PNNL-15202

Pacific Northwest
National Laboratory
Operated by Battelle for the
U.S. Department of Energy

Technical Documentation and Verification for the Buildings Module in the Visual Sample Plan (VSP) Software

R. O. Gilbert
J. E. Wilson
B. A. Pulsipher
N. L. Hassig
L. L. Nuffer

June 2005

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

**Available to DOE and DOE contractors from the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062; ph: (865) 576-8401 fax: (865) 576-5728 email: reports@adonis.osti.gov**

**Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161 ph: (800) 553-6847 fax: (703) 605-6900
email: orders@ntis.fedworld.gov online ordering: http://www.ntis.gov/ordering.htm**

This document was printed on recycled paper.
(9/2003)

PNNL-15202

Technical Documentation and Verification for the Buildings Module in the Visual Sample Plan (VSP) Software

R. O. Gilbert
J. E. Wilson
B. A. Pulsipher
N. L. Hassig
L. L. Nuffer

June 2005

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Acknowledgments

The authors are pleased to acknowledge Rebecca L. Blackmon, who is the Technical Support Working Group program manager for this project, Teresa Lustig of the Department of Homeland Security, and Les Sparks of the U.S. Environmental Protection Agency for their support, advice, and encouragement in the development of the Buildings Module of the Visual Sample Plan (VSP) software. Also, the many contributions made by project team members Donald L. Hadley and Nancy B. Valentine of the Pacific Northwest National Laboratory (PNNL) on important bio-threat issues and sampling methods are very much appreciated. A special "thank you" is also extended to the following individuals in the Statistical Sciences Group at PNNL: Craig A. McKinstry for helping to verify the correctness of statistical procedures, Lucille A. Walker for project administrative and financial accounting support, and Mary H. Cliff for assisting with final preparation of this report and other essential day-to-day needs.

Contents

Figures

Tables

Abbreviations and Acronyms

AL	Action Level
CTTSO	Combating Terrorism Technology Support Office
DHS	Department of Homeland Security
DOE	U.S. Department of Energy
DoD	U.S. Department of Defense
DQA	Data Quality Assessment
DQO	Data Quality Objectives
DPGD	Decision Performance Goal Diagram
EPA	U.S. Environmental Protection Agency
ESTCP	Environmental Security Technology Certification Program
LF	Lilliefors test for normal distribution
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MQO	Measurement Quality Objectives
PNNL	Pacific Northwest National Laboratory
SERDP	Strategic Environmental Research Development Program
SW	Shapiro-Wilk test for normality
TSWG	Technical Support Working Group
QA	Quality Assurance
VSP	Visual Sample Plan

Summary

Visual Sample Plan (VSP) is an easy-to-use visual and graphic statistically-based software tool being developed by the Pacific Northwest National Laboratory (PNNL) to help determine the appropriate number and location of environmental samples so that environmental decisions can be made with the required confidence. The VSP software, which is available free at http://dqo.pnl.gov/vsp, is a significant aid in developing probability-based sampling designs (number and location of samples and measurements) using the Data Quality Objectives (DQO) planning process developed by the U.S. Environmental Protection Agency (EPA). VSP also has the capability of conducting statistical analyses to provide descriptive statistical summaries of data sets, to test whether data are normally distributed, and to compute upper confidence limits on means.

This report is the latest in a series of reports that document the statistical methods used in VSP [Davidson (2001), Gilbert et al. (2001), Gilbert et al. (2002), and Gilbert et al. (2003)] and the quality assurance (QA) activities conducted by PNNL to verify that VSP computations are correct and accurate. This report focuses on the VSP buildings module that was developed with support from the Department of Homeland Security (DHS), Combating Terrorism Technology Support Office (CTTSO), and the Technical Support Working Group (TSWG). Section 1.0 provides an introduction and overview of the buildings module, while Section 2.0 describes the statistical computations and methods used. Section 3.0 provides the results of the QA activities. Section 4.0 is the reference list.

The QA verification results in Section 3.0 demonstrate that VSP is providing correct and accurate computations for

- the number of samples required for the various sampling objectives and design options
- all statistical calculations, including descriptive statistics, statistical tests for evaluating if data are normally distributed, and upper confidence limits on the mean, and
- the graphs of data used to visually evaluate if data are normally distributed.

1.0 Introduction and Overview of the Buildings Module

Visual Sample Plan (VSP) is a software tool under development at the Pacific Northwest National Laboratory (PNNL) with support from the U.S. Department of Energy (DOE), the U.S. Environmental Protection Agency (EPA), the U.S. Department of Defense (DoD) via the Strategic Environmental Research Development Program (SERDP) and the Environmental Security Technology Certification Program (ESTCP), the U.S. Navy, and the Department of Homeland Security (DHS).

This report focuses on the VSP buildings module that was recently added to VSP with support from DHS through the Combating Terrorism Technical Support Office, Technical Support Working Group (TSWG). The motivation for developing the buildings module is the possibility that biological, chemical, or nuclear agents may be used in terrorist attacks on buildings. If these attacks should occur, measurements of building surfaces will be needed to decide whether specific rooms or suites of rooms have been contaminated, require decontamination, have been successfully decontaminated, or have been recontaminated. The VSP buildings module complements other modules in VSP that were developed to determine the number and location of samples needed to assess the levels of contamination in soils and sediments, as documented in the VSP User's Guide [Hassig et al. (2004) and Hassig et al. (2005)].

The buildings module helps the VSP user to quickly determine the number and location of samples needed for a user-defined "target population." The target population is a specified portion of surfaces (walls, floors, ceilings, doors, and windows) in a room or a suite of rooms for which a decision is needed. For example, a target population might consist of all floor, wall, and ceiling surfaces in two adjacent rooms that share an air circulation system.

VSP determines the number and location of samples needed for a defined target population when a decision about the population will be made on the basis of

- comparing the mean or median to an action level (AL) (threshold value)
- comparing individual measurements to an AL, or
- comparing both an average and individual measurements to ALs.

If the data are normally distributed, VSP determines the number and location of samples or in-situ measurements for four statistical tests that compare the *mean* to an AL:

- the one-sample t test
- the Sequential Probability Ratio Test
- Barnards Sequential t test, and
- Collaborative Sampling.

If the data are not normally distributed, VSP determines the number and location of samples or in-situ measurements for two statistical tests that compare the *median* (50^{th} percentile) to an AL:

- MARSSIM Sign test, which is applicable to any non-normally distributed set of data
- Wilcoxon Signed Ranks test, which is applicable to any symmetric non-normally distributed set of data.

MARSSIM is an acronym for Multi-Agency Radiation Survey and Site Investigation Manual. When comparing *individual* samples or in-situ measurements to the AL, VSP determines the number and location of samples or measurements needed for making decisions using the following statistical methods:

- Assessing if circular or elliptical *hot spots* that exceed a specified size and AL are present
- Comparing a *parametric one-sided upper tolerance limit* (UTL) to the AL to assess if the fraction of the target population that exceeds the AL is larger than what can be tolerated. The method assumes that the data are normally distributed
- Comparing a *non-parametric (distribution-free) one-sided UTL* to the AL to assess if the fraction of the target population that exceeds the AL is larger than what can be tolerated. This UTL is valid regardless of the underlying distribution of the data.
- Using *acceptance sampling* to assess if the percentage of the target population that exceeds the AL is larger than what can be tolerated. The target population consists of a finite number of small square geographical "grid units," a subset of which are measured. A grid unit might be the area over which a swipe or swab sample is taken, or the area (location) that is effectively scanned by an in-situ detector.
- Using the *Wright-Grieve Bayesian* method to determine the number of small grid units in the target population that must be measured and found to be less than the AL (or more generally, to be non-defective) to state with specified confidence that *all* of the grid units in the target population are less than the AL.

All of these methods are defined and discussed in Sections 2.1, 2.2 and 2.3. The methods described in these sections were originally developed for other VSP modules (Gilbert et al. 2002). The methods described here in Sections 2.3.2 through 2.3.6 for comparing individual measurements to ALs were developed specifically for the buildings module. The QA activities performed for the previously developed methods are discussed in Section 3.1. QA activities for the new buildings module methods are documented in Section 3.2.

The measurements obtained at locations determined using the buildings module can be entered into VSP in order to compute summary (descriptive) statistics, test for normally distributed data, and compute one-sided upper confidence limits for the mean, and more generally, make decisions on the basis of the data. VSP reports the results of these analyses and automatically prepares a project report that summarizes and describes the sampling objectives, number and location of samples, underlying assumptions, costs, and other information needed to document the basis of data-based decisions. The project report also includes a sensitivity analysis that shows the sensitivity of the required number of samples to changes in VSP user inputs (data quality objectives: DQOs). The latest User's Guide for VSP (Hassig et. al, 2005) includes instructions for using the buildings module. The User's Guide can be downloaded from http://dqo.pnl.gov/vsp.

2.0 Documentation of Statistical Methods and Computations

2.1 Compare Mean to a Threshold when Data are Normally Distributed

2.1.1 One-Sample t Test

The one-sample t-test can be used to test if the true mean of the population exceeds a fixed upper limit (action level). The equation used in VSP to compute the minimum recommended number of samples, n, needed for the test when the VSP user specifies that only r = 1 analytical measurement for each sample will be made is

$$n = \frac{s_{total}^2 \left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5 Z_{1-\alpha}^2 \tag{1}$$

If the VSP Measurement Quality Objectives (MQO) module is used, which allows for r = 1, 2, or 3 analytical replicates of each field sample, then the equation that is used to compute n is

$$n = \frac{\left(s_{sample}^2 + \frac{s_{analytical}^2}{r}\right)\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5 Z_{1-\alpha}^2 \tag{2}$$

The notation used in Equations (1) and (2) is defined in Table 1. These equations are computed using the values of α, β, Δ, r, s_{total}, s_{sample}, and $s_{analytical}$ that are specified by the VSP user.

Table 1. Some Notation used in this Report

Notation	Description
n	the minimum recommended number of samples or observations that should be collected or obtained from the target population, as computed using one of the sample-size equations or algorithms in VSP.
r	the number of measurements (analytical replicates) that will be obtained for each sample.
α	the probability that the VSP user is willing to tolerate that a Type I decision error will be made, i.e., that the data collected and used in the appropriate statistical test will falsely reject the null hypothesis. For example, if the null hypothesis is "the mean concentration of the target population *exceeds* the action level (AL)," then α is the probability the VSP user can tolerate that the statistical test computed using the n data will incorrectly indicate that the mean concentration does not exceed the AL; in short, calling a "dirty" site "clean."
β	the probability the VSP user is willing to tolerate that a Type II decision error will be made, i.e., that the data collected and used in the appropriate statistical test will falsely accept the null hypothesis. For example, if the null hypothesis is that the mean concentration of the target population exceeds the AL, then β is the probability the VSP user can tolerate that the statistical test computed using the n data will falsely indicate the mean concentration does exceed the AL; in short, calling a "clean" site "dirty."

Δ	the width of the "gray region" in the Decision Performance Goal Diagram (DPGD) used in the DQO process (EPA 2000a) and in VSP. For example, if the sampling objective is to compare the true mean of the target population to the true mean of a reference or "background" population, then Δ is the difference between the true target population mean and the true reference mean that the VSP user specifies is important to detect with (high) probability 1 - β. Similarly, if the objective is to compare the mean of the target population to a fixed AL, then Δ is the difference between the true mean and the AL that is important to detect with (high) probability 1 - β.
σ^2_{total}	the true total variance of the population of all possible measurements made on all possible samples collected from the target population. The model of the true total variance used in VSP is $$\sigma^2_{total} = \sigma^2_{sample} + \sigma^2_{analytical}$$ where $\sigma^2_{analytical}$ is the true variance component due to the analytical measurement process in the laboratory and σ^2_{sample} is the true variance component due to all other sources of variation, including variations in true concentrations at different target population locations and the variance added due to selecting, collecting, and transporting samples to the laboratory.
s^2_{total}	the computed *estimate* of the true total variance, σ^2_{total}. If r = 1 for all n samples, then the quantity s^2_{total} is computed as $$s^2_{total} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1},$$ where x_i is the measurement obtained for the single aliquot or measurement from the ith sample and $\bar{x}$ is the arithmetic mean of the n measurements, x_i.
s^2_{sample}	an estimate of the total variance of the data x_i that would be obtained if $\sigma^2_{analytical} = 0$.
$s^2_{analytical}$	an estimate of the total variance of the data x_i that would be obtained if the only variability in the data was due to the analytical process, i.e., if $\sigma^2_{sample} = 0$.
$t_{1-\alpha,df}$	the value of the Student's t-distribution with df degrees of freedom. By definition, the proportion of the distribution to the left of the value $t_{1-\alpha,df}$ is 1- α. A table of the values of $t_{1-\alpha,df}$ is found in most statistics books, e.g., Gilbert (1987, Table A2, page 255).
$Z_{1-\alpha}$	the value from the standard normal distribution for which the proportion of the distribution to the left of $Z_{1-\alpha}$ is 1- α. A table of the values of $Z_{1-\alpha}$ is found in most statistics books, e.g., Gilbert (1987, Table A1, page 254). If the selected probability of a false rejection, α, is made smaller, then $Z_{1-\alpha}$ will be larger, leading to a larger number of samples. If the null hypothesis is that the concentrations in the target population exceed the action level, i.e., that the target population is "dirty," then $Z_{1-\alpha}$ can be thought of as an index number whose magnitude quantifies the strength of our desire to avoid deciding a dirty target population is clean.

$Z_{1-\beta}$	the value of the standard normal distribution for which the proportion of the distribution to the left of $Z_{1-\beta}$ is 1-β. If the selected probability of a false acceptance of the null hypothesis, β, is made smaller, then $Z_{1-\beta}$ will be larger, leading to a larger number of samples. If the null hypothesis is that concentrations for the target population *exceed* the action limit, i.e., that the target population is "dirty," then $Z_{1-\beta}$ can be thought of as an index number whose magnitude quantifies the strength of our desire to avoid deciding a clean target population site is dirty.
$\phi(z)$	the cumulative standard normal distribution function, i.e., $$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx$$ A table of $\phi(z)$ values is provided in most statistics books, e.g., Gilbert (1987, Table A1, page 254).

The assumptions that underlie the derivation of Equation (1) are that the data are normally distributed and representative of the target population, they are not spatially or temporally correlated, and that $\bar{x} - ks_{total}$ is normally distributed with mean $\mu - k\sigma$ and variance $\left(\frac{\sigma^2}{n}\right)\left(1+\frac{k^2}{2}\right)$ where k is a given constant (Guenther 1981). The derivation of Equation (1) is found in Wallis (1947), Guenther (1977), EPA (2000a, Appendix A), and EPA (1992, pp. F-8, F-9, and F-10). Equation (1) is used in the statistics book by Bowen and Bennett (1988, pp. 155, 156), EPA (2000a), EPA(1994, p. 21), and EPA (2000b, pp. 3-7).

Guenther (1981) indicates that although Equation (1) is an approximation to the true minimum sample size required for the one-sample t-test, Equation (1) usually yields the exact solution for n. The exact solution is obtained using an iterative approach using tables of the non-central t distribution found in Owen (1965).

Equation (2) also should provide a very accurate approximation of n for a specified value of r because it is a straightforward extension of Equation (1) to the case of r > 1. For this case, it is easily shown that the total variance is estimated by computing

$$s_{total}^2 = s_{sample}^2 + \frac{s_{analytical}^2}{r}.$$

Inserting this equation for s_{total}^2 into Equation (1) yields Equation (2).

2.1.2 Sequential Sampling (Sequential Probability Ratio Test)

Sequential sampling of small batches of samples or measurements over time is provided in VSP to test if the true mean exceeds a specified threshold value (AL). The sampling is sequential in the sense that several sequential (in time) samplings of the target population are made until the statistical test has the power (ability) specified by the VSP user to decide between the null or alternative hypothesis. Sequential tests are useful if samples can be easily obtained and analyzed (measured) in small batches over time.

The Sequential Probability Ratio Test (SPRT) (Wald 1947; Wetherill 1966) is discussed in this section. This test requires that the standard deviation of the measurements be known with great accuracy before

the sequential sampling study is conducted. In Section 2.1.3, Barnard's sequential test is discussed (Barnard, 1952; Wetherill, 1966). Barnard's test can be used in place of the SPRT when the standard deviation is not known with great accuracy, which is the usual situation.

The assumptions that underlie the SPRT test are that the data are normally distributed and representative of the study site, the data are not spatially or temporally correlated, and the standard deviation of the data to be collected is known with great accuracy.

The SPRT test in VSP works as follows if a map with a specific selected target population, e.g., the walls of a room, is drawn or loaded into VSP:

- The VSP user inputs the DQO parameters (α , β, null hypothesis, width of the gray region, known standard deviation (σ), and the threshold value) in the design dialog box. Also, the number of samples to collect on each sampling excursion is specified. When the "Apply" button in VSP is pressed, VSP places the required number of sampling locations for the first sampling occasion on the map. VSP also provides a listing of the geographical coordinates of the required samples.
- The samples are collected and analyzed for the bio/chem/rad threats of concern.
- The VSP user reopens the SPRT design dialog box, presses the "Input Values" button, and inputs the measurements obtained for the samples collected. The VSP user then closes the data input dialog box and VSP computes the mean and determines whether another round of sampling is needed before a decision can be made by the SPRT whether to accept the null hypothesis or the alternative hypothesis. The sample mean is plotted on a decision graph in the VSP "Graph View" for ease of interpretation.
- If another round of sampling is needed, the Apply button is pressed and VSP places the additional sampling locations on the map (avoiding existing sampling locations) and provides the geographical coordinates of the new samples.

The last three bullets above are repeated until there is enough data so that the SPRT can either accept the null hypothesis or the alternative hypothesis with the probabilities of making decision errors specified by the VSP user. It is useful to watch the VSP Graph View during the sequential sampling and testing process to see how close the SPRT is to making a decision.

If no sample areas on the target population map have been selected by the VSP user, or no map is being used, then the VSP user can enter as few or as many data values as desired. Also, in this case, it is not necessary to close the design dialog box in order to enter more data values.

Suppose the null hypothesis selected by the VSP user is that the target population is dirty. Then the SPRT determines that there is enough evidence to accept the null hypothesis if the mean of the sample values is greater than UL_d , where

$$UL_d = AL - \frac{\Delta}{2} + \frac{\sigma^2 A}{\Delta n} .$$

Also, the SPRT determines that there is enough evidence to accept the alternative hypothesis if the mean of the sample values is less than LL_d , where

$$LL_d = AL - \frac{\Delta}{2} - \frac{\sigma^2 B}{\Delta n}$$

and

$$A = \ln\frac{1-\beta}{\alpha}$$

$$B = \ln\frac{1-\alpha}{\beta}$$

ln is the natural logarithm

α is the maximum acceptable Type I decision error rate (probability of rejecting the null hypothesis when the null hypothesis is true)

β is the maximum acceptable Type II decision error rate (probability of accepting the null hypothesis when the null hypothesis is false)

Δ is the width of the gray region in the Decision Performance Goal Diagram (DPGD)

σ^2 is the variance of the data, assumed to be known with great accuracy

n is the number of samples collected thus far and used in the SPRT

AL is the action level (threshold value).

