109944

United States Environmental Protection Agency

ENTERED

ProUCL Version 4.00.05 User Guide

(Draft)

33934

RESEARCH AND DEVELOPMENT

EPA/600/R-07/038
May 2010
www.epa.gov

ProUCL Version 4.00.05 User Guide
(Draft)

Prepared for:

Brian Schumacher
and
John Nocerino
U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Environmental Sciences Division
Characterization and Monitoring Branch
944 E. Harmon Ave.
Las Vegas, NV 89119

Felicia Barnett

Technology Support Center
U.S. Environmental Protection Agency
Region 4
61 Forsyth Street SW,
Atlanta, GA 30303

Prepared by:

Anita Singh, Ph.D.[1]
Robert Maichle[1]
Narain Armbya[1]
Ashok K. Singh, Ph.D.[2]
Sanghee E. Lee[1]

[1]Lockheed Martin Environmental Services
1050 E. Flamingo Road, Suite N240
Las Vegas, NV 89119

[2]Department of Hotel Management
University of Nevada, Las Vegas
Las Vegas, NV 89154

Notice: Although this work was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency policy. Mention of trade names and commercial products does not constitute endorsement or recommendation for use.

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

129cmb07

Notice

The United States Environmental Protection Agency (EPA) through its Office of Research and Development (ORD) funded and managed the research described here. It has been peer reviewed by the EPA and approved for publication. Mention of trade names and commercial products does not constitute endorsement or recommendation by the EPA for use.

ProUCL software was developed by Lockheed Martin under a contract with the EPA and is made available through the EPA Technical Support Center in Las Vegas, Nevada. Use of any portion of ProUCL that does not comply with the ProUCL User Guide is not recommended.

ProUCL contains embedded licensed software. Any modification of the ProUCL source code may violate the embedded licensed software agreements and is expressly forbidden.

ProUCL software provided by the EPA was scanned with McAfee VirusScan and is certified free of viruses.

With respect to ProUCL distributed software and documentation, neither the EPA nor any of their employees, assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed. Furthermore, software and documentation are supplied “as-is” without guarantee or warranty, expressed or implied, including without limitation, any warranty of merchantability or fitness for a specific purpose.

Changes from ProUCL 4.0 (Version 4.00.00) to ProUCL 4.00.002

Although extensive changes were made in the code for ProUCL 4.0 (version 4.00.00) to produce ProUCL 4.00.02, those changes are transparent to the users. Most of those changes were made so that ProUCL 4.00.02 is compatible with our developing statistical software, Scout (e.g., both programs share the same statistical libraries). ProUCL will also reside as a separate module in Scout as a research tool.

There is a very minor correction of a displayed value in one of the hypothesis tests, the two sample t-test. The p-value associated with the t-test was computed in two different ways: one way is correct and the other way, although it produced subtle differences, is incorrect. The incorrect method has been removed from ProUCL 4.00.02.

Several extra warning messages have been added to ProUCL 4.00.02, mainly in regard to attempting tests when a data set is very small ($n < 5$), when the number of detected values is small (e.g., only zero, one, or two), or when all of the values are non-detected values. For an example, some screens depicting those warning messages are included in the newly added Section 2.11 (page 40) of this ProUCL 4.00.02 User Guide.

The only software files that were changed from ProUCL version 4.0 (4.00.00) to version 4.0.02 were updates in the ProUCL.exe file, and updates to the StatsLib.dll file to produce a more advanced ScoutLib.dll file. Very minor changes were made to this ProUCL 4.00.02 User Guide, including: changes to avoid inappropriate user inputs (warnings), changes to the title page, the inclusion of an acknowledgement page, and the inclusion of a contact information page.

