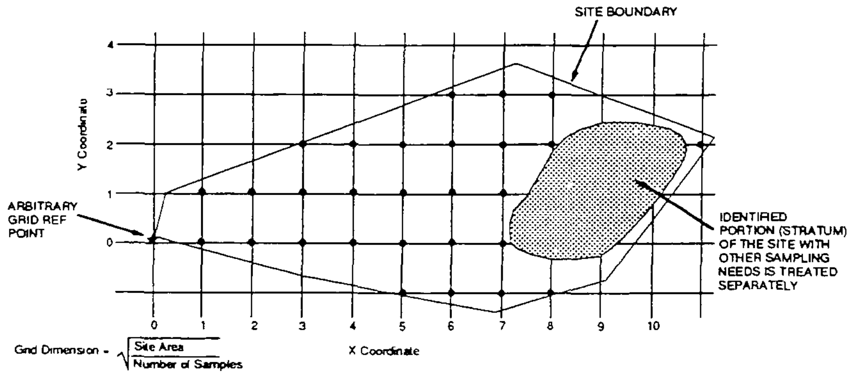


*Whether the variance is considered low, moderate, or high will depend on the magnitude of the standard and the risk level it represents; e.g., a variance that is 10 times the magnitude of the standard may be considered moderate if the standard is conservative (i.e., if the standard is set low).

FIGURE 3 - Strategies for Selecting Sampling Locations

SYSTEMATIC SAMPLING DESIGN EXAMPLE

Systematic sampling distributes the sampling points uniformly over the site area of interest. The systematic sampling plan provides a uniform site coverage with a larger grid spacing.

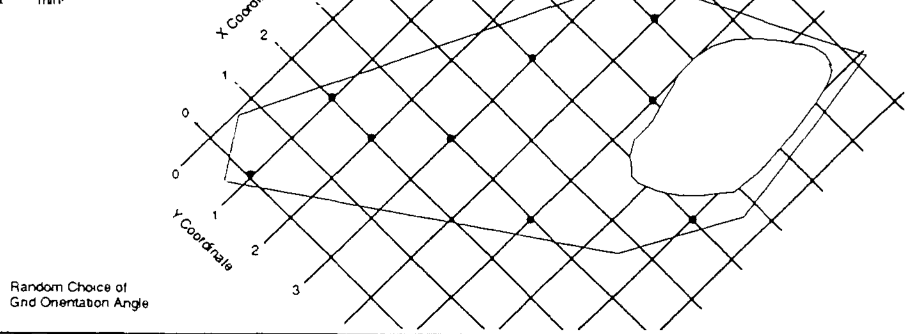


RANDOM SAMPLING DESIGN EXAMPLE

True random selection of sampling points requires that each sample point chosen must be independent of the location of all other sample points. The random sampling plan has a better chance of detecting site anomalies than the systematic sampling plan.

X Coordinate = $X_{min} + (X_{max} - X_{min}) * RND$

Y Coordinate = $Y_{min} = (Y_{max} \; Y_{min}) * RND$



or low contamination of the site being unsampled. In economic terms, this type of sampling gives greater information per unit cost than simple random sampling.

Systematic sampling, however, does have disadvantages. For example, poor precision may result if contamination occurred in a particular pattern that is missed by the system's sampling, i.e., if contamination occurred along a straight line such as a pipeline or trench. Also, there are no trustworthy methods for estimating error under systematic sampling. If contamination is not random, systematic sampling can yield biased results.

Sample Size

Once a general sampling plan is selected, the number of samples to be taken can be determined.

The number of samples necessary to reliably determine site conditions depends on a variety of factors. There is the desired level of confidence, variability of the sampling results (σ^2), and accuracy with respect to cleanup target (ppm). The following equation can be used to determine the minimum sample size required to achieve designated levels of confidence:

$$\text{Number of samples} = \sigma^2 \left(\frac{Z_{(1-\alpha)} + Z_{(1-\beta)}}{C_s - \mu_1} \right)^2$$

where:

σ^2 = variance of the data

C_s = cleanup standards, ppm

μ_1 = alternative clean decision level

$Z_{(1-\alpha)}$ and $Z_{(1-\beta)}$ = the false positive and false negative normal deviates, respectively. See Table 1 for values of Z based on α and β.