If the null hypothesis is that the site is clean, then the SPRT determines that there is enough evidence to accept the null hypothesis if the mean of the sample values is greater than UL_c and that there is enough evidence to accept the alternative hypothesis if the mean of the sample values is less than LL_c, where:

$$UL_c = AL + \frac{\Delta}{2} + \frac{\sigma^2 A}{\Delta n}$$

$$LL_c = AL = + \frac{\Delta}{2} - \frac{\sigma^2 B}{\Delta n}$$

The notation for these equations is defined above.

VSP projects the number of additional samples needed to make a decision by using the following algorithm:

1. Increase the value of n by 1.
2. Recalculate the upper and lower bounds using the new value of n.
3. If the sample mean falls outside the new boundaries, then the increase in n is given by VSP as the number of additional samples needed.
4. Repeat Steps 1 through 3 up to 100 times

2.1.3 Barnard's Sequential t Test

Barnard's sequential t-test (Barnard 1952) can be used in place of the SPRT discussed in Section 2.1.2 when the standard deviation is not known with great accuracy. The assumptions that underlie Barnard's test are that the data are sampled sequentially from a normal distribution and that the data are representative of the study site and are not spatially or temporally correlated.

The VSP user goes through the same steps to use the sequential t-test using the VSP dialogue box as was described above for the SPRT except that the user must initially supply the measurements for 10 samples collected from the target population of interest. VSP uses these data to compute a sample standard deviation for the first iteration of the sequential t-test.

Null Hypothesis: Site is Dirty

If the null hypothesis selected by the VSP user is that the target population is dirty, then the sequential t-test determines that there is enough evidence to accept the null hypothesis if

$$\ln(L_n) \geq \ln\left(\frac{1-\alpha}{\beta}\right)$$

where

L_n is the likelihood ratio test statistic, which is computed using the method described in Appendix A of Gilbert et al., (2002).

ln is the natural logarithm.

The definitions of α and β are given in Section 2.1.2. The sequential t-test determines that there is enough evidence to reject the null hypothesis and accept the alternative hypothesis if

$$\ln(L_n) \leq \ln\left(\frac{\alpha}{1-\beta}\right).$$

Finally, the sequential t-test determines that the information from the n samples is not sufficient to make a decision if

$$\ln\left(\frac{\alpha}{1-\beta}\right) < \ln(L_n) < \ln\left(\frac{1-\alpha}{\beta}\right),$$

in which case additional samples are collected and the sequential test repeated using the full data set of all measurements collected to date.

Null Hypothesis: Site is Clean

If the null hypothesis selected by the VSP user is that the site is clean, then the sequential t-test determines that there is enough evidence to accept the null hypothesis if

$$\ln(L_n) \leq \ln\left(\frac{\beta}{1-\alpha}\right),$$

that there is enough evidence to reject the null hypothesis and accept the alternative hypothesis if

$$\ln(L_n) \geq \ln\left(\frac{1-\beta}{\alpha}\right),$$

and that additional samples are needed if

$$\ln\left(\frac{\beta}{1-\alpha}\right) < \ln(L_n) < \ln\left(\frac{1-\beta}{\alpha}\right).$$

Appendix A in Gilbert et al., (August 2002) provides additional discussion of Barnard's sequential t test.

2.1.4 Collaborative Sampling

The Collaborative Sampling (CS) design uses two measurement techniques to obtain a cost effective estimate of the mean of the characteristic of interest for the specified target population. One measurement technique is the "standard analysis" (referred to in VSP as the expensive analysis method), and the other is a less expensive and less accurate measurement method (referred to in VSP as the inexpensive analysis method). The idea behind CS is to replace the need for collecting so many expensive analyses with obtaining a fewer number of those analyses to allow one to obtain a relatively large number of the inexpensive analyses. It works like this: At *n'* field locations selected using simple random sampling or grid sampling, the inexpensive analysis method is used. Then, for each of *n* of the *n'* locations, the expensive analysis method is also conducted. The data from these two analysis methods are used to estimate the mean and the standard error (SE: the standard deviation of the estimated mean). The method of estimating the mean and SE assumes there is a linear relationship between the inexpensive and expensive analysis methods. If the linear correlation between the two methods is sufficiently high (close to 1), and if the cost of the inexpensive analysis method is sufficiently less than that of the expensive analysis method, then CS is expected to be more cost effective at estimating the population mean than if the entire measurement budget was spent on obtaining only expensive analysis results at locations selected using simple random sampling or grid sampling.

The VSP CS module computes the values of *n'* and *n* that should be used to estimate the CS mean and SE. VSP also computes the mean, standard error and other outputs that can be used to assess the validity of the assumptions that underlie CS. The equations used for these computations are provided below. It should be noted that Gilbert (1987, Chapter 9) and other statisticians use the term Double Sampling instead of Collaborative Sampling. The term Collaborative Sampling rather than Double Sampling is used in VSP in order to prevent potential users from thinking that a Double Sampling design requires doubling the number of samples.

Method Used to Determine if Collaborative Sampling is Cost Effective

Before VSP computes *n'* and *n* it determines if CS is cost effective compared with using the entire measurement budget to obtain only expensive analysis results at target population locations selected using simple random sampling or grid sampling. If CS is found to be cost effective, then VSP computes *n'* and *n*. If CS is not cost effective, then VSP computes the number of field locations that should be collected and analyzed using only the expensive analysis method to estimate the mean and SE.

VSP declares CS to be cost effective if

$$\rho^2 > \frac{4R}{(1+R)^2}, \tag{3}$$

where

ρ = the true correlation coefficient between the expensive and inexpensive analysis measurements,

$R = c_{Ex} / c_{Inex}$,

c_{Ex} = the per unit cost of an expensive analysis, including the cost of collecting, handling, preparing, and measuring the sample, and

c_{Inex} = the per-unit cost of an inexpensive analysis, including finding the field location and conducting the inexpensive analysis method.

Equation (3) above is from Gilbert (1987, page 108). It is assumed that the following cost equation applies:

$C = c_{Ex} n + c_{Inex} n' =$ total dollars available for doing n' inexpensive analyses and n expensive analyses

Note that C does not include what might be termed "overhead" costs of project management, preparing the Quality Assurance Project Plan or the Sampling and Analysis Plan, QA/QC, and other such costs.

Method VSP Uses to Compute the Number of Expensive and Inexpensive Analysis Measurements when CS is Cost Effective

VSP computes *n'* and *n* such that the product of the total measurement cost, *C*, and the variance of the estimated mean, $\sigma^2_{\bar{x}_{cs}}$ is minimized. VSP uses the following formulas that were derived using the same method of proof used in Appendix A of EPA (2000a):

$$n' = \left[\frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \sigma^2_{total,ex}}{\Delta^2} + \frac{1}{2} Z^2_{1-\alpha} \right] \rho \left(\sqrt{R(1-\rho^2)} + \rho \right), \quad (4)$$

and

$$n = \left[\frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \sigma^2_{total,ex}}{\Delta^2} + \frac{1}{2} Z^2_{1-\alpha} \right] \left[1 - \rho^2 + \rho \sqrt{\frac{(1-\rho^2)}{R}} \right], \quad (5)$$

where

n' is the recommended minimum number of samples to measure with the inexpensive method,

n is the recommended minimum number (subset) of the *n'* samples to also measure with the expensive method,

α is the acceptable probability that the statistical test will falsely reject the null hypothesis,

β is the acceptable probability that the statistical test will falsely accept the null hypothesis,

Δ is the width of the gray region in the Decision Performance Goal Diagram (DPGD),

$\sigma_{total,ex}$ is the total standard deviation of the expensive measurements, including analytical error,

$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,

$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$,

ρ is the assumed correlation between the expensive and inexpensive measurements obtained on the same samples,

C_{ex} is the cost of making a single expensive measurement,

C_{inex} is the cost of making a single inexpensive measurement,

R is the cost ratio C_{ex}/C_{inex}, and

C is the total measurement cost, i.e., $C = C_{inex}\, n' + C_{ex}\, n$.

Method VSP Uses to Compute the Number of Expensive Analysis Measurements when CS is Not Cost-Effective,

VSP uses the following equation to compute the required number of expensive measurements, *n*, needed to compare a mean to a threshold using the Z test described below:

$$n = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \sigma^2_{total,ex}}{\Delta^2} + \frac{1}{2} Z^2_{1-\alpha} \tag{6}$$

which is derived in EPA (2000a, Appendix A). The parameters in Equation (6) are defined in Table 1.

Method VSP Uses to Estimate the Mean and Standard Error when CS is Cost-Effective

The estimated mean and SE (standard deviation of the estimated mean) are computed assuming that there is a linear relationship between the expensive and inexpensive measurements obtained on the same set of n samples. This assumption should be verified by the VSP user before CS in VSP is used. After the VSP CS design is determined and the resulting measurements are entered into VSP, VSP shows the linear regression plot of these data. This plot should be examined to verify that the assumption of a linear relationship does indeed seem reasonable. Also, VSP computes the correlation coefficient, ρ, using the inexpensive and expensive data. The VSP user should use this correlation and the cost ratio, R, in the CS module of VSP to see if CS is still considered to be more cost efficient than simple random sampling.

The process used by VSP in the CS module to compute the mean and SE is given in the following steps:

1. The *n'* inexpensive and *n* expensive analysis measurements are made by the VSP user and entered into VSP. Let x_{Ex_i} and x_{Inex_i} denote the expensive and inexpensive measurements, respectively, on the i^{th} unit.
2. VSP estimates the mean by computing $\bar{x}_{cs}$ (cs stands for collaborative sampling) as follows

$$\bar{x}_{cs} = \bar{x}_{Ex} + b\left(\bar{x}_{n'} - \bar{x}_{Inex}\right), \tag{7}$$

   where $\bar{x}_{Ex}$ and $\bar{x}_{Inex}$ are the means of the n expensive and inexpensive measurements, respectively, $\bar{x}_{n'}$ is the mean of the n' inexpensive values, and b is the slope of the estimated regression of expensive on inexpensive values.
3. VSP computes the estimated standard error (standard deviation of $\bar{x}_{cs}$) as follows:

$$SE = \sqrt{s^2(\bar{x}_{cs})} = \sqrt{s^2_{Ex \bullet Inex}\left[\frac{1}{n} + \frac{(\bar{x}_{n'} - \bar{x}_{Inex})^2}{(n-1)s^2_{Inex}}\right] + \frac{s^2_{Ex} - s^2_{Ex \bullet Inex}}{n'}}, \quad (8)$$

where s^2_{Ex} and s^2_{Inex} are the variances of the *n* expensive and inexpensive measurements, respectively, and $s^2_{Ex \bullet Inex}$ is the residual variance about the estimated linear regression line. The equations used to calculate the quantities in Equations (7) and (8) are:

$$\bar{x}_{Ex} = \frac{1}{n}\sum_{i=1}^{n} x_{Ex_i}$$

$$\bar{x}_{Inex} = \frac{1}{n}\sum_{i=1}^{n} x_{Inex_i}$$

$$\bar{x}_{n'} = \frac{1}{n'}\sum_{i=1}^{n'} x_{Inex_i}$$

$$b = \frac{\sum_{i=1}^{n}\left(x_{Ex_i} - \bar{x}_{Ex}\right)\left(x_{Inex_i} - \bar{x}_{Inex}\right)}{\sum_{i=1}^{n}\left(x_{Inex} - \bar{x}_{Inex}\right)^2}$$

$$s^2_{Ex} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{Ex_i} - \bar{x}_{Ex}\right)^2$$

$$s^2_{Inex} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{Inex_i} - \bar{x}_{Inex}\right)^2$$

$$s^2_{Ex \bullet Inex} = \frac{n-1}{n-2}\left(s^2_{Ex} - b^2 s^2_{Inex}\right)$$

Method VSP Uses to Compute the Mean and SE when CS is Not Cost-Effective

The process used by VSP to compute the estimated mean and SE (standard deviation of the estimated mean) when CS is not cost effective compared to simple random sampling is given in the following steps:

1. After the n samples are collected and the ***n*** *expensive* measurements have been obtained, the VSP user enters them into VSP. Let x_i denote the expensive measurement on the ith unit.

2. VSP estimates the mean by computing $\bar{x}$ as follows:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (9)$$

3. VSP computes the standard deviation of the ***n*** measurements as follows:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

4. VSP computes the standard error (standard deviation of $\bar{x}$) as follows:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{10}$$

Method VSP Uses to Test if the True Mean Exceeds a Specified Threshold Value When CS is Cost-Effective

If the null hypothesis is H_o: true mean ≥ threshold value, then VSP computes

$$Z = \frac{\bar{x}_{cs} - ThresholdValue}{\sqrt{s^2_{\bar{x}_{cs}}}} \tag{11}$$

and H_o is rejected if $Z \le -z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)^{th}$ percentile of the standard normal distribution. For example, if the VSP user specifies that $\alpha = 0.05$, then H_o is rejected if $Z \le -1.645$.

If the null hypothesis is H_o: true mean ≤ threshold value, then VSP computes Equation (11) and H_o is rejected if $Z \ge z_{1-\alpha}$,where $z_{1-\alpha}$ is defined in Table 1.

Method VSP Uses to Test if the True Mean Exceeds a Specified Threshold Value when CS is Not Cost-Effective

If the null hypothesis is H_o: true mean ≥ threshold value, then VSP computes

$$Z = \frac{\bar{x} - ThresholdValue}{\sqrt{s^2_{\bar{x}}}}, \tag{12}$$

where $\bar{x}$ and $s^2_{\bar{x}}$ are computed using Equations (9) and (10), respectively, and H_o is rejected if $Z \le -z_{1-\alpha}$.

If the null hypothesis is H_o: true mean ≤ threshold value, then VSP computes Equation (12) and H_o is rejected if $Z \ge z_{1-\alpha}$

Statistical Assumptions

The assumptions that underlie the equations used to compute n' and n and to test the hypotheses are:

1. There is an underlying linear relationship between the expensive and inexpensive analysis methods
2. The true correlation coefficient, ρ, is well known from prior studies or has been well estimated using a preliminary sampling study of the target population or a very similar target population
3. Collaborative sampling is more cost effective than simple random sampling
4. The optimum values of n' and n are used to estimate the true mean, i.e., the values of n' and $\boldsymbol{n}$ computed, assuming the value of ρ used is valid
5. The field sampling locations are selected using simple random sampling or systematic grid sampling
6. The costs C_{Ex} and C_{Inex} are appropriate
7. The measurements are normally or approximately normally distributed.

2.2 Compare Median to a Threshold when Data are not Normally Distributed

2.2.1 MARSSIM Sign Test

The sign test can be used to test whether the true median concentration of the target population exceeds a fixed action level (upper limit value). The assumptions that underlie this test are that the data are representative of the study site and that the data are not spatially or temporally correlated. The probability distribution of the data need not be known. Hence, the Sign test is a distribution-free test.

The formula used to compute the approximate number of samples, n, needed for the sign test when only r = 1 analytical replicates per field sample will be obtained is given in the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (EPA 1997, p. 5-33):

$$n = 1.20\left[\frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{4\left(SignP - 0.50\right)^2}\right], \tag{13}$$

where

$$SignP = \phi\left(\frac{\Delta}{s_{total}}\right). \tag{14}$$

If the VSP user makes use of the MQO module, in which case r = 1, 2, or 3 analytical replicates per field sample can be used, then

$$SignP = \phi\left[\frac{\Delta}{\left(s^2_{sample} + \frac{s^2_{analytical}}{r}\right)^{1/2}}\right].$$

Equation (14) is derived in Gogolak, Powers, and Huffert (1997, pp. 9-3 and 9-7).

VSP denotes the sign test as the MARSSIM sign test to indicate that Equation (13) is taken from the MARSSIM document (EPA 1997). The function ϕ in Equation (14) is defined in Table 1 as denoting the cumulative distribution function of the standard normal distribution (the normal distribution with mean zero and standard deviation 1). Hence, by definition, $\phi\left(\frac{\Delta}{s_{total}}\right)$ is the fraction of the bell-shaped standard normal distribution that is less than or equal to the value of $\frac{\Delta}{s_{total}}$. Note that $\frac{\Delta}{s_{total}} > 0$ because Δ, the width of the gray region in the DPGD, must be greater than zero. VSP computes Sign P = $\phi\left(\frac{\Delta}{s_{total}}\right)$ using the value of $\frac{\Delta}{s_{total}}$ specified by the VSP user. EPA (1997, Table 5.4, p. 5-32) provides values of Sign P for selected values of $\frac{\Delta}{s_{total}}$ between 0.1 and 3.0.

Equation (13) is based on the formula for n proposed by Noether (1987, Section 2.1) for the sign test. His equation for n is identical to Equation (13) except that Noether used the constant 1.00 in place of 1.20. The assumptions that underlie Noether's equation are that the data are representative of the underlying population, the data are not correlated, and the computed sign test statistic (the quantity computed using the data to make the test) is approximately normally distributed.

Noether (1987) indicates that the value of n computed using his method should achieve the performance requirements for the sign test (as specified by α, β and Δ) unless the computed n is "quite small." The MARSSIM report (EPA 1997) used 1.20 instead of 1.00 in Equation (13) to provide subjective added confidence that the larger value of n computed would result in achieving the performance requirements specified for the test by the VSP user.

Equation (13) is used in VSP primarily because it is used in EPA (1997), which is a multi-agency consensus document that was developed collaboratively by four federal agencies having authority and control over radioactive materials: the U.S. Departments of Defense and Energy, the U.S. Nuclear Regulatory Commission, and the EPA.

2.2.2 Wilcoxon Signed Ranks Test

The Wilcoxon Signed Ranks test can be used to test whether the true median or mean of the target population exceeds the fixed action level. The assumptions needed for this test are that the data are representative of the study site, are not spatially or temporally correlated, and have a symmetric (but not necessarily normal) distribution. Note that the test applies to either the mean or median when the assumption of symmetry is true, because those two parameters have identical values for symmetric distributions.

The equation used in VSP to compute the minimum recommended number of samples, n, needed for the test when the VSP user specifies that only r = 1 analytical replicates from each field sample will be obtained, is

$$n = 1.16\left[\frac{s_{total}^2\left(Z_{1-\alpha}+Z_{1-\beta}\right)^2}{\Delta^2}+0.5Z_{1-\alpha}^2\right]. \quad (15)$$

This equation is also used in EPA (2000b, pp. 3-12).

If the MQO module is used, in which case r = 1, 2, or 3 can be used, then the equation for n is

$$n = 1.16\left[\frac{\left(s_{sample}^2+\frac{s_{analytical}^2}{r}\right)\left(Z_{1-\alpha}+Z_{1-\beta}\right)^2}{\Delta^2}+0.5Z_{1-\alpha}^2\right]. \quad (16)$$

Equation (15) is identical to Equation (1) and Equation (16) is identical to Equation (2) except for the 1.16 multiplier. The constant 1.16 is used because it is known (Conover 1999, p. 363) that the Wilcoxon Signed Ranks test will require no more than 1.16 times as many samples as the t-test to achieve the α and β decision error rate test performance specifications when the data are normally distributed.

Noether (1987, Section 2.2) developed an alternative to Equation (15) for computing n for the Wilcoxon Signed Ranks test. His method does not require that the data have a symmetric distribution. His method is based on the assumptions that the data are representative of the underlying target population, the data are not correlated, and the computed Wilcoxon Signed Ranks test statistic (the quantity computed using the data to make the test) is approximately normally distributed. Noether (1987) indicates that the value of n computed using his method should achieve the performance requirements for the test (as specified by α and β) unless the n computed using his method is "quite small."