Changes from ProUCL 4.00.02 to ProUCL 4.00.04

ProUCL 4.00.04 is an upgrade of ProUCL Version 4.00.02 which represents an upgrade of ProUCL 4.0 (EPA, 2004). ProUCL 4.00.04 contains all statistical methods as available in ProUCL 4.00.02 to address various environmental issues for both full data sets without nondetects (NDs) and for data sets with NDs (also known as left-censored data sets). In addition to having all methods available in ProUCL 4.00.02, ProUCL 4.00.04 has extended version of Shapiro-Wilk (S-W) test that can perform normal and lognormal goodness-of-fit tests for data sets of sizes upto 2000. Moreover, ProUCL 4.00.04 can compute upper prediction and upper tolerance limits based upon gamma distribution. Some modifications have also been made in decision tables and recommendations made by ProUCL to estimate the EPC terms. Specifically, based upon recent experience, developers of ProUCL are re-iterating that the use of lognormal distribution to estimate EPC terms should be avoided, as the use of lognormal distribution yields unrealistic and highly unstable UCLs. In an effort to simplify the EPC estimation process, for highly skewed lognormally distributed data sets, developers are recommending the use of appropriate nonparametric Chebyshev *(mean Sd) UCLs*. These changes have been incorporated in various decision tables included in ProUCL 4.00.04 Technical Guide and ProUCL 4.00.04 User Guide. Recommendations made by ProUCL 4.00.02 have been changed accordingly in ProUCL 4.00.04. Some minor bugs as suggested by ProUCL 4.0 and ProUCL 4.00.02 users have also been addressed in this upgraded version of ProUCL software package.

Changes from ProUCL 4.00.04 to ProUCL 4.00.05

ProUCL version 4.00.05 is an upgrade of ProUCL Version 4.00.04 (EPA, 2008). ProUCL 4.00.05 consists of all of the statistical and graphical methods that are available in previous ProUCL 4.0 versions to address various environmental issues for full data sets without nondetect (ND) observations and also for data sets with NDs and below detection limit observations. Several additions (e.g., sample size determination module), enhancements (File module), modifications (e.g., p-values of WRS/WMW test, Gehan test) have been made in ProUCL 4.00.05. Some bugs (e.g., correction in adjusted Gamma UCLs) as suggested and found by users of previous ProUCL 4.0 versions have been addressed in this version of ProUCL. With the inclusion of the sample size determination module, ProUCL 4.00.05 will serve as a comprehensive statistical software package equipped with statistical methods and graphical tools needed to address environmental sampling and statistical issues described in various CERCLA (EPA 2002a, 2002b, 2006) and RCRA (EPA 1989b, 1992b, 2002c, 2009) guidance documents. For data sets with and without nondetect observations, ProUCL 4.00.05 also provides statistical methods to address reference area and survey unit sampling issues described in MARSSIM (EPA 2000) document. In addition to sample size determination methods, ProUCL 4.00.05 offers parametric and nonparametric statistical methods (e.g., Sign test, Wilcoxon Rank Sum test) often used to address statistical issues described in MARSSIM (EPA 2000) guidance document. The user friendly sample size determination module of ProUCL 4.00.05 has a straight forward mechanism to enter the desired/pre-specified decision parameters needed to compute appropriate sample size(s) for the selected statistical application. The Sample Size module of ProUCL 4.00.05 provides sample size determination methods for most of the parametric and nonparametric one-sided and two-sided hypotheses testing approaches available in the Hypothesis Testing module of ProUCL 4.0. Theoretical details of the Sample Size module are given in "Supplement to ProUCL 4.0 Technical Guide: Determining Minimum Sample Sizes for User Specified Decision Parameters." Some specific changes made in ProUCL 4.00.05 are listed as follows.

- File Option: This option has been upgraded to open *.xls files by default. In the earlier versions of ProUCL, this option was available only for *.wst, *.ost and *.gst files and Excel files had to be imported. Now you can use the import option to read multiple worksheets from one Excel file. ProUCL 4.00.05 will import worksheets until all worksheets are read or a blank or empty worksheet is encountered.

- Displaying All Menu Options: ProUCL 4.00.05 now displays all available menu options even before opening a valid (e.g., non empty) data file. However, the user has to open a valid data file before activating a menu option and using a statistical or graphical method available in ProUCL 4.00.05. Obviously, no statistical method can be used on an empty (without any data) spreadsheet.