$Z_{(1-\alpha)}$ is the normal deviate point associated with the error of saying the site attains C_s when in fact it does not. Under the ***null hypothesis***, H_o, (i.e., under the assumption that the site does not attain C_s) the probability of exceeding the $Z_{(1-\alpha)}$ value is desired to be α. μ_o is the mean concentration under the null hypothesis.

Similarly $Z_{(1-\beta)}$ is the normal deviate point associated with the error of saying the site does not attain C_s when in fact it does. Under this ***alternate hypothesis***, H_1, that the site attains the cleanup standard, the probability of exceeding the $Z_{(1-\beta)}$ value is desired to be β. μ_1 is the mean concentration under the alternate hypothesis H_1.

Hence, $\mu_1 \leq C_s$. The relationship of μ_1, C_s, α, and β is illustrated in Figure 4.

The *variance* is generally not known at the time that the sample size is being calculated but can be estimated from any data that does exist or crudely approximated using the formula:

$$\hat{\sigma}^2 \text{ (estimated variance)} = \text{Range}/6$$

where Range is the expected spread between the smallest and largest values.

Box 2 shows a sample calculation of sample size.

Evaluation of Attainment

The mean of the sampling data is an estimate of the mean contamination of the entire sample area; it does not convey information regarding the reliability of the estimate. Through the use of a "confidence interval," it is possible to provide a range of values within which the true mean is located.

The formula for an upper one-sided $100(1\text{-}\sigma)$ percent confidence limit around the population mean is presented below:

$$\mu_{U\alpha} = \overline{X} + t_{(1-\alpha, df)}(s/\sqrt{n})$$

where:

$\overline{X}$ = computed mean level of contamination

S = the ***standard deviation*** of the sampling data

df = the degrees of freedom (= n-1)

$\mu_{\mu\alpha}$ = confidence limit

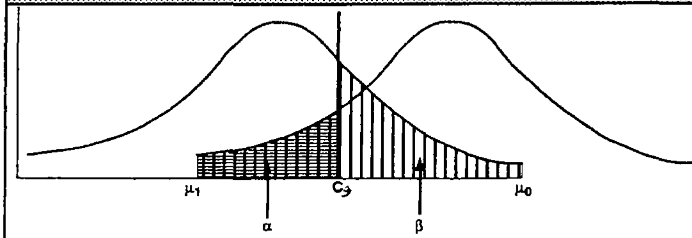
The appropriate value of $t_{(1-\alpha, df)}$ can be obtained from Table 2. The one-sided confidence interval can be used to test whether the site has attained the cleanup standard.

To determine whether the site meets a specified cleanup standard, use the upper one-sided confidence limit U, defined in the above equation. If $\mu_{U\alpha} < C_s$, conclude that the area attains the cleanup standard. If $\mu_{U\alpha} \geq C_s$, conclude that the area does not attain the cleanup standard.

EXAMPLE CALCULATION USING THE POWER CURVES

At a former wood processing plant it is desirable to determine if the average concentrations of PAH compounds in the surface soil are below 50 ppm (the cleanup standard C_s). The project managers have decided that the dangers from long-term exposure can be reasonably controlled if the mean concentration in the sample area is less than the cleanup standard. The false positive rate for the test is to be at most 5% (i.e., $\alpha = .05$). The false negative rate is desired to be no more than 20% (i.e., $\beta = .20$). The coefficient of variation of the data is thought to be about 1.2. The power curves for $\alpha = .05$ (see Figure 5) and the approximate sample sizes for random sampling were reviewed.

FIGURE 4 - Relationship of μ_1 and C_s



H_0: The site exceeds the cleanup standard (i.e., $\mu_0 \geq C_s$)
H_1: The site attains the cleanup standard (i.e., $\mu_1 < C_s$)
C_s = Cleanup Standard

BOX 2 - Example Sample Size Calculation

If the site cleanup target (C_S) is 12 ppm, the alternative clean decision level (μ_1) is 11 ppm, and the expected variance (σ^2) of the data is 8, we can obtain a 95% confidence level (false positive rate = .05) at a risk of 10% (false negative rate = 0.10) of declaring the site clean by determining the mean site sample concentration from:

$$8\left(\frac{1.645 + 1.282}{12\text{-}11}\right)^2 = 68.53 = 69 \text{ samples}$$

$$Z_{(1-\alpha)} = Z_{.95} = 1.645$$
$$Z_{(1-\beta)} = Z_{.90} = 1.282$$

Note: If the false negative risk is decreased from 10% to 5%, the number of samples required would increase to 87. The reduction of risk always requires increasing sample size.