2.3 Compare Individual Measurements to a Threshold Value

2.3.1 Detecting Hot Spots of Contamination

Suppose a VSP user wants to determine the spacing of sampling or measurement locations in the target population that should be used to detect a hot spot of specified size. The VSP user can specify that the sampling locations should be laid out in a square, rectangular, or triangular pattern. VSP determines the optimum spacing between sampling locations using an implementation of a computer program called ELIPGRID-PC (Davidson 1995b). The events that led to the development of this program are now briefly summarized.

Singer and Wickman (1969) published an algorithm for calculating the probability of locating elliptical hot spots when sampling is done on a square, rectangular, or triangular grid pattern over space. Singer (1972) published a FORTRAN IV computer program, ELIPGRID, to automate the hot spot probability calculations. He also evaluated the efficiency of square and triangular grids in the search for elliptically-shaped hot spots (Singer 1975). Zirschky and Gilbert (1984) developed nomographs for answering the same questions addressed by ELIPGRID. These nomographs were published in Gilbert (1987, Chapter 10) along with examples of the calculations. Davidson (1995b) wrote and published ELIPGRID-PC for the personal computer, an upgraded and corrected version of the original ELIPGRID algorithm. ELIPGRID-PC was subsequently incorporated into VSP.

The assumptions that underlie the ELIPGRID-PC and the VSP implementation of that code are, from Gilbert (1987, pp. 119-120):

- The target (hot spot) is circular or elliptical. For subsurface targets, this assumption applies to the projection of the target to the ground surface.
- Samples or measurements are taken on a square, rectangular, or triangular pattern.
- The distance between grid points is much larger than the area sampled, measured, or cored at grid points; that is, a very small proportion of the area being studied can actually be measured.
- The definition of "hot spot" is clear and unambiguous; the types of measurement and the levels of contamination that constitute a hot spot are clearly defined.
- There are no measurement misclassification errors; that is, no errors are made in deciding when a hot spot has been found.

The computations conducted by VSP to determine the spacing of sampling locations are described in Singer (1972) and Davidson (1995b).

2.3.2 One-Sided Upper Tolerance Limits when Data are Normally Distributed

Tolerance limits can be used to statistically test whether the target population, e.g., a specified area or room in a building, is contaminated with biological agents, chemicals, or radionuclides at concentrations greater than their respective fixed ALs. The statistical meaning, use, and computation of tolerance limits are discussed in Hahn and Meeker (1991) and Helsel (2005, Chapter 6). VSP computes the number of measurements, n, needed to compute a tolerance limit to statistically test if a fixed AL has been exceeded. A discussion of this use of tolerance limits is given by Millard and Neerchal (2001, page 339). If the VSP user inputs the n measurements into VSP on the Data Analysis tab of the dialog box, then VSP computes the tolerance limit and conducts the statistical test.

VSP computes a one-sided upper tolerance limit, which is identical to a one-sided upper confidence limit on a specified percentile of the population of measurements. The P^{th} percentile is the value above which (1-P)% of the population lies and below which P% of the population lies, where 0< P <100.

A one-sided upper tolerance limit on the P^{th} percentile of a population, denoted by $UTL_{P,\alpha}$, is a value computed using the n measurements such that at least P% of the population of measurements is less than $UTL_{P,\alpha}$ with 100(1-α)% confidence. For example, if P = 90 and α = 0.05, then at least 90% of the population is less than the computed value $UTL_{90,0.05}$ with 95% confidence.

The method VSP uses to compute tolerance limits using n data depends on the probability distribution of the population of measurements. It is assumed in this section that the measurements made in buildings are normally distributed. If this assumption is false, the computed tolerance limits will not be accurate and decisions based on that computed limit may be in error. Statistical goodness-of-fit tests and graphical plots within VSP should always be used to evaluate if the measurements are from a normal distribution before using the normal tolerance limits. Also, VSP can compute upper tolerance limits when the underlying distribution is uncertain or unknown (Section 2.3.3).

The other assumptions that underlie the use of tolerance limits are:

- Representative measurements have been obtained from a defined target population (e.g., a wall, section of a wall, the floor in a hallway, the walls and floors of a selected set of rooms, etc.) using

simple random sampling, systematic sampling on a grid, or some other suitable probability-based sampling design.

- The measurements are statistically independent, i.e., there is no spatial correlation (no spatial patterns) of contaminant levels throughout the target population.

The assumption of statistical independence implies that tolerance limits may be most useful for building areas that are not expected to contain "hot spots" or other dominant spatial patterns. Hence, they may be most useful after decontamination has occurred and the objective is to test if the building is ready to be re-occupied.

Method Used in VSP to Compute a Percentile of a Normal Distribution

The P[th] percentile of a normal distribution, denoted by x_P, is computed from n measurements as follows (Millard and Neerchal, 2001, page 276, Equation 5.138):

$$x_P = \bar{x} + Z_P s, \tag{17}$$

where

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = mean of the n measurements, (18)

x_i = i[th] measurement,

Z_P = P[th] percentile of the standard normal distribution (normal distribution with mean zero and standard deviation 1), e.g., if P = 95, then Z_{95} = 1.645. Tables of Z_P values are in many statistical books, e.g., Gilbert (1987, Table A1, page 254), (19)

$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$ = standard deviation of the n measurements. (20)

Method Used in VSP to Compute a One-Sided Upper Tolerance Limit on the P[th] Percentile of a Normal Distribution

A one-sided upper tolerance limit for the P[th] percentile of a normal distribution is computed as follows:

$$UTL_{P,\alpha} = \bar{x} + t_{n-1,Z_P\sqrt{n},1-\alpha}\, s, \tag{21}$$

where $\bar{x}$ and s are computed using Equations (18) and (20), respectively, and

$t_{n-1,Z_P\sqrt{n},1-\alpha}$ = the 100(1-α)[th] percentile of the non-central t distribution with (n-1) degrees of freedom and non-centrality parameter $Z_P\sqrt{n}$.

The null hypothesis being tested by comparing the $UTL_{P,\alpha}$ to the AL is

H_o: P^{th} percentile of the population ≥ AL.

The H_o is rejected and the statement "Conclude Site is Clean" is stated if $UTL_{P,\alpha}$ < AL.

The statement "Conclude Site is Dirty" is stated if $UTL_{P,\alpha} \geq$ AL.

Method Used in VSP to Compute the Number of Measurements Needed to Test H_o Using a One-Sided Upper Tolerance Limit on the P^{th} Percentile of a Normal Distribution

VSP uses the exact method given by Lyles and Kupper (1996, Equation 5) to compute the number of measurements, n, needed to test the null hypothesis H_o. The procedure is to find the smallest integer n such that

$$t_{n-1,-Z_P\sqrt{n},\alpha} - t_{n-1,Z_{1-\theta}\sqrt{n},1-\beta} \geq 0, \tag{22}$$

where

$t_{n-1,-Z_P\sqrt{n},\alpha}$ = the $100(\alpha)^{th}$ percentile of the non-central t distribution with (23)
n-1 degrees of freedom and non-centrality parameter $-Z_P\sqrt{n}$,

$t_{n-1,-Z_{1-\theta}\sqrt{n},1-\beta}$ = the $100(1-\beta)^{th}$ percentile of the non-central t distribution with (24)
n-1 degrees of freedom and non-centrality parameter $-Z_{1-\theta}\sqrt{n}$,

$$\theta = 1 - \phi\left[\frac{UBGR - LBGR}{\sigma} + Z_P\right] \tag{25}$$

α is the false rejection rate specified by the VSP user, i.e., α is the probability the VSP user can tolerate that the data will falsely indicate that the null hypothesis, H_o, should be rejected,

β is the false acceptance rate specified by the VSP user, i.e., β is the probability the VSP user can tolerate that the data will falsely indicate that the null hypothesis should be accepted,

$\phi(x)$ is the probability that a measurement from a standard normal distribution falls below the value x,

UBGR is the upper bound of the gray region of the Decision Performance Goal Diagram (DPGD) specified by the VSP user,

LBGR is the lower bound of the gray region of the DPGD specified by the VSP user,

Z_P is defined by Equation (19) above,

σ is the true standard deviation of all possible measurements from the target population. In practice, an estimate of σ is used in Equation (25).

2.3.3 Nonparametric (Distribution-Free) One-Sided Upper Tolerance Limits

Nonparametric (distribution-free) tolerance limits can be used to statistically test whether a specified area or room in a building is contaminated with biological agents, chemicals or radionuclides at concentrations greater than their respective fixed action levels (ALs). The VSP user can have VSP compute the number of measurements needed to compute a nonparametric tolerance limit to statistically test if a fixed AL has been exceeded. VSP will also determine the tolerance limit using the n measurements obtained and does the statistical test. This use of tolerance limits is discussed in Millard and Neerchal (2001, page 339).

A nonparametric tolerance limit is valid regardless of the probability distribution of the population of measurements. That is, the data distribution need not be known. If the distribution is known with confidence to be a normal distribution, the tolerance limits for the normal distribution should be used. If the distribution is known with confidence to be a lognormal distribution then tolerance limits for the lognormal distribution should be used.

A one-sided upper tolerance limit is identical to a one-sided confidence limit on a specified percentile P of the population of measurements. The P^{th} percentile is the value above which (1-P)% of the population lies and below which 100P% of the population lies.

A one-sided upper tolerance limit on the P^{th} percentile of a population is a value, denoted here by $UTL_{P,\alpha}$, such that at least P% of the population of measurements is less than $UTL_{P,\alpha}$ with 100(1-α)% confidence. For example, if P = 90 and α = 0.05, then at least 90% of the population is less than the computed value $UTL_{90,0.05}$ with 95% confidence.

The following assumptions are needed when using nonparametric tolerance limits:

- Representative measurements have been obtained from a defined target population (e.g., a wall, section of a wall, the floor in a hallway, the walls and floors of a selected set of rooms, etc.) using simple random sampling, systematic sampling on a grid, or some other suitable probability-based sampling design.
- The measurements are statistically independent, i.e., there is no spatial correlation (no spatial patterns of contaminant levels throughout the target population).
- There are no "outliers," i.e., there are no observations that are mistakes or that do not belong to the population being studied.

The assumption of statistical independence implies that tolerance limits may be most useful for building areas that are not expected to contain "hot spots" or other dominant patterns. Hence, they may be most useful after decontamination has occurred and the objective is to test if the building is ready to be re-occupied.

Method Used in VSP to Determine a Nonparametric One-Sided Upper Tolerance Limit on the P^{th} Percentile of Any Distribution

Nonparametric tolerance limits are determined in VSP as follows:

1. The VSP user specifies the desired values of P and α and the action level, AL.

2. VSP computes the number of measurements that should be obtained using the following equation (from Hahn and Meeker, 1991, page 169):

$$n = \frac{\ln(\alpha)}{\ln(P)}$$

3. The VSP user obtains the n representative measurements from the target population using a probability-based design, e.g., using simple random sampling or sampling on a square or triangular grid pattern.

4. Then

   $UTL_{P,\alpha}$ = the largest of the n measurements obtained is the nonparametric upper 100(1-α)% tolerance limit on the P^{th} percentile.

5. If $UTL_{P,\alpha} \geq$ AL, then the null hypothesis

   H_o: P^{th} percentile of the population $\geq$ AL

   is not rejected and the statement "Conclude Site is Dirty" is given by VSP.

6. If $UTL_{P,\alpha}$ < AL, then the null hypothesis is rejected and the statement "Conclude the Site is Clean" is given by VSP.

Clearly, the assumptions of representative measurements and "no outliers" are important because if one or more outliers occur, the maximum of the n measurements will not be a valid nonparametric upper tolerance limit on a percentile.

2.3.4 Nonparametric (Distribution-Free) Compliance Sampling for Attributes

There may be occasions when decisions about the need for decontamination, additional decontamination, or some other action will depend on how much of a room or set of rooms is contaminated above an action level (AL) or is "defective" in some way. Compliance sampling for attributes (Schilling 1982, Chapter 17, pages 474-482) can be used to statistically test whether the fraction of a room (or suite of rooms) that is contaminated above the action level (AL) is less than a prescribed upper limit. VSP determines the number, n, and location of grid units (defined below) in the room that must be measured or inspected to make this determination. If one or more of the grid units equal or exceed the AL or are defective as defined by the VSP user, then the required confidence is not achieved. This section documents the compliance sampling methodology in Schilling (1982), which is implemented in VSP.

Definitions

Grid Unit	A grid unit is a small unit area on a room surface that will be inspected or measured to determine if it exceeds the action level (AL) or is otherwise defective. For example, a single grid unit might be a 10cm by 10cm area that will be swiped and then measured for a biological or chemical agent.
N	N is the total number of grid units in the target population.
n	n is the number of grid units that are randomly selected and measured or inspected. The value of n is computed in VSP as described below.

AL	AL is the action level, e.g., a contaminant concentration or the value of some other metric for a grid unit that indicates the grid unit is not acceptable, i.e., that the grid unit is "defective."
P_o	P_o is the maximum tolerable proportion of defective grid units in the target population of N grid units.
H_o	H_o is the null hypothesis H_o: The proportion of the N grid units that are defective $\leq P_o$
H_a	H_a is the alternative hypothesis H_a: The proportion of the N grid units that are defective $> P_o$
C	C is the "acceptance number," i.e., the number of the n measured grid units that can exceed the AL (i.e., allowed to be defective) without rejecting the null hypothesis. For compliance sampling, C is always equal to zero.
Decision Rule	If one or more of the n grid units measured or inspected exceed the AL or are otherwise defective, then reject H_o and accept H_a and take the action needed when H_o is rejected.

Assumptions Underlying Compliance Sampling

- The size of the grid unit has been determined to be appropriate for the measurement (inspection) method to be performed. For example, an appropriate grid unit size might be a 10cm by 10cm surface area.
- The total number of grid units in the target population, N, is known. All N grid units are the same size.
- n of the N grid units are selected using simple random sampling or perhaps in a square, rectangular, or triangular pattern that has a randomly selected starting point.
- The n grid units selected must be representative of the total population of N grid units.
- Each of the n grid units are measured or inspected using an approved method that has a very low chance of making measurement or inspection mistakes.

Calculations

VSP uses the following steps to calculate the number, n, of the N grid units that need to be randomly selected and measured (inspected) [Schilling (1978) and Schilling (1982, pages 476-479)]:

1. The VSP user specifies N and P_o, which are defined above.
2. The VSP user specifies the % confidence required that less than 100 P_o % of the N grid units in the target population are defective.
3. VSP determines the factor P_g from Table 2, which is based on Table 17-2 in Schilling (1982, page 478). This table gives the correspondence between the required confidence and P_g.

4. VSP computes $D = N(P_o / P_g)$.
5. VSP computes $n = N * f$, where VSP selects f from Table 3, which is Table 17-1 in Schilling (1982, page 477). The cells of Table 3 give values of *D*. VSP determines *f* from the row and column headings that correspond to the value of *D* computed in Step 4 above.

Table 2. Values of the Factor P_g Needed to Compute n for Compliance Sampling

Required Confidence	50%	75%	90%	95%	97.5%	99%	99.5%	99.9%
P_g	0.301	0.602	1.000	1.301	1.602	2.000	2.300	2.996

Table 3. Values of the Factor D Needed to Compute n for Compliance Sampling

f	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
.0	∞	229.1053	113.9741	75.5957	56.4055	44.8906	37.2133	31.7289	27.6150	24.4149
.1	21.8543	19.7589	18.0124	16.5342	15.2668	14.1681	13.2064	12.3576	11.6028	10.9272
.2	10.3189	9.7682	9.2674	8.8099	8.3902	8.0039	7.6471	7.3165	7.0093	6.7231
.3	6.4557	6.2054	5.9705	5.7496	5.5415	5.3451	5.1594	4.9836	4.8168	4.6583
.4	4.5076	4.3640	4.2270	4.0963	3.9712	3.8515	3.7368	3.6268	3.5212	3.4196
.5	3.3219	3.2278	3.1372	3.0497	2.9652	2.8836	2.8047	2.7283	2.6543	2.5825
.6	2.5129	2.4454	2.3797	2.3159	2.2538	2.1933	2.1344	2.0769	2.0208	1.9660
.7	1.9125	1.8601	1.8088	1.7586	1.7093	1.6610	1.6135	1.5667	1.5207	1.4754
.8	1.4307	1.3865	1.3428	1.2995	1.2565	1.2137	1.1711	1.1286	1.0860	1.0432
.9	1.0000	0.9562	0.9117	0.8659	0.8184	0.7686	0.7153	0.6567	0.5886	0.5000

2.3.5 Nonparametric (Distribution-Free) Acceptance Sampling for Attributes

Acceptance sampling for attributes can be used instead of compliance sampling (Section 2.3.4) to test if a room is sufficiently free from contamination. The two methods are very similar, except that acceptance sampling permits one or more grid units to be defective, e.g., exceed the action level, without concluding that the allowed level of contamination has been exceeded. VSP determines the "acceptance number," C, and the location and number, n, of grid units (defined below) in the room that must be measured or inspected to make this determination. C is the number of grid units that are allowed to be defective without concluding that the allowed level of contamination has been exceeded.

Definitions

Grid Unit, N, n, AL Defined in Section 2.3.4

H_o H_o is the null hypothesis that is being tested:

H_o: The number of the N grid units that are defective = $D_o = NP_o$

H_o is assumed to be true unless the data strongly indicate otherwise.

P_o P_o is the maximum proportion (fraction) of the target population of N grid units that are allowed to be defective, e.g., contaminated above the AL. P_o is an input of the VSP user.

H_a H_a is the alternative hypothesis

H_a: The number of the N grid units that are defective = $D_a = NP_a$

where $D_a > D_o$. H_a is accepted as being true if H_o is rejected.

P_a — P_a is an unacceptable proportion of the N grid units that have contamination greater than the AL. P_a is an input of the VSP user.

C — C is the "Acceptance Number," i.e., the number of grid units that are allowed to exceed the AL (allowed to be defective) without rejecting the null hypothesis. C is computed in VSP as described below.

α — α (alpha) is the probability that can be tolerated that the n data indicate H_o should be rejected when it is really true.

β — β (beta) is the probability that can be tolerated that the n data indicate H_o should not be rejected when H_o is really false.

$1-\beta$ — $1-\beta$ is the probability that the n data indicate H_o should be rejected when H_o is really false. $1-\beta$ is the "power" of the statistical test.

Decision Rule — If more than C of the n grid units measured or inspected are defective, e.g., exceed the AL, then reject H_o and accept H_a and take the needed action.

Assumptions Underlying Schilling's Acceptance Sampling for Attributes

- The size of the grid unit has been determined to be appropriate for the measurement (inspection) method to be performed. For example, an appropriate grid unit size might be a 10cm by 10cm surface area.
- The total number of grid units in the target population, N, is known. All N grid units are the same size.
- n of the N grid units are selected using simple random sampling or perhaps in a square, rectangular, or triangular pattern that has a randomly selected starting point.
- The n grid units selected must be representative of the total population of N grid units.
- Each of the n grid units are measured or inspected using an approved method that has a very low chance of making measurement or inspection mistakes.