- Sample Size Module: Several parametric (assuming normal distribution) and nonparametric sample size determination formulae as used and described in various EPA guidance documents (e.g., EPA 1989a, 1989b, 1992, 2000, 2002a, 2002b, 2002c, 2006, and 2009) have been incorporated in ProUCL 4.00.05. Inclusion of this module will help the users to develop/design DQOs based sampling plans with pre-specified values of decision error rates (α = Type I and β =Type II) and width of the gray region, Δ around the parameter of interest (e.g., mean concentration, proportion of sampled observations exceeding the action level). Basic sample size determination formulae have been incorporated for sampling of continuous characteristics (lead, Ra 226) as well as for attributes (e.g., proportion exceeding a specified threshold). Additionally, sample size formulae for acceptance sampling of discrete objects (e.g., drums) have also been

incorporated in this module. The detailed description of the statistical methods and formulae used in this module are described in the supplement document to ProUCL 4.0 Technical Guide.

- Adjusted Gamma UCL: There was a minor bug in the computation of adjusted α level of significance (called β level) used for calculating Adjusted Gamma UCLs. This error has been corrected.

- Computation of Nonparametric Percentiles: There are several ways to compute nonparametric percentiles; and percentiles obtained using different methods can differ slightly. For graphical displays, ProUCL 4.00.05 uses development software, ChartFX. Thus boxplots generated by ProUCL display percentiles (e.g., median and quartiles) as computed by ChartFX. In order to avoid confusion, the percentile algorithm used in ProUCL has been modified so that it now computes and displays comparable percentiles as computed by ChartFX.

- UCL based upon Winsorization Method: In the computation of Winsorized UCLs, the sample standard deviation of the winsorized data was being used instead of the approximate unbiased estimate of the population standard deviation from the winsorized data as detailed in ProUCL 4.00.04 Technical Guide. This has been corrected in ProUCL 4.00.05.

- Displaying K values used in UTLs: The tolerance factor, K, based on the number of valid observations, level of confidence coefficient, and coverage percentage is displayed along with UTL statistics and other relevant input parameters.

- Fixes in the p- values associated with WSR/WMW Test, Gehan Test, and Equality of Variances Test: More efficient algorithms have been incorporated in ProUCL 4.00.05 to compute p-values associated with the test statistics associated with these two tests.

- Additional Critical values associated with Land's H Statistic: In addition to 0.9 and 0.95 confidence coefficients; 0.975, 0.99 and 0.995 confidence levels have been incorporated in ProUCL 4.00.05 to compute H-UCLs based upon a lognormal distribution.

- Adjustment in Precision Associated with Lognormal and Gamma ROS Methods: The lower bound associated with Lognormal ROS and Gamma ROS extrapolated estimates have been extended from 1e-7 to 1e-10. ProUCL 4.00.05 issues a warning message when extrapolated ROS estimates lie below 1e-10.

- Some terminology changes have been made in single sample hypotheses approaches available in the Hypothesis Testing module. Specifically the phrase “Compliance Limit” have been replaced by the phrase “Action Level”.

Contact Information for ProUCL 4.00.05

The ProUCL software is developed under the direction of the Technical Support Center (TSC). All comments and questions concerning ProUCL should be addressed to:

Felicia Barnett, (HSTL)
US EPA, Region 4
61 Forsyth Street, S.W.
Atlanta, GA 30303-8960
barnett.felicia@epa.gov
(404) 562-8659
Fax: (404) 562-8439

Executive Summary

Statistical inference, including both estimation and hypotheses testing approaches, is routinely used to:

1. Estimate environmental parameters of interest, such as exposure point concentration (EPC) terms, not-to-exceed values, and background level threshold values (BTVs) for contaminants of potential concern (COPC),
2. Identify areas of concern (AOC) at a contaminated site,
3. Compare contaminant concentrations found at two or more AOCs of a contaminated site,
4. Compare contaminant concentrations found at an AOC with background or reference area contaminant concentrations, and
5. Compare site concentrations with a cleanup standard to verify the attainment of cleanup standards.

Several exposure and risk management and cleanup decisions in support of United States Environmental Protection Agency (EPA) projects are often made based upon the mean concentrations of the COPCs. A 95% upper confidence limit (UCL95) of the unknown population (e.g., an AOC) arithmetic mean (AM), μ_1, can be used to:

- Estimate the EPC term of the AOC under investigation,
- Determine the attainment of cleanup standards,
- Compare site mean concentrations with reference area mean concentrations, and
- Estimate background level mean contaminant concentrations. The background mean contaminant concentration level may be used to compare the mean of an area of concern. It should be noted that it is not appropriate to compare individual point-by-point site observations with the background mean concentration level.