These curves illustrate the relationship between cleanup level and probability of attainment for various sample sizes. Approximate sample sizes for a range of coefficients of variation are presented below the figure as a guide to determining which curve is appropriate for the situation under consideration. Using this information, the following conclusions can be made:

- While it would be desirable to have a test with power curves similar to E and F, the sample sizes of more than 100 will cost too much.

- Power curves A, B, and C have unacceptably low power (i.e., the power, $1\text{-}\beta$, is too low) when the mean concentration is roughly 75% of the cleanup level (i.e., 37 ppm). For example, at .75 on the x axis, curves A, B, and C give power (on the y axis) of approximately .15 to .40 (i.e., β error rates of .85 to .60). This clearly is undesirable in most situations. Viewing the table in Figure 5, we see that in order to have a false negative rate of 20% or less the site mean concentration would have to be approximately 25% of the cleanup level for curve A to 57% of the cleanup level for curve C.

- Consequently, a reasonable compromise between high power and low sample size is to have a test with a power curve similar to D.

Based on specifications above and the table at the bottom of Figure 5, the information needed to calculate the sample size is:

$\alpha = .05$;
$\beta = .20$; and
$\mu_1 = C_s * .69 = 34.5$ ppm.

These values can be used to calculate sample size. From Table 1:

$$Z_{1-\alpha} = 1.645$$
$$Z_{1-\beta} = 0.842$$
$$\text{c.v.} = \frac{\sigma}{\mu_1} \approx 1.2$$
$$\sigma \approx (1.2)(34.5) = 44.4$$
$$\sigma^2 \approx 1971.36$$

$$\text{Number of samples} = \sigma^2 \left(\frac{Z_{(1-\alpha)} + Z_{(1-\beta)}}{C_S - \mu_1}\right)$$

$$\text{Number of samples} = 1971.36 \left(\frac{1.645 + .842}{50 - 34.5}\right)^2$$

$$= 50.75 \approx 51 \text{ samples}$$

This number is smaller than the numbers presented below Figure 5 because the numbers in Figure 5 are calculated to be conservative estimates (C_s was used to calculate σ rather than μ_1).

Once the samples are taken, attainment can be evaluated as follows:

The following data are known or calculated:

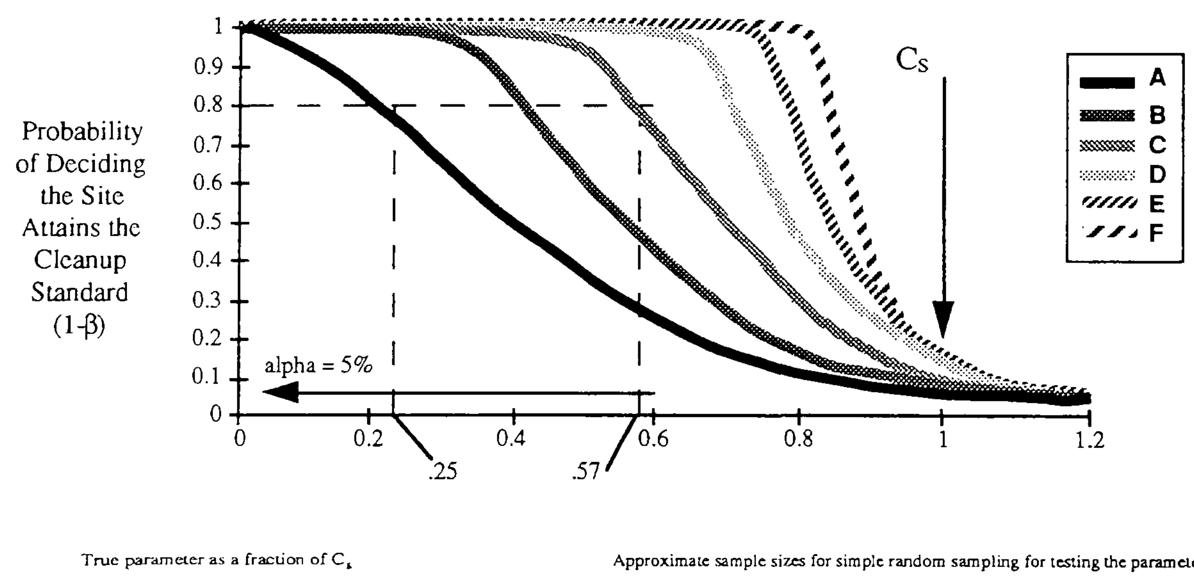
C_s = Cleanup Standard = 50 ppm
$\bar{x}$ = Mean concentration = 38 ppm
s = Standard deviation = 15

$$t_{1-\alpha,df} = \frac{1.684 + 1.671}{2} = 1.677 = 41.52$$

The upper one-sided 95% confidence interval goes to $\mu_{u\alpha} = \bar{x} + t_{1-\alpha,df}\frac{S}{\sqrt{n}} = 38 + 1.677\frac{15}{\sqrt{51}} = 41.52$