Iterative Method Used in VSP to Compute the Number of Grid Units to be Measured, n, and the Acceptance Number, C

The method provided here is based on Equations (26) and (27) below, which are from Desu and Raghavarao (1990, pages 66 and 67) and Bowen and Bennett (1988). The idea is to determine the acceptance number C and the minimum value of n such that Equations (26) and (27) are satisfied:

$$\text{Probability(More than C Grid Units are Defective} \mid H_o \text{ is True}) =$$

$$\sum_{x=C+1}^{\min(D_0,n)} \frac{\binom{D_0}{x}\binom{N-D_0}{n-x}}{\binom{N}{n}} \equiv S_0 \le \alpha\,, \tag{26}$$

which is equivalent to

$$\sum_{x=C+1}^{\min(D_o,n)} \frac{D_o!(N-D_o)!n!(N-n)!}{x!N!(D_o-x)!(n-x)!(N+x-D_o-n)!} \equiv S_0 \le \alpha$$

and

Probability(More than C Grid Units are Defective$|H_a$ is True) =

$$\sum_{x=C+1}^{\min(D_A,n)} \frac{\binom{D_A}{x}\binom{N-D_A}{n-x}}{\binom{N}{n}} \equiv S_A \ge 1-\beta\,, \tag{27}$$

which is equivalent to

$$\sum_{x=C+1}^{\min(D_o,n)} \frac{D_a!(N-D_a)!n!(N-n)!}{x!N!(D_a-x)!(n-x)!(N+x-D_a-n)!} \equiv S_a \ge 1-\beta\,.$$

A solution is obtained by choosing successively larger values of C beginning with C = 0, and then determining the minimum value of n for which both Equation (26) and Equation (27) are satisfied.

Iterative Algorithm used by VSP

1. Begin with setting C = 0.

2. Search for an n between 0 and N that will satisfy both Equation (26) and Equation (27).

   This is done by performing the following binary search on the range $[n_L, n_u]$ starting with $n_L = 0$ and $n_u = N$

   Binary Search
   Evaluate Equations (26) and (27) with n = midpoint of the range $[n_L, n_u]$.
   - If both $S_0 \le \alpha$ and $S_A \ge 1-\beta$ are satisfied, then go to Step 3
   - If only $S_0 \le \alpha$ is satisfied, then a larger n is needed. Search on the range $[n, n_u]$ (the upper half of the current range).
   - If only $S_A \ge 1-\beta$ is satisfied, then a smaller n is needed. Search on the range $[n_L, n]$ (the lower half of the current range).
   - If neither condition is satisfied, then increment C and begin the search again at Step 2.

3. Find the minimum n that will satisfy Equations (26) and (27).

Perform another binary search on the range $[n_L, n_u]$, where n_u is the n found in Step 2 that satisfies Equations (26) and (27), and n_L is the last n_L from Step 2. Note that n_L is the largest value of $n < n_u$ that was found not to satisfy Equations (26) and (27), and n_u is the smallest value of n found that does satisfy those equations.

Binary Search
Evaluate Equations (26) and (27) with n set at the midpoint of the range $[n_L, n_u]$

- If both equations are satisfied, then a smaller satisfactory n has been found. Make the range $[n_L, n]$, where $n = n_u$.
- Otherwise, a larger unsatisfactory n has been found. Change the range to $[n, n_u]$, where $n = n_L$.

When the length of the range is 1, that is, $n_u - n_L \leq 1$, then $n = n_u$ is the minimum value of n that satisfies both Equations (26) and (27) for the minimum value of C.

Example 1 of the Iterative Method

$N = 5$
$P_o = 0.05$ $(D_o = 0)$
$P_a = 0.20$ $(D_a = 1)$
$\alpha = 0.01$
$\beta = 0.05$

Steps	C	n	S_0	S_A	$S_0 \leq \alpha$	$S_A \geq 1-\beta$
Start the search with C = 0 and n = midpoint of the full range [0,5]	0	3	0.00000000000000000	0.60000000000000064	Yes	No
S_A condition not satisfied, so take n = midpoint of upper half [3,5]	0	4	0.00000000000000000	0.80000000000000004	Yes	No
S_A condition not satisfied, so take n = midpoint of upper half [4,5]	0	5	0.00000000000000000	1.0000000000000000	Yes	Yes
Both satisfied, and final range [4,5] is of length 1, so minimum n for C = 0 is n = 5	**0**	**5**				

Example 2 of the Iterative Method

N = 10
P_o = 0.10 (D_o = 1)
P_a = 0.20 (D_a = 2)
α = 0.05
β = 0.10

Steps	**C**	**n**	**S_0**	**S_A**	**$S_0 \leq \alpha$**	**$S_A \geq 1-\beta$**
Start the search with C = 0 and n = midpoint of the full range [0,10]	0	5	0.49999999999999906	0.77777777777777513	No	No
Neither satisfied, so start over with C=1 and n = midpoint of the full range [0,10]	1	5	0.00000000000000000	0.22222222222222132	Yes	No
S_A not satisfied, so take n = midpoint of upper half [5,10]	1	8	0.00000000000000000	0.62222222222221812	Yes	No
S_A not satisfied, so take n = midpoint of upper half [8,10]	1	9	0.00000000000000000	0.79999999999999505	Yes	No
S_A not satisfied, so take n = midpoint of upper half [9,10]	1	10	0.00000000000000000	1.0000000000000000	Yes	Yes
Both satisfied, and final range [9,10] is of length 1, so solution is C = 1 and n = 10	**1**	**10**				

Example 3 of the Iterative Method

N = 50
P_o = 0.02 (D_o = 1)
P_a = 0.10 (D_a = 5)
α = 0.05
β = 0.10

Steps	C	n	S_0	S_A	$S_0 \leq \alpha$	$S_A \geq 1-\beta$
Start the search with C = 0 and n = midpoint of the full range [0,50]	0	25	0.49999999999999906	0.97492401215806124	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [0,25]	0	13	0.26000000000001089	0.79426787366194951	No	No
Neither satisfied, so start over with C = 1 and n = midpoint of the full range [0,50]	1	25	0.00000000000000000	0.82566217976552680	Yes	No
S_A not satisfied, so take n = midpoint of upper half [25,50]	1	38	0.00000000000000000	0.99074836224960339	Yes	Yes
Both satisfied, so now begin search for minimum by taking n = midpoint of [25,38]	1	32	0.00000000000000000	0.94974041420452349	Yes	Yes
Both still satisfied, so take n = midpoint of [25,32]	1	29	0.00000000000000000	0.90847760010573875	Yes	Yes
Both still satisfied, so take n = midpoint of [25,29]	1	27	0.00000000000000000	0.87127659574468153	Yes	No
Conditions not satisfied, so change range to [27,29]	1	28	0.00000000000000000	0.89090128188185957	Yes	No
Conditions not satisfied, and the new range [28,29] has length 1, so solution is C = 1 and n = 29	**1**	**29**				

Example 4 of the Iterative Method

N = 45
P_o = 0.07 (D_o = 3)
P_a = 0.30 (D_a = 14)
α = 0.05
β = 0.05

Steps	**C**	**n**	**S_0**	**S_A**	**$S_0 \leq \alpha$**	**$S_A \geq$ 1-β**
Start with C = 0 and n = midpoint of the full range [0,45]	0	23	0.89147286821702953	0.99999808373317167	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [0,23]	0	12	0.61550387596899303	0.99509317043459966	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [0,12]	0	6	0.35595489781536127	0.90960398081781979	No	No
Neither satisfied, so start again with C = 1 and n = midpoint of the full range [0,45]	1	23	0.51705426356587558	0.99992952396523571	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [0,23]	1	12	0.16899224806201746	0.95387580208521339	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [0,12]	1	6	0.042635658914728709	0.61755530346000809	Yes	No
S_A not satisfied, so take n = midpoint of upper half [6,12]	1	9	0.097251585623678000	0.85262056178852386	No	No
Neither satisfied, so start again with C=2 and n = midpoint of the full range [0,45]	2	23	0.12480620155038495	0.99894911928375341	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [0,23]	2	12	0.015503875968992432	0.81354047651468697	Yes	No
S_A not satisfied, so take n = midpoint of upper half [12,23]	2	18	0.057505285412260757	0.98177734791460636	No	Yes
S_0 not satisfied, so take n = midpoint of lower half [12,18]	2	15	0.032064834390415346	0.93393946867721767	Yes	No
S_A not satisfied, so take n = midpoint of upper half [15,18]	2	17	0.047921071176886168	0.97115147427113779	Yes	Yes
Both conditions satisfied, so now begin search for minimum n within range [15,17]	2	16	0.039464411557435991	0.95570902814568315	Yes	Yes
Both conditions satisfied, and the new range [16,17] has length	**2**	**16**				

1, so the solution is C=2 and n=16						

2.3.6 Wright and Grieve's Bayesian Method for Attributes

Grieve (1994) developed an equation (his Equation 2.5) that can be used to compute the number of grid units, n, that should be selected from the total set of N equal-sized grid units in the target population and found to have contamination less than the AL (or to be non-defective) in order to be 100(1- ε) percent confident that all N grid units are less than the AL or are non-defective. Grieve's equation, which is based on the methods in Wright (1992), is coded into the VSP software.

For the VSP building module, grid units must be square and of a size such that an appropriate sample (e.g., a swipe or swab sample) or in-situ measurement can be made for each grid unit. If the VSP user draws or loads a map of the room or rooms of concern into VSP, then VSP asks the user to specify the grid size in meters, feet or inches, whereupon VSP computes N. If a map is not drawn or loaded, then VSP asks the user to specify the grid unit size and N.

Note that the Wright and Grieve method is also used in the VSP UXO module. However, in that case a grid unit is a swath (a long, narrow rectangular transect) along which a geophysical detector is moved to look for anomalies that may indicate the presence of UXO.

Wright and Grieve's method is "Bayesian" because it requires that the stakeholders provide a quantitative measure of their *belief* that the target population has grid units that contain contamination greater than the AL (or are defective). This belief should be based on all information and data collected about the target population and the conceptual site model developed for the population. This "belief" is quantified by choosing a specific Beta probability distribution for the fraction, f, of the N grid units that are contaminated above the AL. In other words, there is uncertainty about the fraction of the grid units that are "defective," but there is agreement that the probability that f takes on various values can be modeled by a specific Beta distribution, the shape of which is determined by the value of the two parameters of the distribution: a and b. The expected (true average) value, δ, of f for a Beta distribution with parameter values a and b is $\delta = a / (a + b)$. The Beta distribution is described in many books, including Rothschild and Logothetis (1986, pages 50-51), Patil et al (1976) and Johnson and Kotz (1970).

The VSP user can choose among seven possible Beta distributions. These distributions are listed in Table 4 and are illustrated in Figures 1 through 6. The shape of each distribution and the expected value, δ, for each distribution is determined by the values of the two parameters, a and b.

Table 4. The Seven Beta Distributions Available for Selection in VSP to Model the Uncertainty in the Fraction of Grid Units that are Contaminated above the Action Level or are Defective in Some Other Manner.

Parameter Values of the Seven Beta Distributions in VSP	**Expected Value, δ,* of the Fraction, f, of the Grid Units that are Defective**	**English Characterization of the Beta Distribution used in the VSP Software**
1. a = 1, b = 999	0.001	Extremely low fraction
2. a = 1, b = 99	0.01	Very low fraction
3. a = 1, b = 9	0.1	Low fraction
4. a = 1, b = 1	0.5	All fractions equally likely
5. a = 9, b = 1	0.9	High fraction
6. a = 99, b = 1	0.99	Very high fraction
7. a= 999, b = 1	0.999	Extremely high fraction

* $\delta = a/(a+b)$

The VSP user selects one of the seven distributions and VSP computes n using the following equation (derived from Equation 2.5 in Grieve 1994):

$$n \geq N - (N+b)\left\{1 - (1-\varepsilon)^{(1-\delta)/(b\delta)}\right\},$$

where N, a, b, δ and 1-ε have been defined above. If all of the n randomly selected grid units have contamination measurements less than the AL (or are non-defective), then one can state with 100(1-ε) percent confidence that none of the remaining N-n grid units are contaminated above the AL. As is the case for Schilling's method in Section 2.3.5, it is assumed that all grid units are equal (or approximately) equal in size.

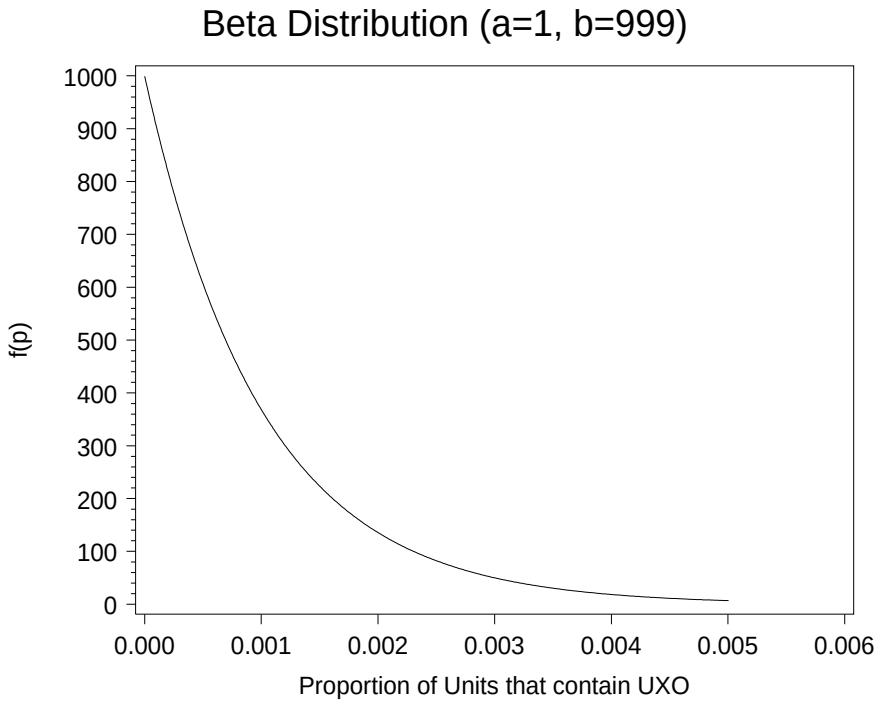

Figure 1. Beta Distribution with Parameters a = 1, b = 999 and Expected Value $\delta = 0.001$

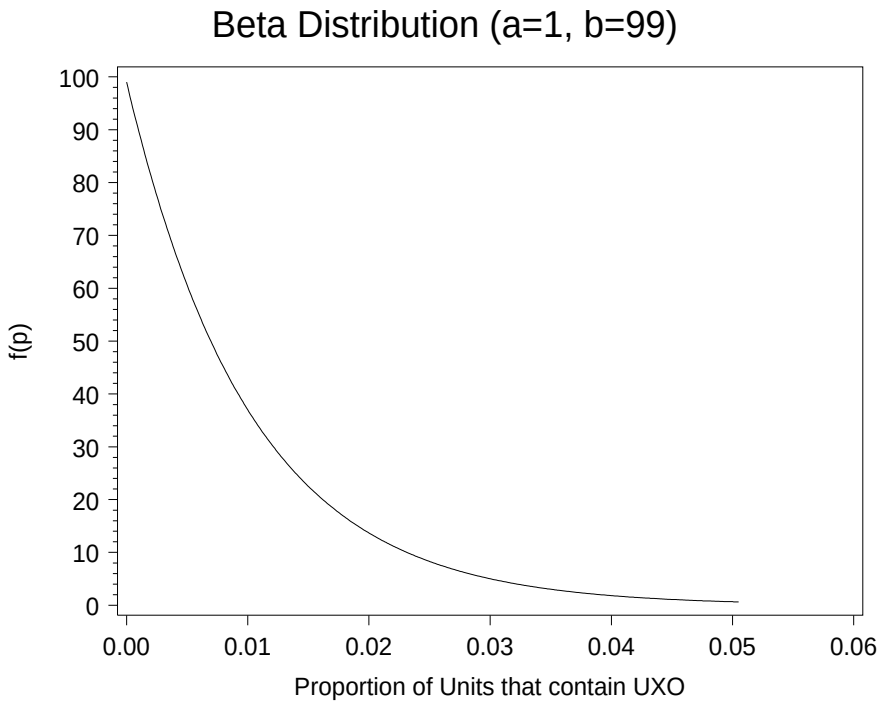

Figure 2. Beta Distribution with Parameters a = 1, b = 99 and Expected Value $\delta = 0.01$

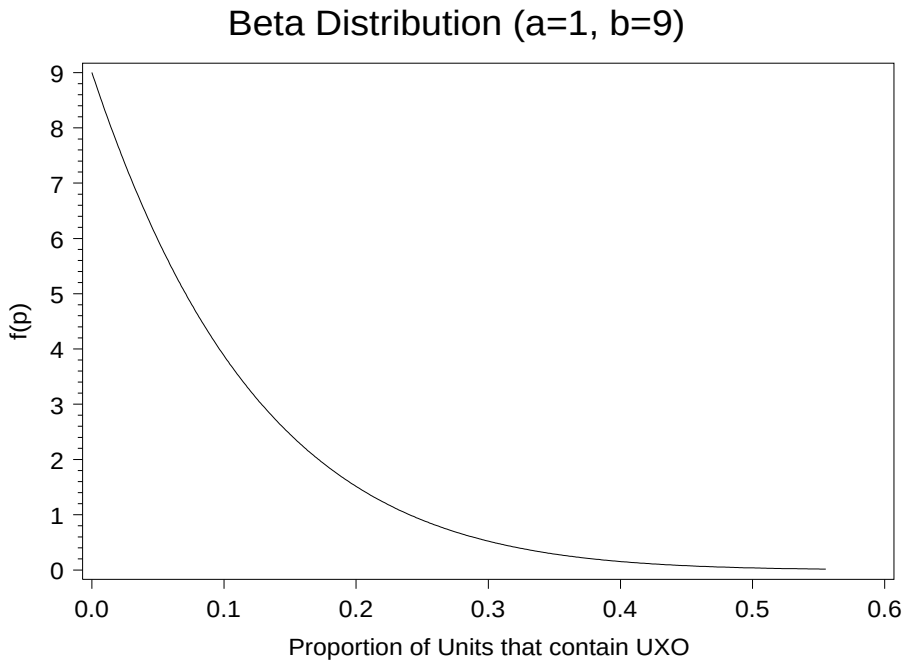

Figure 3. Beta Distribution with Parameters a = 1, b = 9 and Expected Value $\delta = 0.1$