It is important to compute a reliable and stable *UCL95* of the population mean using the available data. The *UCL95* should approximately provide the 95% coverage for the unknown population mean, μ_1. Based upon the available background data, it is equally important to compute reliable and stable upper percentiles, upper prediction limits (*UPLs*), or upper tolerance limits (*UTLs*). These upper limits based upon background (or reference) data are used as estimates of BTVs, compliance limits (*CL*), or not-to-exceed values. These upper limits are often used in site (point-by-point) versus background comparison evaluations.

Environmental scientists often encounter trace level concentrations of COPCs when evaluating sample analytical results. Those low level analytical results cannot be measured accurately and, therefore, are typically reported as less than one or more detection limit (*DL*) values (also called nondetects). However, practitioners need to obtain reliable estimates of the population mean, μ_1, and the population standard deviation, σ_1, and upper limits including the *UCL* of the population mass or mean, the *UPL*, and the *UTL* based upon data sets with nondetect (*ND*) observations. Additionally, they may have to use hypotheses testing approaches to verify the attainment of cleanup standards, and compare site and background concentrations of COPCs as mentioned above.

Background evaluation studies, BTVs, and not-to-exceed values should be estimated based upon defensible background data sets. The estimated BTVs or not-to-exceed values are then used to identify the COPCs, to identify the site AOCs or hot spots, and to compare the contaminant concentrations at a site

with background concentrations. The use of appropriate statistical methods and limits for site versus background comparisons is based upon the following factors:

- Objective of the study,
- Environmental medium (e.g., soil, groundwater, sediment, air) of concern,
- Quantity and quality of the available data,
- Estimation of a not-to-exceed value or of a mean contaminant concentration,
- Pre-established or unknown cleanup standards and BTVs, and
6. Sampling distributions (parametric or nonparametric) of the concentration data sets collected from the site and background areas under investigation.

In background versus site comparison evaluations, the environmental population parameters of interest may include:

- Preliminary remediation goals (PRGs),
- Soil screening levels (SSLs),
- RBC standards,
- BTVs, not-to-exceed values, and
- Compliance limit, maximum concentration limit (MCL), or alternative concentration limit (ACL), frequently used in groundwater applications.

When the environmental parameters listed above are not known or pre-established, appropriate upper statistical limits are used to estimate those parameters. The UPL, UTL, and upper percentiles are used to estimate the BTVs and not-to-exceed values. Depending upon the site data availability, point-by-point site observations are compared with the estimated (or pre-established) BTVs and not-to-exceed values. If enough site and background data are available, two-sample hypotheses testing approaches are used to compare site concentrations with background concentrations levels. These statistical methods can also be used to compare contaminant concentrations of two site AOCs, surface and subsurface contaminant concentrations, or upgradient versus monitoring well contaminant concentrations.

ProUCL 4.00.05 is an upgrade of ProUCL Version 4.00.02 which represents an upgrade of ProUCL 4.0 (EPA, 2007). ProUCL 4.00.05 contains all statistical methods as available in ProUCL 4.00.02 to address various environmental issues for both full data sets without nondetects (NDs) and for data sets with NDs (also known as left-censored data sets). In addition to having all methods available in ProUCL 4.0 as described below, ProUCL 4.00.05 has extended version of Shapiro-Wilk (S-W) test that can perform normal and lognormal goodness-of-fit tests for data sets of sizes upto 2000. Moreover, ProUCL 4.00.05 can compute upper prediction and upper tolerance limits based upon gamma distribution. Some modifications have also been made in decision tables and recommendations made by ProUCL to estimate the EPC terms. Specifically, based upon recent experience, developers of ProUCL are re-iterating that the use of lognormal distribution to estimate EPC terms should be avoided, as the use of lognormal distribution yields unrealistic and highly unstable UCLs. In an effort to simplify the EPC estimation process, for highly skewed lognormally distributed data sets, developers are recommending the use of appropriate nonparametric Chebyshev *(mean Sd) UCLs.* These changes have been incorporated in various decision tables included in ProUCL 4.0 Technical Guide and ProUCL 4.0 User Guide. Recommendations made by ProUCL 4.00.02 have been changed accordingly in ProUCL 4.00.05. Some minor bugs as suggested by ProUCL 4.0 users have also been addressed in this upgraded version of ProUCL software package.