Since 41.5 < 50, there is a 95% confidence that the mean concentration of the sample area attains the cleanup standard of 50 ppm.

FIGURE 5 - Power Curves for $\alpha = 5\%$



Parameters for the Power Curves	Power Curve: A	B	C	D	E	F
$\alpha =$	.05	.05	.05	.05	.05	.05
$\beta =$	.20	.20	.20	.20	.20	.20
$\mu_1 =$	$.25 * C_s$	$.43 * C_s$	$.57 * C_s$	$.69 * C_s$	$.77 * C_s$	$.84 * C_s$

Approximate sample sizes for simple random sampling for testing the parameters indicated

Parameters being tested	Power Curve: A	B	C	D	E	F
Mean						
with cv (data) = .5	4	5	9	17	30	61
with cv (data) = 1	11	20	34	65	117	242
with cv (data) = 1.5	25	43	76	145	264	544

TABLE 1
Z Values for Selected Alpha and Beta

β / α	$Z_{1-\beta}$ / $Z_{1-\alpha}$
0.450	0.124
0.400	0.253
0.350	0.385
0.300	0.524
0.250	0.674
0.200	0.842
0.100	1.282
0.050	1.645
0.025	1.960
0.010	2.326
0.0050	2.576
0.0025	2.807
0.0010	3.090

TABLE 2
Table of t for Selected Alpha and Degrees of Freedom

Use alpha to determine which column to use based on the desired parameter, $t_{1-\alpha,df}$. Use the degrees of freedom to determine which row to use. The t value will be found at the intersection of the row and column. For values of degrees of freedom not in the table, interpolate between those values provided.

Degrees of Freedom (df)	α for determining $t_{1-\alpha,df}$							
	.25	.10	.05	.025	.01	.005	.0025	.001
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160
400	0.675	1.284	1.649	1.966	2.336	2.588	2.823	3.111
infinite	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090

GLOSSARY

Distribution – The frequencies with which measurements in a data set fall within specified intervals.

False Negative (β) – The probability of mistakenly concluding that the sample area has not attained the cleanup level when it has. It is known as the probability of making a Type II error.

False Positive (α) – The probability of mistakenly concluding that the sample area has attained the cleanup level when it has not. It is known as the probability of making a Type I error.

Hypothesis – An assumption about a property or characteristic of a population under study. The goal of statistical inference is to decide which of two complementary hypotheses is likely to be true. In the context of this document, the *null* hypothesis is that the sample area has not achieved the cleanup standard and the *alternative* hypothesis is that it has.

Normal Distribution – A family of "bell-shaped" distributions, or curves, where each individual distribution is uniquely defined by its ***mean*** and ***variance***.

Sample Area – The specific area within a waste site for which a separate decision on attainment is to be reached.

Sample Mean – The arithmetic average of a set of sample measurements, $x_1, x_2, \ldots x_n$, defined to be:

$$\overline{x} = \sum_{i=1}^{n} X_i / n$$

Sample Population – The total number of soil/solid media units at a waste site for which inferences regarding attainment of cleanup standards are to be made.

Sample Standard Deviation – The more commonly used measure of dispersion of the sample measurements, defined to be:

$$S = \sqrt{S^2}$$

(See definition for ***variance***)

Sequential Test Procedures – Sampling process that terminates when enough evidence is obtained to either accept or reject the null hypothesis.

Simple Random Sample – A sample of n units collected from a population of interest (for example, all possible samples of soil units at a site) such that each unit has an equal chance of being selected.

Variance – A measurement of dispersion of the sample measurements, $x_1, x_2, \ldots x_n$, defined to be:

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 / n - 1$$

EPA

First Class Mail
Postage and Fees Paid
EPA
Permit No. G-35

United States
Environmental Protection
Agency (OS-120)
Washington, D.C. 20460