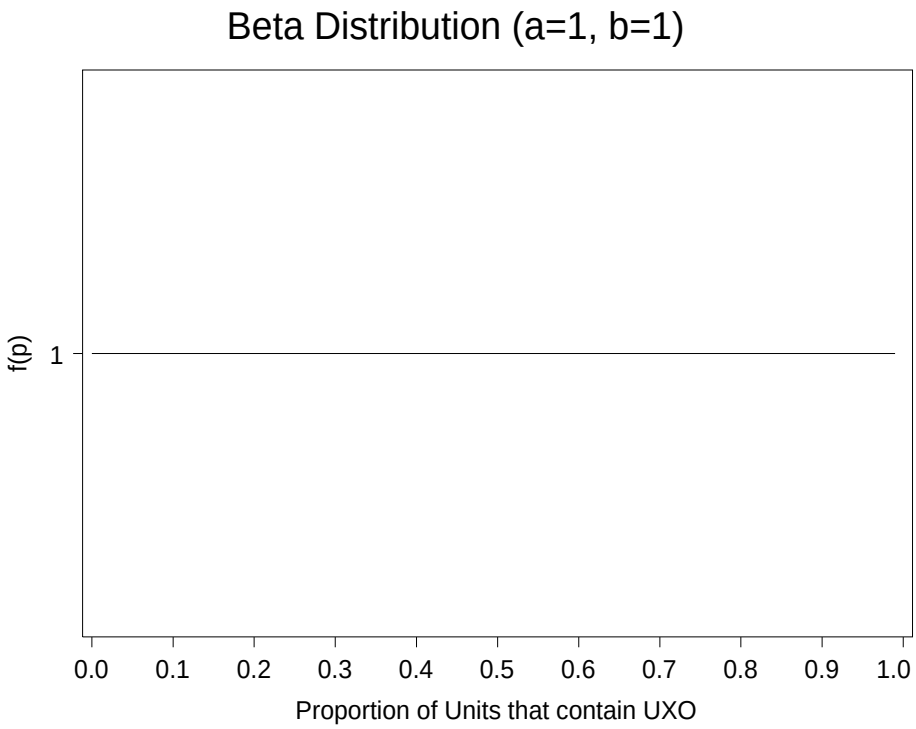

Figure 4. Beta Distribution with Parameters a = 1, b = 1 and Expected Value $\delta = 0.5$

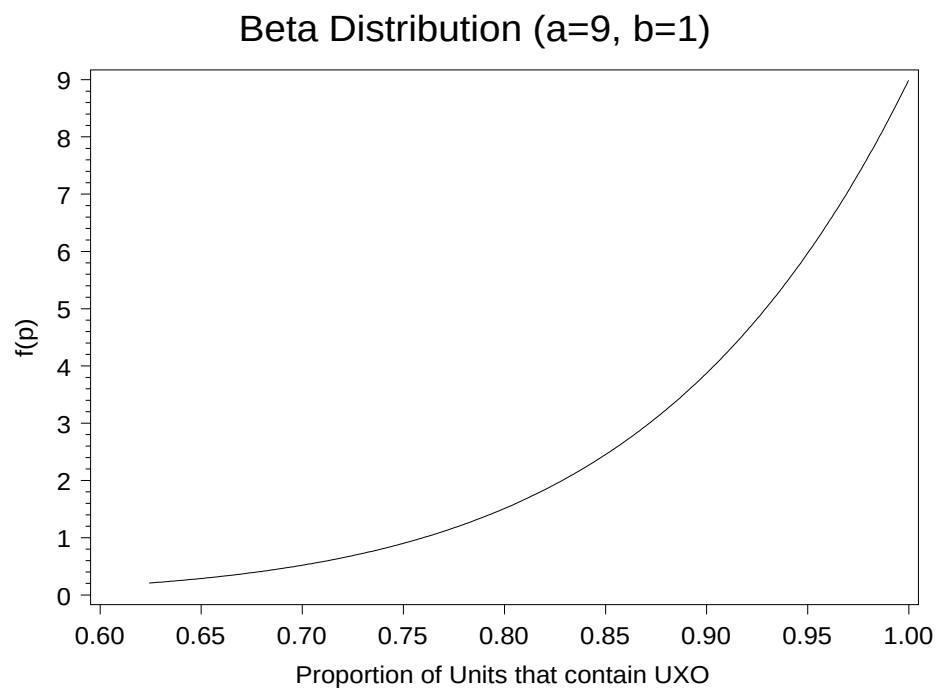

Figure 5. Beta Distribution with Parameters a = 9, b = 1 and Expected Value $\delta = 0.9$

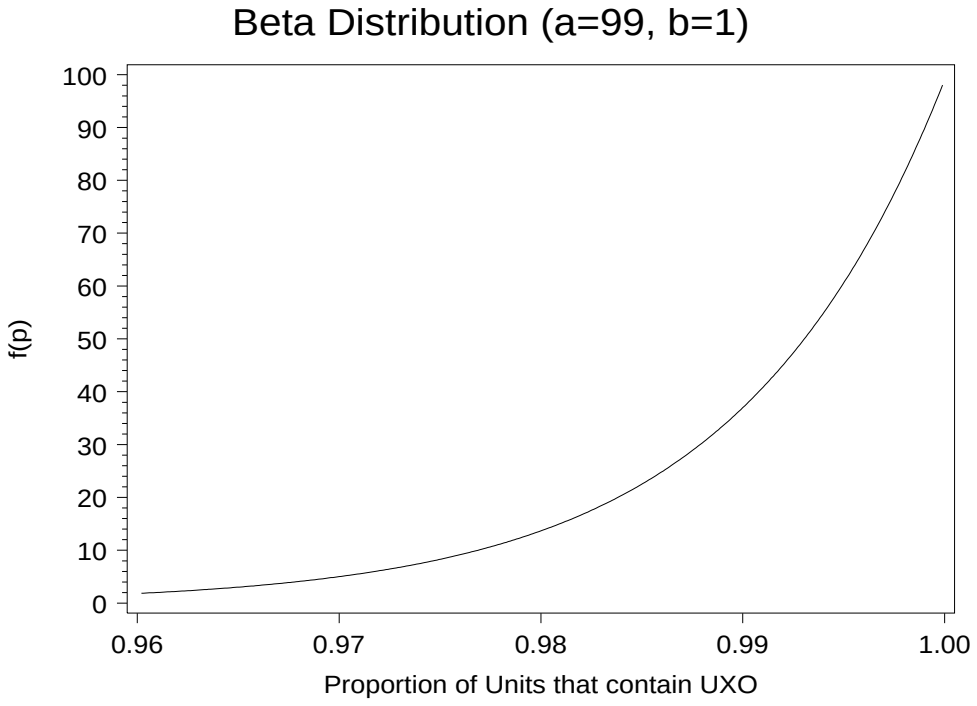

Figure 6. Beta Distribution with Parameters a = 99, b = 1 and Expected Value $\delta = 0.99$

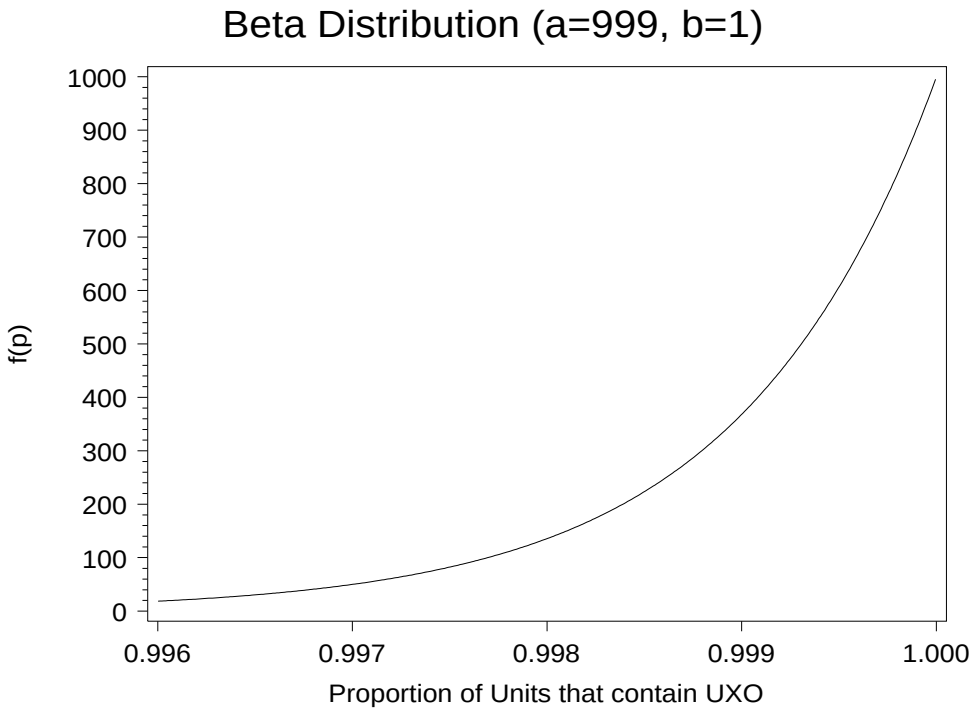

Figure 7. Beta Distribution with Parameters a = 999, b = 1 and Expected Value $\delta = 0.999$

2.4 Data Quality Assessment (DQA)

Data quality assessment (DQA) is the scientific and statistical evaluation of data to determine if data obtained are of the right type, quality, and quantity to support their intended use (EPA 2000b). Although VSP has been designed to determine the number and location of samples needed to make statistically defensible decontamination decisions for rooms in buildings, VSP also has some DQA capability. Once the required n data are obtained, they can be entered into VSP using the Data Analysis tab on the dialog box. Then VSP will compute summary (descriptive) statistics and statistical tests and graphs to help assess if the data are normally distributed. The normality assumption is needed for the statistical methods in VSP discussed in Sections 2.1.1, 2.1.2, 2.1.3, 2.1.4, and 2.3.2. The DQAs conducted in VSP are documented in Sections 2.4.1, 2.4.2 and 2.4.3, below.

2.4.1 Summary (Descriptive) Statistics

The summary statistics computed by VSP are as follows:

n the number of measurements in a data set

Min the minimum of the n data values

Max the maximum of the n data values

Mean the arithmetic mean of the n data values:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Median — the 50th percentile of the data set, i.e., the value above which and below which half the data lay. The sample median is computed from the ordered data, which are denoted by $x_{[1]} \le x_{[2]} \le \cdots \le x_{[n]}$, as follows:

Median = $x_{[(n+1)/2]}$ if n is an odd number

= $\frac{1}{2}\left(x_{[n/2]} + x_{[(n+2)/2]}\right)$ if n is an even number

Range — the maximum data value minus the minimum data value, i.e.,

$$\text{Range} = x_{[n]} - x_{[1]}$$

P^{th} Percentile — the value below which P% of the n data lie and above which (1 – P)% of the n data lay, where $0 < P < 100$.

VSP computes the P^{th} percentile by first computing $k = (P/100)(n+1)$.
If k is an integer, the P^{th} percentile is $x_{[k]}$, which is the k^{th} largest of the n data values.
For example, to compute the 50^{th} percentile of a data set of n = 9 measurements, we have P = 50 and $k = (P/100)(n+1) = 0.50(10) = 5$, which is an integer. Hence, the 50^{th} percentile (the median) is the 5^{th} largest datum, $x_{[5]}$.

If k is not an integer, the P^{th} percentile is obtained by linear interpolation between the two closest ordered data values. For example, suppose n = 11 and P = 70. Then $k = (P/100)(n+1) = 0.70(12) = 8.4$. The 70^{th} percentile is computed using linear interpolation between the 8^{th} and 9^{th} largest data values, i.e., between $x_{[8]}$ and $x_{[9]}$.

Interquartile Range (IQR) — the 75^{th} percentile of the data set minus the 25^{th} percentile of the data set

Standard Deviation (s) — the spread of the data, s, computed as:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Variance — the square of the standard deviation:

$$\text{Variance} = s^2$$

Standard Error — the standard deviation of the estimated mean computed as

$$SE = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

Skewness Coefficient — a measure of the symmetry of the data set computed as

$$\text{SKEW} = \frac{\frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^3}{s^3},$$

where $\bar{x}$ and s are computed as given previously.

Figure 8 shows the summary (descriptive) statistics computed and displayed by VSP for a set of 150 data generated from a normal distribution.

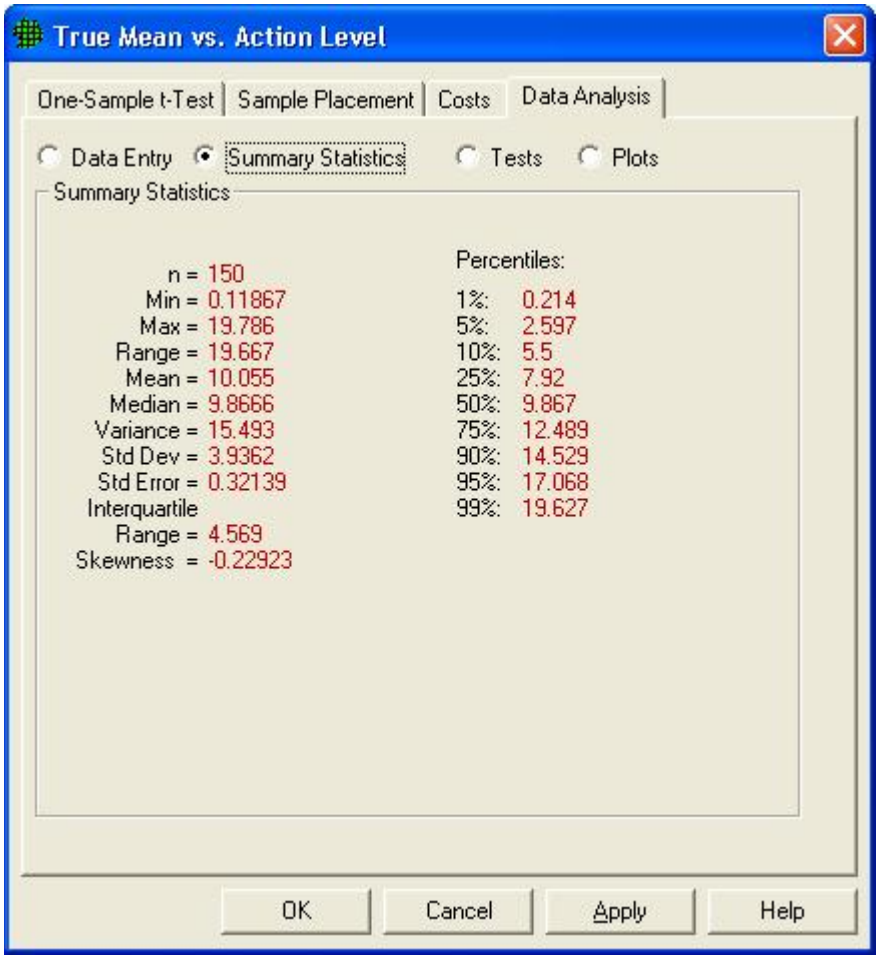

Figure 8. VSP Summary (Descriptive) Statistics Computed using 150 Data from a Normal Distribution

2.4.2 Testing Whether Data are Normally Distributed

Shapiro-Wilk Test

The Shapiro-Wilk test, which is denoted by W, tests whether the n data are normally distributed. VSP computes this test when the number of data does not exceed 50. The null and alternative hypotheses are

H_o: The data are normally distributed
H_a: The data are not normally distributed

The test is conducted as follows (from Gilbert 1987, pp. 158-160):

1. Compute
$$d = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$$
2. Order the n data from smallest to largest to obtain the sample order statistics $x_{[1]} \le x_{[2]} \le \cdots \le x_{[n]}$
3. Compute k, where
$$k = \frac{n}{2} \quad \text{if n is even}$$
$$k = \frac{n-1}{2} \quad \text{if n is odd}$$
4. Find coefficients $a_1, a_2, a_3, \ldots, a_k$ from Table A6 in Gilbert (1987, pages 259-260).
5. Compute
$$W = \frac{1}{d}\left[\sum_{i=1}^{k} a_i (x_{[n-i+1]} - x_i\right]^2$$
6. Reject H_o and accept H_a at the α significance level if W is less than the percentile, $W_{\alpha,n}$, of the W statistic given in Table A7 of Gilbert (1987, page 261)

Figure 9 shows the VSP results for the Shapiro-Wilk test for the 30 data used in Figure 8. For these data, null hypothesis is not rejected, i.e., the test indicates that there is insufficient evidence to conclude that the data are not normally distributed.

Note that Figure 9 also shows the 95% upper confidence limit on the mean that is computed assuming the data are normally distributed. Since the Shapiro-Wilk test indicated that the assumption of normally distributed data cannot be rejected, VSP automatically recommends on the screen shot in Figure 9 that this UCL be compared to the action level (AL) for making decontamination decisions. If the Shapiro-Wilk test had rejected the null hypothesis, then VSP would have automatically recommended that the UCL computed using the nonparametric (distribution-free) Chebyshev method (shown in Figure 9) be compared to the AL.

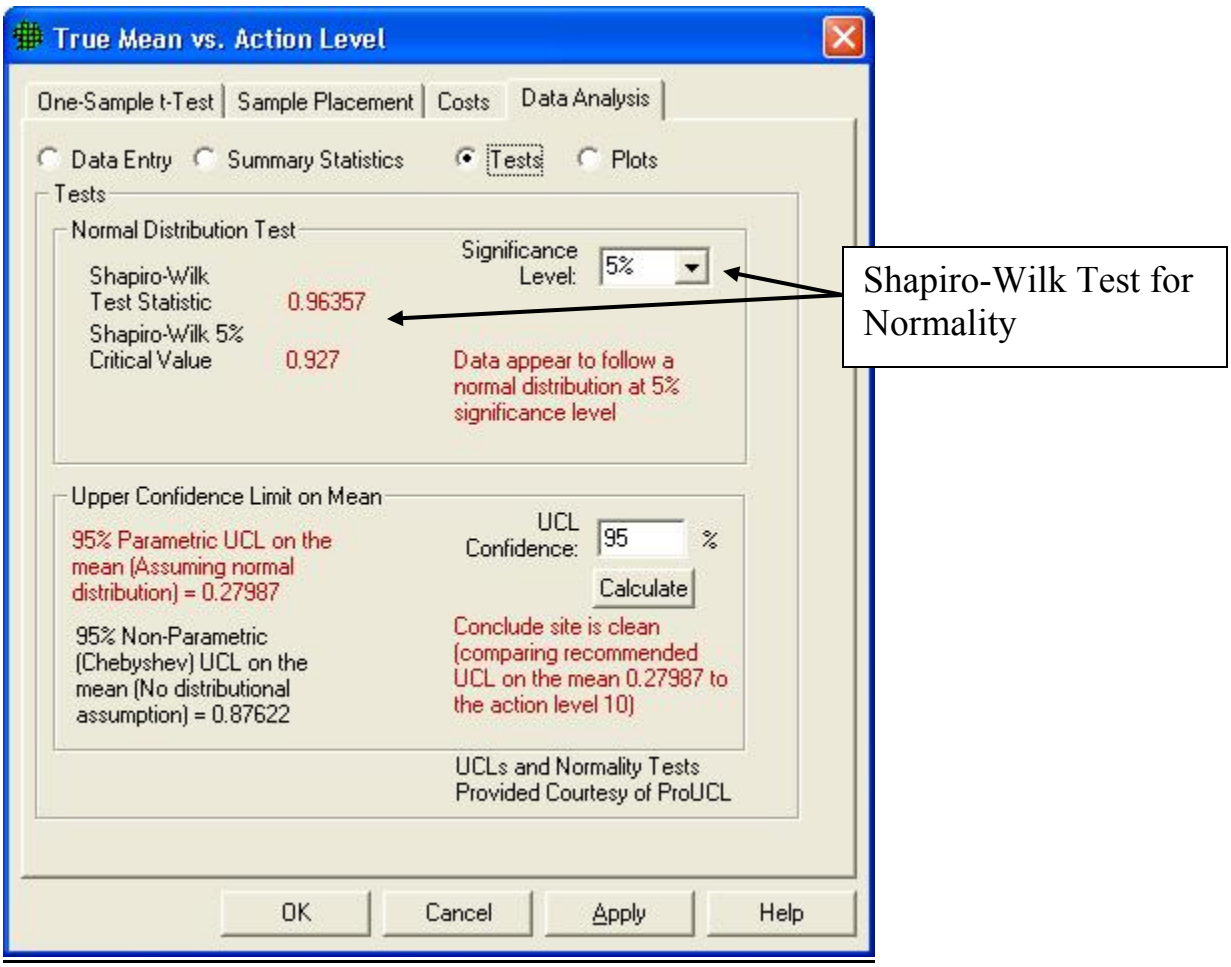

Figure 9. The VSP Results for the Shapiro-Wilk Test for Normality using 30 Data Generated from a Normal Distribution

Lilliefors Test

The Lilliefors test (Conover 1999, pages 443-447) tests the following null and alternative hypotheses for data sets for which $50 < n \leq 1000$:

H_o: The data are normally distributed
H_a: The data are not normally distributed

The test is conducted as follows:

1. Compute the mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and the standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

2. Compute the "normalized" sample values Z_i:

$$Z_i = \frac{x_i - \bar{x}}{s} \qquad i = 1, 2, \ldots, n \tag{28}$$

3. Compute the Lilliefors test statistic T as follows:

$$T = \sup \left| F^*(x) - S(x) \right|,$$

where

T is the supremum over all x of the absolute value of the difference $F^*(x) - S(x)$, and

$F^*(x)$ is the cumulative distribution function of a normal distribution with mean zero and standard deviation one, and

$S(x)$ is the empirical distribution function of the values of Z_i, which are computed using Equation (28) above.

4. Reject H_o and accept H_a at the α significance level if T exceeds the critical value for the test, which can be obtained from Table A14 in Conover (1999, page 548).

Figure 10 shows the VSP results for the Lilliefors test for a set of 150 normally distributed data. For these data, the null hypothesis is not rejected, i.e., the test indicates that there is insufficient evidence to conclude that the data are not normally distributed. Figure 10 also provides the 95% upper confidence limit on the mean that is computed assuming the data are normally distributed, as well as when the data are assumed not to be normally distributed.

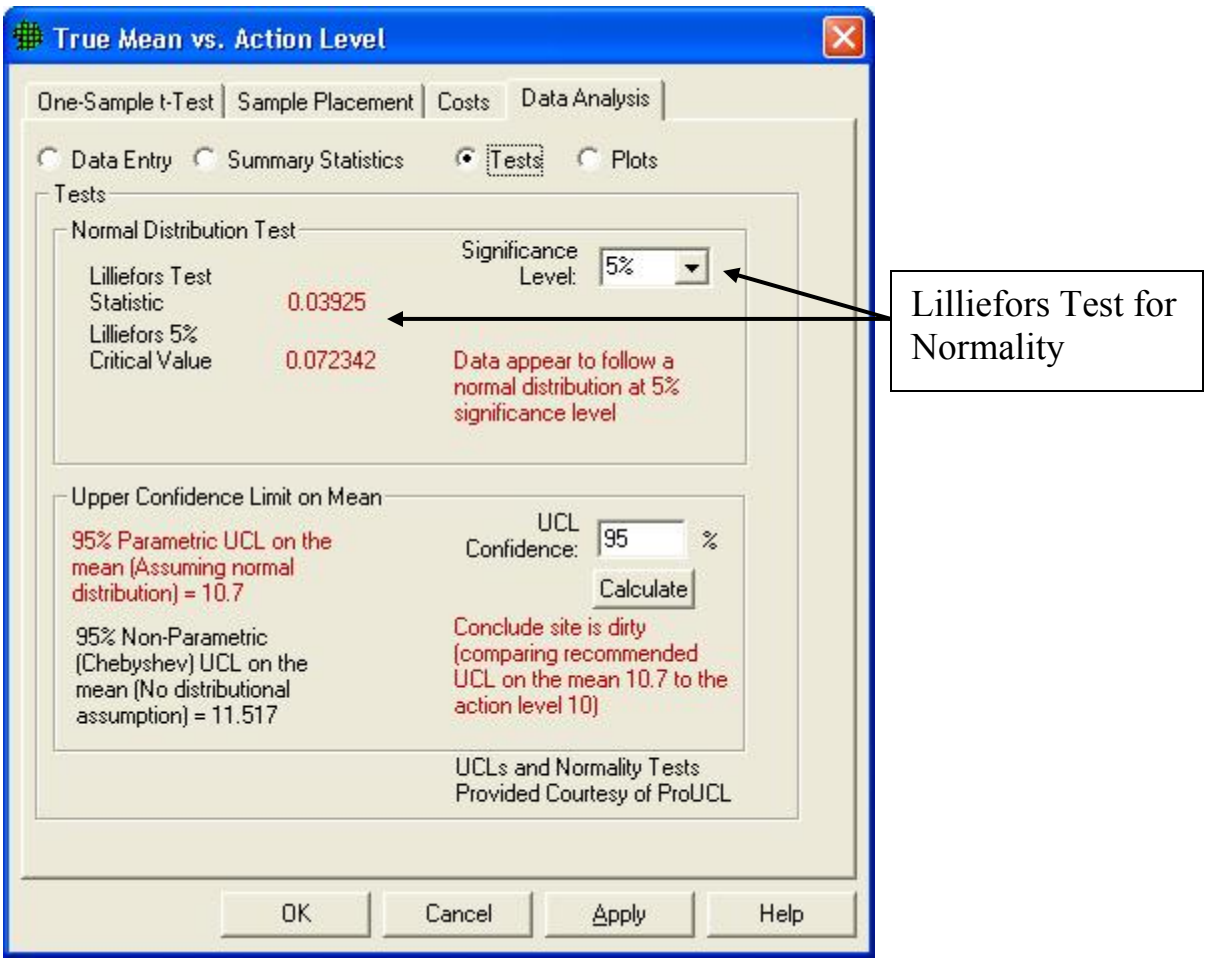

Figure 10. VSP Results for the Lilliefors Test for Normality using 150 Data Generated from a Normal Distribution

2.4.3 Graphical Tools for Assessing Whether Data are Normally Distributed

VSP provides three graphical summaries of the data that provide visual assessments of the normality assumption:

- Histograms
- Box-and-whisker plots
- Quantile-quantile plots (also called "probability plots")

Examples of these graphs are shown in Figures 11, 12, and 13, which were produced by VSP. These graphs display the 150 measurements used in Figure 10.

All three plots are consistent with the conclusion of the Lilliefors test, that is, they provide no strong evidence that the data are not normally distributed. The shape of the histogram in Figure 11 is not inconsistent with bell shape of the normal distribution. The box-and-whisker plot in Figure 12 indicates that the median is approximately midway between the 25th and 75th percentiles and that the data are scattered to the same extent in both tails of the distribution, which suggests normality. Finally, the Q-Q plot in Figure 13 shows that there is a strong linear relationship between the observed data and the data that are theoretically expected if the data are normally distributed.

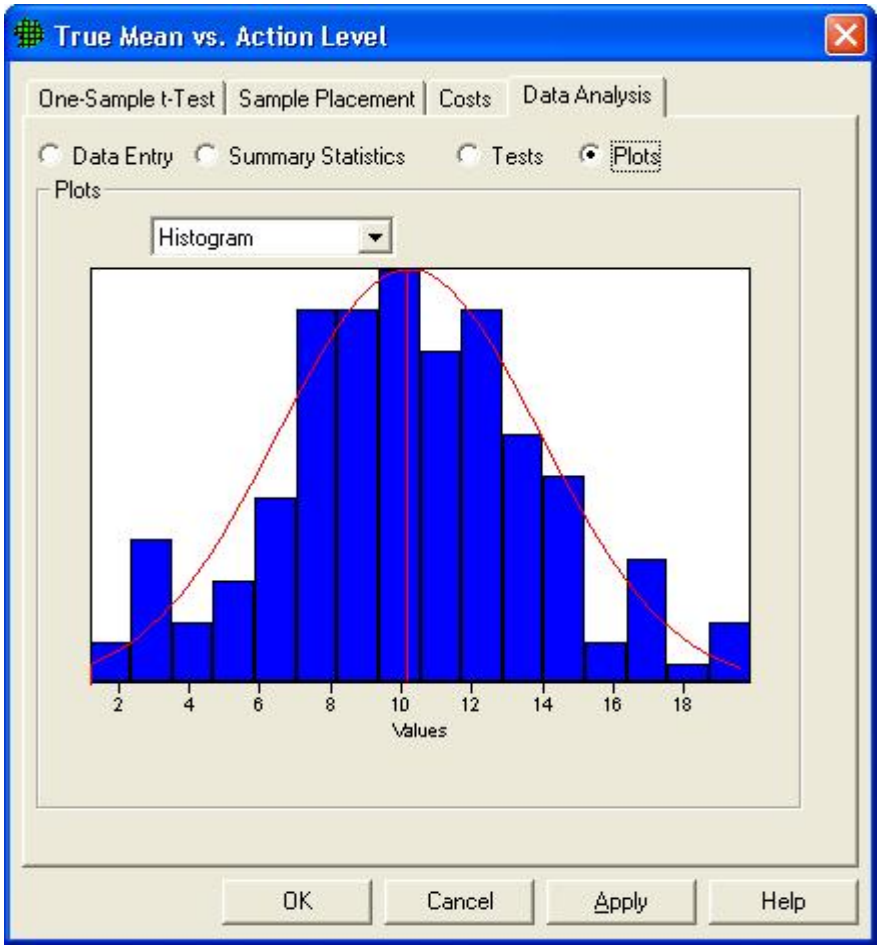

Figure 11. Histogram of n = 150 Data Compared to the Normal Distribution

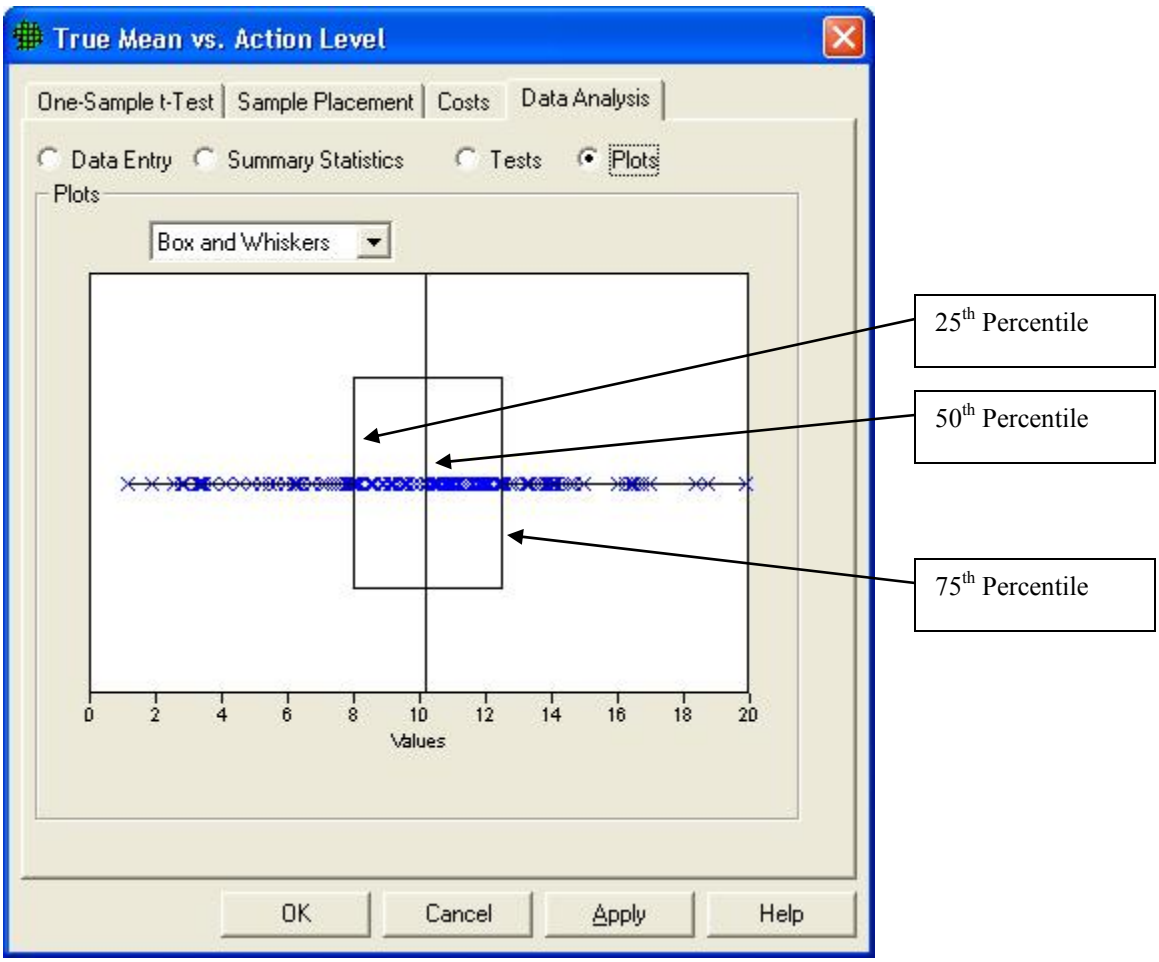

Figure 12. Box-and-Whisker Plot of n = 150 Data

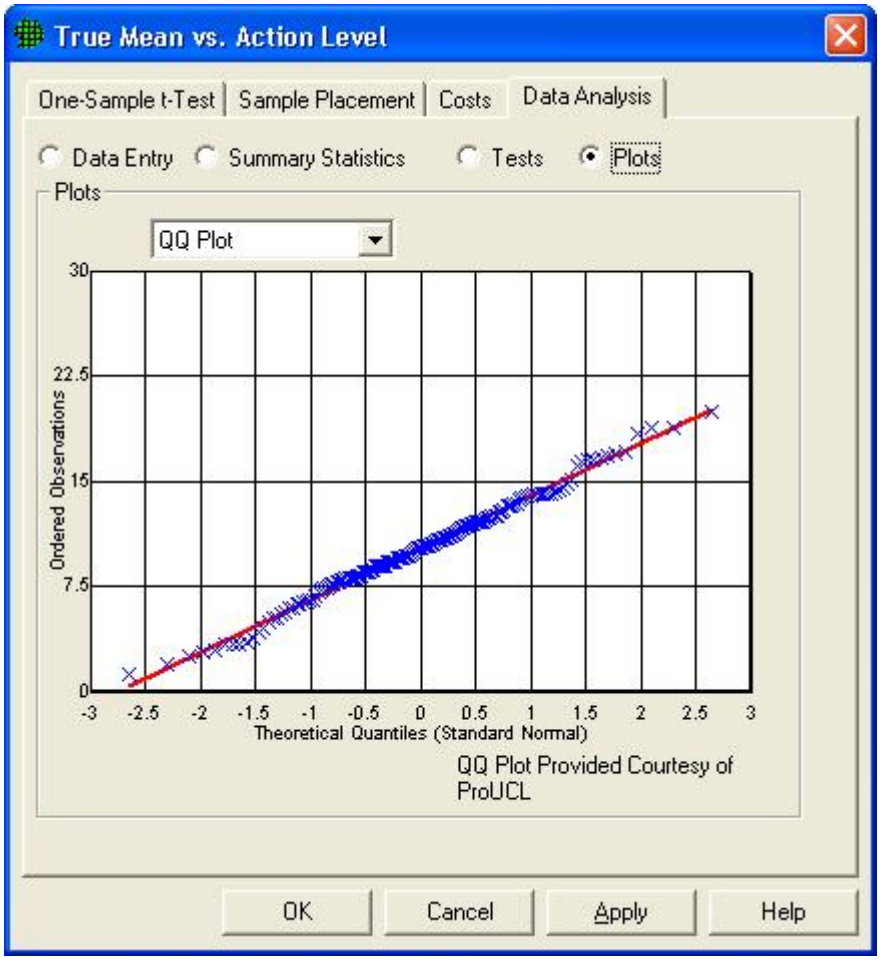

Figure 13. Quantile-Quantile Plot of n = 150 Data

2.5 One-Sided Upper Confidence Limits on the Mean

VSP also automatically computes one-sided upper confidence limits (UCL) on the mean for any data set entered into the VSP building module. Two methods for computing the UCL are provided. One method assumes the data are normally distributed, and one does not. The formulas in VSP for these two methods of computing UCLs are identical to those used in the ProUCL (2004) computer code.

The equation used to compute the UCL when the data are assumed to be normally distributed is the usual formula (Gilbert 1987, Equation 11.6, page 139):

$$UCL = \bar{x} + t_{1-\alpha,n-1}\frac{s}{\sqrt{n}},$$

where

$\bar{x}$ = mean of the n measurements
s = standard deviation of the n measurements
$t_{1-\alpha,n-1}$ = 100(1-α) percentile of the t distribution with n-1 degrees of freedom

n = number of measurements

VSP uses the following Chebyshev formula to compute the UCL when no assumption is made about the distribution of the data:

$$UCL = \bar{x} + \sqrt{\frac{1}{\alpha} - 1}\ \frac{s}{\sqrt{n}} \quad ,$$

where $\bar{x}$, s, α and n have been defined above.

Both of these methods are also used in the EPA-developed software ProUCL (2004, pages A-24 and A-32).

3.0 Verification of VSP Computations and Outputs

3.1 Previously Developed Statistical Methods in VSP that are Applicable to the VSP Buildings Module

Gilbert et al. (2002) documents the computations conducted to verify that VSP is correctly and accurately computing the number of samples needed for the statistical methods described in Sections 2.1, 2.2, and 2.3.1 of this report. PNNL staff compared VSP computations with hand calculations and the results of PNNL-prepared (S-PLUS® software (http://www.insightful.com/products/splus/default.asp) and SAS® (Statistical Analysis System) software (http://www.SAS.com). Exact or sufficiently good agreement was obtained in all cases considered as described in Sections 3.2.1, 3.2.3, 3.2.4, and 3.2.5 in Gilbert et al. 2002. In addition, computations conducted independently by the Research Triangle Institute (RTI) for the "one-sample" t test (as well as other VSP methods not discussed in this report) agreed exactly with those of VSP as described in Section 3.2.2 of Gilbert et al (2002). Davidson (2001) verified the accuracy of sample-size equations in an early version of VSP.

PNNL also checked the accuracy of the four graphical displays in VSP outputs:

- the "Map View," which shows on the map of the study site the sampling locations determined by VSP,
- the "Graph View," which shows the Decision Performance Goal Diagram, a graph of the quantitative DQO parameters specified by the VSP user,
- the "Report View," which documents the sampling objectives, number and location of samples, underlying assumptions, costs, and other information, and
- the "Coordinate View," which lists the geographical coordinates of the sampling locations.

Three minor display problems were uncovered and corrected.

Finally, extensive testing by PNNL verified that the following non-statistical portions of VSP were operating correctly in Sections 4.1 - 4.7 of Gilbert et al. (2002):

- installation success for various computer platforms
- file import, export, and removal of sampling locations
- drawing functions
- correspondence between dialog box values and values in view windows
- documentation of algorithms to determine sampling locations
- pseudo-random and quasi-random number generators
- the largest unsampled spot location algorithm

3.2 New Statistical Methods Developed for the VSP Buildings Module

3.2.1 One-Sided Upper Tolerance Limits for Normally Distributed Data

Verifying Computations of n

As indicated in Section 2.3.2, a one-sided upper tolerance limit on the P^{th} percentile of a population is a value, denoted here by $UTL_{P,\alpha}$, computed using the n measurements such that at least P% of the population of measurements is less than $UTL_{P,\alpha}$ with 100(1-α)% confidence, where 0 < P < 100. For

example, if P = 95 and α = 0.05, then at least 95% of the population is less than the computed value $UTL_{95,0.05}$ with 95% confidence.

VSP uses the iterative method in Section 2.3.2, Equation (22), to determine the number of samples or measurements, n, needed to compute $UTL_{P,\alpha}$. The VSP user inputs the parameters P, α, β, AL, Δ, and s. Then VSP computes n. The values of n computed by VSP were verified as being accurate by using VSP to compute n for 24 combinations of input parameters used in Lyles and Kupper (1996, Table II). The values of n computed by VSP and by Lyles and Kupper (1996) were identical for all 24 cases. The inputs and computed values of n are given in Table 5. Tables 6 and 7 show the iterations VSP conducted to compute n = 58 and n = 20 in Table 5.

Table 5. Number of Samples or Measurements, n, Computed by both VSP and Lyles and Kupper (1996) for Computing a One-Sided Upper Tolerance Limit on the 95th Percentile of a Normal Distribution*

$\frac{UBGR}{LBGR}$	**Ln(UBGR)-Ln(LBGR) ****	**Variance of the Natural Logarithms of the Measurements (σ_y^2)**					
		0.50	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**
1.5	0.40546	58	107	154	202	249	295
2.0	0.69315	24	42	59	76	93	109
2.5	0.91629	16	27	37	47	57	67
3.0	1.09861	13	20	28	35	42	49

* The values of n in this table were obtained for the following VSP DQO inputs:
P = 95, α = 0.05, β = 0.20, action level (AL) = 3, Ln (3) = 1.0986 = UBGR.
** Width of the gray region (in natural logarithms) of the Decision Performance Goal Diagram.

Table 6. Iterations Computed by VSP in Determining the Number of Samples Needed for a Normal Distribution One-Sided Upper Tolerance Limit when P = 95, α = 0.05, β = 0.20, Action Level = 1.0986, Width of Gray Region = 0.405, and Variance = 0.50

n	$t_{n-1,-Z_P\sqrt{n},\alpha}$ *	$t_{n-1,-Z_{1-\theta}\sqrt{n},1-\beta}$ **	$t_{n-1,-Z_P\sqrt{n},\alpha} - t_{n-1,Z_{1-\theta}\sqrt{n},1-\beta}$ ***
10000	-437.0290	-596.286	159.258
5001	-212.7520	-293.9670	81.21500
2501	-100.3410	-142.6840	42.34280
1251	-60.7745	-76.9311	16.15660
626	-43.7835	-53.9880	10.20450
314	-31.8240	-37.8151	5.99113
158	-23.4242	-26.4179	2.99365
80	-17.5704	-18.41150	0.84107
41	-13.5638	-12.8207	-0.74314
61	-15.7653	-15.0133	0.14800
51	-14.7119	-14.4421	-0.26980
56	-15.2490	-15.19390	-0.05518
59	-15.5611	-15.6291	0.06798
58	-15.4579	-15.48530	0.02736
57	-15.3539	-15.3402	-0.01369
58	-15.4579	-15.48730	0.02736

* Equation (23)
** Equation (24)
*** Equation (22)

Table 7. Iterations Computed by VSP in Determining the Number of Samples Needed for a Normal Distribution One-Sided Upper Tolerance Limit when P = 95, $\alpha = 0.05$, $\beta = 0.20$, Action Level = 1.0986, Width of Gray Region = 1.09861 and Variance = 1.0

n	$t_{n-1,-Z_P\sqrt{n},\alpha}$ *	$t_{n-1,-Z_{1-\theta}\sqrt{n},1-\beta}$ **	$t_{n-1,-Z_P\sqrt{n},\alpha} - t_{n-1,Z_{1-\theta}\sqrt{n},1-\beta}$ ***
10000	-437.0290	-741.044	304.016
5001	-212.7520	-367.19	154.438
2501	-100.3410	-180.178	79.8368
1251	-60.7745	-95.2448	34.4703
626	-43.7835	-66.8698	23.0863
314	-31.8240	-46.8679	15.0439
158	-23.4242	-32.7722	9.348
80	-17.5704	-22.8703	5.29983
41	-13.5638	-15.9557	2.39186
21	-10.8672	-11.0335	0.166307
11	-9.33628	-7.64787	-1.6884
16	-10.0946	-9.46975	-0.62488
19	-10.5618	-10.4322	-0.12955
20	-10.7152	-10.7365	0.0212574

* Equation (23)
** Equation (24)
*** Equation (22)

As explained in Section 2.3.2, the computation of normal distribution upper tolerance limits, as well as the number of samples required, requires using percentiles from the non-central t distribution. These percentiles used in the VSP software were verified as being accurate by comparison with percentiles computed using non-central t values computed using the SAS® (Statistical Analysis System) software (http://www.SAS.com). One hundred comparisons were made for various values of the non-centrality parameter and degrees of freedom. For all cases the two methods agreed to at least 5 digits.

Note that Lyles and Kupper (1996) assumed that the data were lognormally distributed, rather than normally distributed as is assumed here in this section. However, the natural logarithms of lognormally distributed data are normally distributed. Hence, to make the Lyles and Kupper (1996) computations of n applicable to the normal distribution, the natural logarithms of the standard deviation, s, and of the AL were used in VSP. Note that VSP can be used to determine n for tolerance limits when the data are either normally or lognormally distributed. For the lognormal case, the VSP user must be sure to input the natural logarithms of the standard deviation and AL

Verifying VSP Computations of Percentiles and One-Sided Upper Tolerance Limits

Hand calculations were conducted to verify that VSP was correctly computing estimates of percentiles and one-sided upper tolerance limits. These comparisons were conducted for the three sets of input parameters shown in Table 8.0. Both VSP and hand calculations of n were exactly the same, and computations of percentiles and UTLs agreed to at least 4 decimal places.

Table 8. Comparing VSP and Hand Calculations for Computing One-Sided Upper Tolerance Limits for the Normal Distribution

VSP Inputs						VSP and Hand Calculations		
Percentile of Interest	**α**	**β**	**Δ**	**s**	**AL**	**n**	**Estimated Percentile***	**UTL****
0.90	0.05	0.10	2	1	10	9	12.49	11.28
0.99	0.05	0.20	2	1.4	5	20	7.557	6.521
0.70	0.01	0.05	5	2	30	8	29.396	33.579

* Computed using Equation (17)
** Computed using Equation (21)

3.2.2 Nonparametric (Distribution-Free) One-Sided Upper Tolerance Limits

As indicated in Section 2.3.4, VSP computes the number of samples or measurements needed to determine the 100(1-α) % nonparametric upper tolerance limit on the P^{th} percentile using the following equation from Hahn and Meeker (1991, page 169):

$$n = \frac{\ln(\alpha)}{\ln(P)}$$

Numerous hand calculations verified that VSP was accurately computing n using this equation for various combinations of α and P.

3.2.3 Nonparametric (Distribution-Free) Compliance Sampling for Attributes

The acceptance sampling methodology in the VSP buildings module is identical to the acceptance sampling methodology used in the UXO Module of VSP except that grid units in the UXO module are long, narrow rectangles (transects) suitable for use with geophysical detectors, whereas units in the building module are relatively small square units of a size specified by the VSP user. Hence, the computations reported in Gilbert et al. (2003, Section 7.2, Table 7.2, pages 29-30) to verify that VSP is correctly computing n for the UXO module also apply to the buildings module. There was perfect agreement between the VSP and hand calculations of n (the number of grid units that need to be selected and measured) for 11 combinations of VSP input parameters used in Gilbert et al. (2003).

3.2.4 Nonparametric (Distribution-Free) Acceptance Sampling for Attributes

As discussed in Section 2.3.5, acceptance sampling can be used to statistically test whether the fraction of a room (or suite of rooms) that is contaminated above a level (AL) is less than a prescribed upper limit. VSP determines the number (n) and location of grid units (defined below) in the room that must be measured or inspected to make this determination. VSP also computes the integer C. If more than C of the n measured grid units equal or exceed the AL or are defective in some other way as defined by the VSP user, then the required confidence is not achieved and additional investigations may be needed.

Two comparisons (examples) were conducted to verify the accuracy of VSP's computations of n and C. VSP computations were compared with those in Schilling (1982, pages 120-121) and Bowen and Bennett (1987, pages 884-886). Figures 14 and 15 show VSP software results that document the VSP input parameters and the resulting computed values of n and C for these two examples. The results are also summarized in Table 9. The values of n and C computed by VSP were exactly the same as those computed by Bowen and Bennett (1987) and Schilling (1982).

Table 9. Comparing VSP and Scientific Literature Computations of the Number of Samples (n) and the Acceptance Number (C) for Acceptance Sampling

	Acceptance Sampling Input Parameters				VSP Results*	
	Acceptable % of Grid Units Allowed to be Contaminated	**Unacceptable % of Grid Units Allowed to be Contaminated**	**Acceptable False Rejection Rate (Alpha) of the Null Hypothesis**	**Acceptable False Acceptance Rate (Beta) of the Null Hypothesis**	**Number of Grid Units that Must be Sampled, n**	**Accept-ance Number C**
Schilling (1982) Example	10	20	0.24	0.30	10	1
Bowen and Bennett (1987) Example	5	20	0.10	0.10	14	1

* VSP results agree exactly with those of Schilling (1982) and Bowen and Bennett (1987)

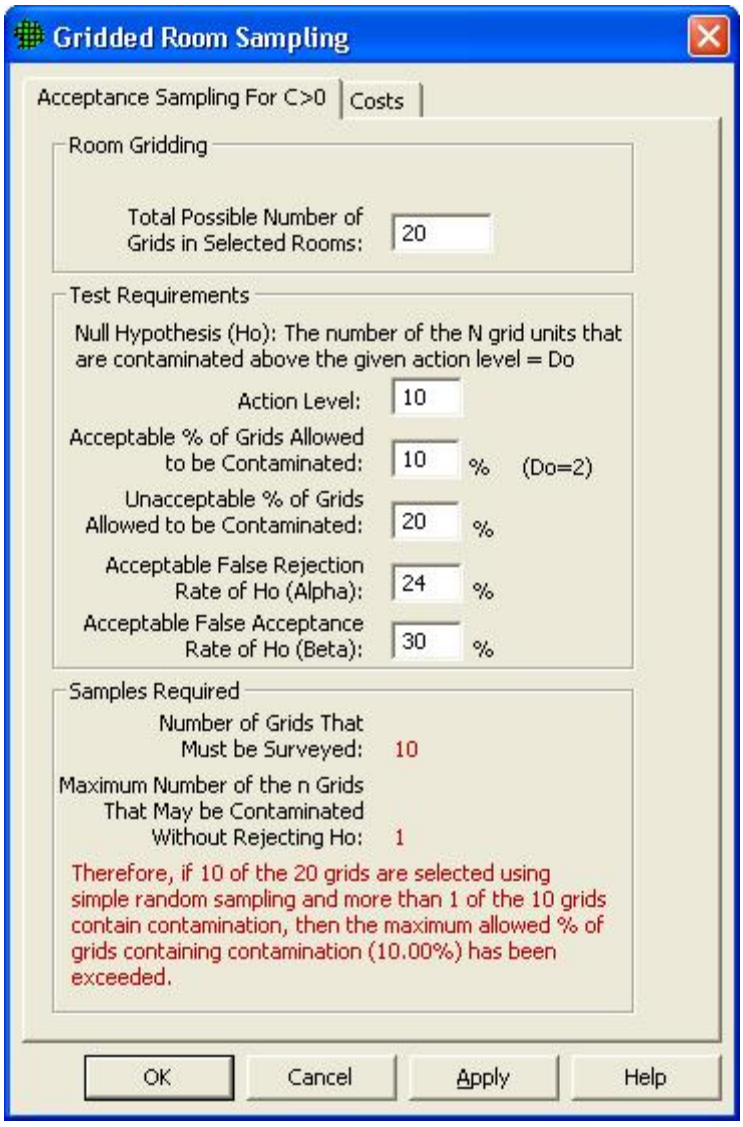

Figure 14. VSP Acceptance Sampling Inputs and Computation of n and C using Input Parameters for the Example in Schilling (1982, Pages 120-121)

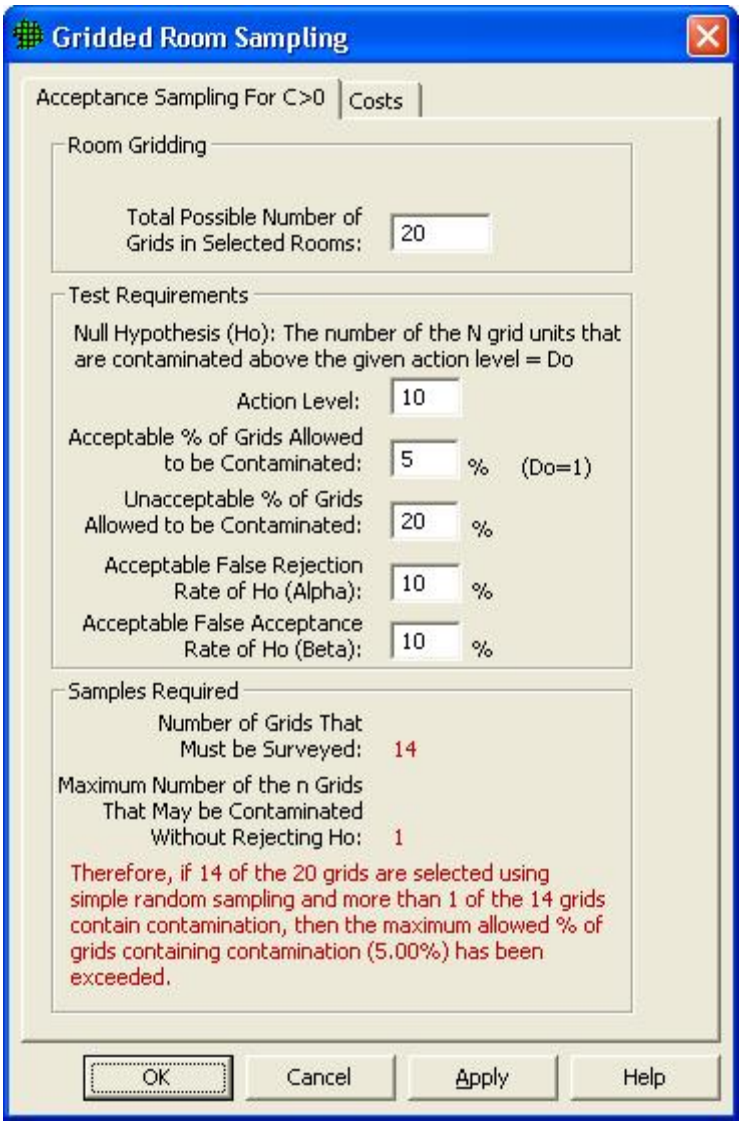

Figure 15. VSP Acceptance Sampling Inputs and Computation of n and C using Input Parameters for the Example in Bowen and Bennett (1987, Pages 884-886)

3.2.5 Wright and Grieve's Bayesian Method for Attributes

The Wright and Grieve method in the VSP buildings module for computing the number of square grid units to sample or measure is identical to the Wright and Grieve method in the VSP UXO module. Hence, the computations reported in Gilbert et al. (2003, Section 7.3, Tables 7.3 – 7.6, pages 30-31) to verify that VSP was correctly computing n for the UXO module also apply to the buildings module. There was perfect agreement between VSP and hand calculations for the 20 UXO cases considered in Gilbert et al. (2003).

3.2.6 Data Quality Assessment (DQA) Methods

The DQA methods available for the buildings module of VSP were described and illustrated in Section 2.4. The accuracy of VSP computations was confirmed using two generated data sets: a subjectively

selected small data set of size n = 10 and a larger data set of size n = 150 generated from a Uniform distribution (i.e., all data values between zero and 1 are equally likely) using the Minitab® software. The n = 10 data are (ordered from smallest to largest): 3, 3, 4, 5, 5, 5, 5, 7, 8, 9. The n = 150 data are shown in Table 10 below.

Table 10. Data Values for the n = 150 Data from a Uniform Distribution that were used to Confirm VSP Calculations

0.393239	0.217942	0.884915	0.591513	0.448162	0.848726	0.449710
0.060060	0.907290	0.306071	0.967525	0.397397	0.306274	0.701034
0.473084	0.961279	0.496642	0.324835	0.097984	0.769982	0.659596
0.368637	0.841605	0.409483	0.546913	0.291464	0.806556	0.096745
0.687879	0.721830	0.544905	0.004014	0.626819	0.569690	0.055153
0.267127	0.723115	0.352003	0.237320	0.755126	0.420712	0.843423
0.068970	0.304536	0.528710	0.500253	0.746839	0.940291	0.773938
0.075600	0.306106	0.797992	0.838047	0.031409	0.314868	0.676918
0.305022	0.699674	0.915306	0.096862	0.238844	0.678509	0.584849
0.654240	0.069360	0.684844	0.718156	0.042515	0.137714	0.360637
0.504661	0.176554	0.269683	0.670299	0.330245	0.639786	0.518187
0.926014	0.123607	0.319425	0.016896	0.691269	0.602390	0.923842
0.475836	0.521439	0.886162	0.213561	0.986175	0.821703	0.088199
0.412209	0.155279	0.607685	0.709771	0.756929	0.623754	0.764255
0.063380	0.597041	0.963442	0.647390	0.259159	0.580860	0.725436
0.384467	0.747730	0.412887	0.789853	0.349645	0.793819	0.831777
0.514551	0.484066	0.464942	0.254986	0.594485	0.294630	0.847482
0.865880	0.364579	0.653673	0.855020	0.112503	0.128002	0.118295
0.446828	0.926365	0.541171	0.772976	0.402897	0.781555	0.134294
0.061969	0.478819	0.407399	0.104149	0.795857	0.503805	0.466339
0.068171	0.963593	0.817505	0.765519	0.657113	0.108781	0.056217
		0.258344	0.933720	0.633266		

Descriptive Statistics

For both the small and large data sets, the VSP and Minitab® (2003) software calculations of the descriptive statistics listed in Table 11 gave the same computed values as displayed in the table. Note that the 1st, 5th, 95th, and 99th percentiles in Table 11 were calculated by hand because Minitab does not compute those percentiles. In those cases, both the VSP and hand calculations gave the same values shown in the table.

Table 11. Values of Descriptive Statistics for the Small and Large Data Sets Used to Verify the Accuracy of Computations in VSP

Descriptive Statistic	Small Data Set (n = 10)	Large Data Set (n = 150)
Minimum	3	0.0040
Maximum	9	0.9862
Range/minimum minus maximum	6	0.9822
Mean	5.4	0.507
Standard Deviation	2.011	0.2786
Variance	4.04	0.0776
Standard Error (Standard	0.636	0.0227

Deviation of the Estimated Mean)		
Interquartile Range (75th Percentile minus 25th Percentile)	3.5	0.456
1st Percentile	3*	0.011*
5th Percentile	3*	0.061*
25th Percentile	3.75	0.294
50th Percentile (Median)	5	0.516
75th Percentile	7.25	0.75
95th Percentile	9*	0.93*
99th Percentile	9*	0.977*

*Computed by hand

Upper Confidence Limits on the Mean

Recall from Section 2.0 that VSP computes UCLs in two ways: when data are assumed or known to be normally distributed and when no assumption about the distribution is made. Hand computations were conducted to verify the accuracy of VSP computations of the UCLs. The UCLs were computed using the two data sets (small and large) described above. The UCLs computed by VSP are given in Table 12. Hand calculations gave identical results in all cases.

Table 12. Comparing VSP and Hand Calculations of UCLs for Two Data Sets

Number of Samples	**Assume Data are Normally Distributed***		**No Data Distribution Assumption****	
	95% Confidence	**99% Confidence**	**95% Confidence**	**99% Confidence**
n = 10	6.566	7.194	8.172	11.73
n = 150	0.545	0.560	0.606	0.733

*Computed as follows: $UCL = \bar{x} + t_{1-\alpha,n-1}\frac{s}{\sqrt{n}}$

**Computed as follows: $UCL = \bar{x} + \sqrt{\frac{1}{\alpha} - 1}\ \frac{s}{\sqrt{n}}$

Tests that Data are Normally Distributed

VSP computes two statistical tests to evaluate if data sets are normally distributed: the Shapiro-Wilk (SW) test and the Lilliefors (LF) test. These tests were described in Section 2.4.2. VSP computes the SW test when $n \leq 50$ and the LF test when $n > 50$. Minitab® computes the LF test, but not the SW test. Hence, the accuracy of VSPs computations for the SW test was verified using hand calculations rather than Minitab®.

As both Minitab® and VSP compute the LF test statistic, Minitab® was used to verify the accuracy of VSPs LF test results. VSP also computes LF test critical values for three different significance levels (values of α, the false rejection decision error rate). As Minitab does not compute these critical values, hand calculations were conducted to verify the accuracy of VSP's computation of critical values. The accuracy of VSP calculations for the SW and LF tests was evaluated for the two data sets (small and large) described above. The results are reported in Table 13.

The VSP computations for the SW test are all very close in value to those computed by hand. The agreement of VSP and Minitab in computing the LF test statistic was excellent. The agreement of VSP and hand computations for the LF critical values was somewhat less for the $\alpha = 0.10$ critical value than for $\alpha = 0.01$ and $\alpha = 0.05$ critical values. However, these critical values were computed by hand using the approximate formula at the bottom on Table A14 in Conover (1999, page 548). The slight disagreement between VSP and hand-computed critical values is believed to be due to using the approximate formula rather than to errors within VSP.

Table 13. Comparing VSP, Minitab®, and Hand Calculations for the Shapiro-Wilk and Lilliefors Tests that the Data are Normally Distributed for a Small Data Set (n = 10) and a Large Data Set (n = 150)

	Test Statistic		**Test Critical Values**					
			$\alpha = 0.01$		**$\alpha = 0.05$**		**$\alpha = 0.10$**	
	VSP	**Minitab or by Hand**	**VSP**	**Hand**	**VSP**	**Hand**	**VSP**	**Hand**
Shapiro-Wilk Test (n = 10)	0.9024	0.9024*	0.781	0.781†	0.842	0.842†	0.869	0.869†
Lilliefors Test (n = 150)	0.0742	0.074**	0.0842	0.0841 ††	0.0723	0.0723† †	0.0657	0.06676 ††

*Computed by hand
**Computed by Minitab®
†Read from Table of Critical Values in Gilbert (1987, Table A7, page 261)
††Computed by hand using Table A14 in Conover (1999, page 548)

All methods above except the summary statistics were directly transferred to VSP from the ProUCL (2004) software (Version 3.0). The summary statistics and statistical/graphical tests for normal distributions listed above are described in ProUCL (2004). The UCL formula for normally distributed data is described in ProUCL (2004, Equation 32, page A-24). The nonparametric Chebyshev UCL is described in ProUCL (2004, Equation 46, page A-32). These UCL methods are also described in EPA (2002b).

Comparing Graphical Displays of VSP

Both VSP and Minitab® produce histogram, box-and-whisker plots, and quantile-quantile plots. These plots were constructed for both the small (n = 10) and the large (n = 150) data set. The graphs produced by Minitab® were identical to those produced by VSP aside from different display formats. For example, the quantile-quantile plots (also called "probability plots") for the large data set for VSP and Minitab® are given in Figures 16 and 17, respectively. The similarity of the two figures is obvious. Both indicate that the normality assumption for these n = 150 data should be rejected.

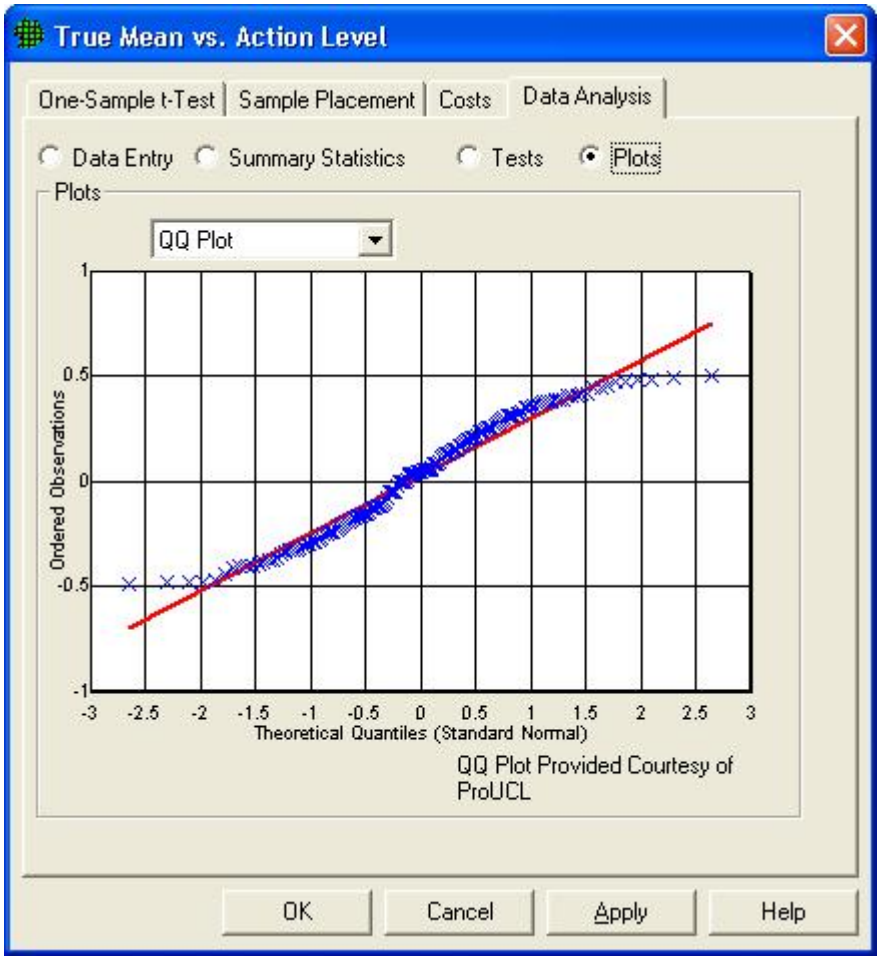

Figure 16. Quantile-Quantile Plot for the Large Data Set (n = 150) Created in VSP

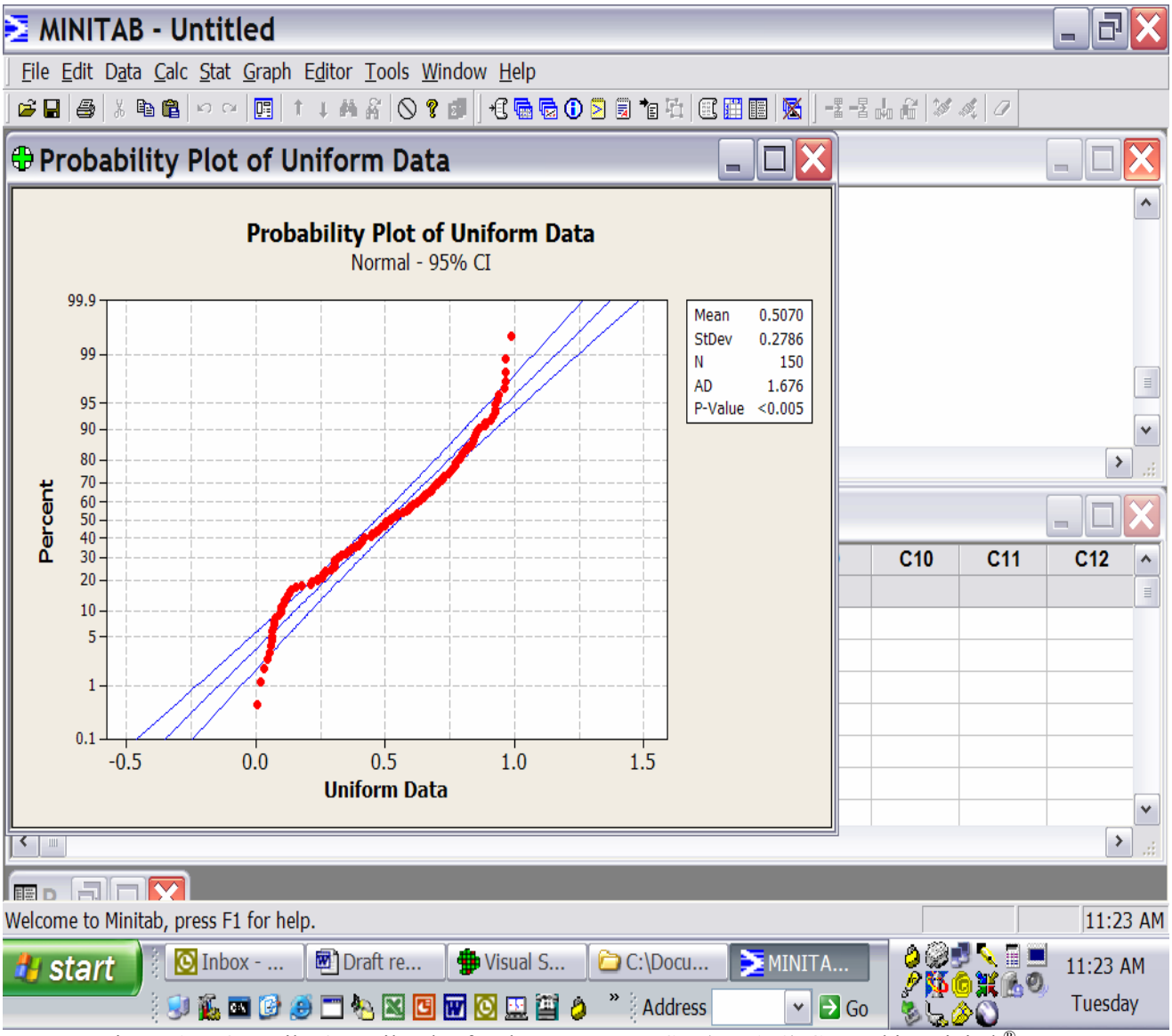

Figure 17. Quantile-Quantile Plot for the Large Data Set (n = 150) Created in Minitab®

4.0 References

Barnard, G.A. 1952. "The Frequency Justification of Certain Sequential Tests," *Biometricka* 39:144-150.

Bowen, W.M. and C.A. Bennett. 1988. *Statistical Methods for Nuclear Material Management*. NUREG/CR-4604. U.S. Nuclear Regulatory Commission, Washington, D.C.

Conover, W.J. 1999. *Practical Nonparametric Statistics*, 3rd edition, John Wiley & Sons, Inc, New York.

Davidson, J.R. 1995b. *ELIPGRID-PC: Upgraded Version.* ORNL/TM-13103, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Davidson, J.R. Jr. 2001. *Verification of the Accuracy of Sample-Size Equation Calculations for Visual Sample Plan Version 0.9C.* PNNL-13434. Pacific Northwest National Laboratory, Richland, WA.

Desu, M.M. and D. Raghavarao. 1990. Sample Size Methodology. Academic Press, New York.

EPA. 1992. *Methods for Evaluating the Attainment of Cleanup Standards, Volume 2: Ground Water.* EPA 230-R-92-014. Office of Policy, Planning, and Evaluation. U.S. Environmental Protection Agency, Washington, D.C.

EPA. 1994. *Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT)-User's Guide,* Version 4.0. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 1997. *Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)*. EPA 402-R-97-016, NUREG-1575, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 2000a. *Guidance for the Data Quality Objectives Process, EPA QA/G-4.* EPA/600/R-96/055. Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C. August 2000.

EPA. 2000b. *Guidance for Data Quality Assessment – Practical Methods for Data Analysis – EPA QA/G-9, QA00 Update*. EPA/600/R-96/084, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 2002a. *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*, OSWER 9285.6-10, U.S. Environmental Protection Agency, Washington, D.C. December 2002.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*, Wiley & Sons, New York, NY.

Gilbert, R.O., J.R. Davidson, Jr., J.E. Wilson, and B.A. Pulsipher. 2001. *Visual Sample Plan (VSP) Models and Code Verification.* PNNL-13450. Pacific Northwest National Laboratory, Richland, WA, February 2001.

Gilbert, R.O., J.E. Wilson, R.F. O'Brien, D.K. Carlson, D.J. Bates, B.A. Pulsipher, and C.A. McKinstry. 2002. *Version 2.0 Visual Sample Plan (VSP): Models and Code Verification.* PNNL-13991. Pacific Northwest National Laboratory, Richland, WA, August 2002.

Gilbert, R.O., J.E. Wilson, R.F. O'Brien, D.K. Carlson, B.A. Pulsipher, and D.J. Bates. 2003. *Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification*. PNNL-14267. Pacific Northwest National Laboratory, Richland, WA, April 2003.

Gogolak, C.V., G.E. Powers and A.M. Huffert. 1997. *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys*. NUREG-1505, U.S. Nuclear Regulatory Commission, Washington, D.C.

Grieve, A.P. 1994. "A Further Note on Sampling to Locate Rare Defectives with Strong Prior Evidence," *Biometrika* 81(4):787-789.

Guenther, W.C. 1977. *Sampling Inspection in Statistical Quality Control*. Griffin's Statistical Monographs and Courses, No. 37, London: Charles Griffin.

Guenther, W.C. 1981. "Sample Size Formulas for Normal Theory T=Tests." *The American Statistician* 35(4):243-244.

Hahn, G.J. and W.Q. Meeker. 1991. *Statistical Intervals*. Wiley & Sons, Inc, New York, NY.

Hassig, N.L., J.E. Wilson, R.O. Gilbert and B.A. Pulsipher. 2004. *Visual Sample Plan Version 3.0 User's Guide*. PNNL-14970. Pacific Northwest National Laboratory, Richland, WA.

Hassig, N.L., J.E. Wilson, R.O. Gilbert, B.A. Pulsipher, L.L. Nuffer. 2005. Visual Sample Plan Version 3.2 User's Guide, (Pending Publication), Pacific Northwest National Laboratory, Richland, Washington.

Helsel, D.R. 2005. *Nondetects and Data Analysis, Statistics for Censored Environmental Data*, Wiley & Sons, New York, NY.

Johnson, N.L. and S. Kotz. 1970. *Continuous Univariate Distributions-2*, Houghton Mifflin Company, Boston, MA.

Lyles, R.H. and L.L. Kupper. 1996. "On Strategies for Comparing Occupational Exposure Data to Limits," *American Industrial Hygiene Association Journal* 57:6-15.

Millard, S.P. and N.K. Neerchal. 2001. *Environmental Statistics with S-Plus*. CRC Press, New York.

Minitab® (2003). Minitab Software, Version 14. Information and software available at http://www.minitab.com.

Noether, G.E. 1987. "Sample Size Determination for Some Common Nonparametric Tests." *Journal of the American Statistical Association* 82:645-647.

Owen, D.R. 1965. *Handbook of Statistical Tables.* Addison-Wesley, Reading, MA.

Patil, J.K., C.H. Kapadia and D.B. Owen. 1976. *Handbook of Statistical Distributions*, Marcel Dekker, Inc., New York.

ProUCL. 2004. *ProUCL Version 3.0 User Guide April 2004*. Available for download from http://www.epa.gov/nerlesd1/tsc/tsc.htm.

Rothschild, V. and N. Logothetis. 1986. *Probability Distributions*, John Wiley & Sons, New York, NY.

SAS® (Statistical Analysis System) software (http://www.SAS.com).

Schilling, E.G. 1978. "A Lot Sensitive Sampling Plan for Compliance Testing and Acceptance Inspection," *Journal of Quality Technology* 10(2):47-51.

Schilling, E.G. 1982. *Acceptance Sampling in Quality Control*, Marcel Dekker, New York.

Singer, D.A. 1972. "ELIPGRID, a FORTRAN IV Program for Calculating the Probability of Success in Locating Elliptical Targets with Square, Rectangular and Hexagonal Grids." *Geocom Programs* 4:1-16.

Singer, D.A. 1975. "Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets." *Journal of Research of the U.S. Geological Survey* 3(2):163-167.

Singer, D.A. and F.E. Wickman. 1969. *Probability Tables for Locating Elliptical Targets with Square, Rectangular and Hexagonal Point Nets*. Pennsylvania State University, University Park, Pennsylvania.

S-PLUS® software (http://www.insightful.com/products/splus/default.asp).

Wald, A. 1947. *Sequential Analysis*. Dover, NY.

Wallis, W.A. 1947. "Uses of Variables in Acceptance Inspection for Percent Defective." *Techniques of Statistical Analysis*, eds. C. Eisenhart, M.W. Hastay, and W.A. Wallis. McGraw-Hill, New York.

Wetherill, G.B. 1966. *Sequential Methods in Statistics*. Chapman and Hall, New York.

Wright, T. 1992. "A Note on Sampling to Locate Rare Defectives with Strong Prior Evidence," *Biometrika* 79(4):685-691.

Zirschky, J. and R.O. Gilbert. 1984. "Detecting Hot Spots at Hazardous Waste Sites." *Chemical Engineer*, July 9, pp. 97-100.

Distribution

No. of Copies

OFFSITE

Technical Support Working Group

Rebecca Blackmon

Department of Homeland Security

Lance Brooks
Teresa Lustig

Department of Energy

David Biancosino
George Detsis
Chip Langdon

U.S. Environmental Protection Agency

Mike Carter
JoAnn Griffith
Les Sparks
John Warren

General Accounting Office

Sushil Sharma

No. of Copies

ONSITE

Pacific Northwest National Laboratory

R.O. Gilbert	K6-08
D.L. Hadley	K5-16
N.L. Hassig	K6-08
J.E. Hathaway	K6-08
S.W. Martin	K8-21
B.M. Matzke	K6-08
C.A. McKinstry	K6-08
L.L Nuffer	K1-85
B.A. Pulsipher	K6-08
B.A. Seiders	K8-46
M.L. Spradling	K8-21
N.B. Valentine	K2-10
J.E. Wilson	K6-08