ProUCL 4.0/ProUCL 4.00.02 contains:

- Rigorous parametric and nonparametric (including bootstrap methods) statistical methods (instead of simple ad hoc or substitution methods) that can be used on full data sets without nondetects and on data sets with below detection limit (BDL) or ND observations.

2. State-of-the-art parametric and nonparametric UCL, UPL, and UTL computation methods. These methods can be used on full-uncensored data sets without nondetects and also on data sets with BDL observations. Some of the methods (e.g., Kaplan-Meier method, ROS methods) are applicable on left-censored data sets having multiple detection limits. The UCL and other upper limit computation methods cover a wide range of skewed data sets with and without the BDLs.

3. Single sample (e.g., Student's t-test, sign test, Proportion test, Wilcoxon Singed Rank test) and two-sample (Student's t-test, Wilcoxon-Mann-Whitney test, Gehan test, quantile test) parametric and nonparametric hypotheses testing approaches for data sets with and without ND observations. These hypothesis testing approaches can be used to: verify the attainment of cleanup standards, perform site versus background comparisons, and compare two or more AOCs, monitoring wells (MWs).

4. The single sample hypotheses testing approaches are used to compare site mean, site median, site proportion, or a site percentile (e.g., 95th) to a compliance limit (action level, regularity limit). The hypotheses testing approaches can handle both full-uncensored data sets without nondetects, and left-censored data sets with nondetects. Simple two-sample hypotheses testing methods to compare two populations are available in ProUCL 4.0, such as two-sample t-tests, Wilcoxon-Mann-Whitney (WMW) Rank Sum test, quantile test, Gehan's test, and dispersion test. Variations of hypothesis testing methods (e.g., Levene's method to compare dispersions, generalized WRS test) are easily available in most commercial and freely available software packages (e.g., MINITAB, R).

5. ProUCL 4.0 also includes graphical methods (e.g., box plots, multiple Q-Q plots, histogram) to compare two or more populations. ProUCL 4.0 can also be used to display a box plot of one population (e.g., site data) with compliance limits or upper limits (e.g., UPL) of other population (background area) superimposed on the same graph. This kind of graph provides a useful visual comparison of site data with a compliance limit or BTVs. Graphical displays of a data set (e.g., Q-Q plot) should be used to gain insight knowledge contained in a data set that may not otherwise be clear by looking at simple test statistics such as t-test, Dixon test statistic, or Shapiro-Wilk (S-W) test statistic.

6. ProUCL 4.0 can process multiple contaminants (variables) simultaneously and has the capability of processing data by groups. A valid group column should be included in the data file.

7. ProUCL 4.0 provides GOF test for data sets with nondetects. The user can create additional columns to store extrapolated (estimated) values for nondetects based upon normal ROS, gamma ROS, and lognormal ROS (robust ROS) methods.

ProUCL 4.0/ProUCL 4.00.02 retains all of the capabilities of ProUCL 3.0, including goodness-of-fit (GOF) tests for a normal, lognormal, and a gamma distribution and computation of UCLs based upon full data sets without nondetects. Graphical displays and GOF tests for data sets with BDL observations have also been included in ProUCL 4.0. It is re-emphasized that the computation of appropriate UCLs, UPLs, and other limits is based upon the assumption that the data set under study represents a single a single population. This means that the data set used to compute the limits should represent a single statistical population. For example, a background data set should represent a defensible background data set free of outlying observations. ProUCL 4.0 includes simple and commonly used classical outlier identification procedures, such as the Dixon test and the Rosner test. These procedures are included as an aid to identify outliers. These simple classical outlier tests often suffer from masking effects in the presence of multiple outliers. Description and use of robust and resistant outlier procedures is beyond the scope of ProUCL 4.0 software.

It is suggested that the classical outlier procedures should always be accompanied by graphical displays including box plots and Q-Q plots. The use of a Q-Q plot is useful to identify multiple or mixture samples that might be present in a data set. However, the decision regarding the proper disposition of outliers (e.g., to include or not to include outliers in statistical analyses; or to collect additional verification samples) should be made by members of the project team and experts familiar with site and background conditions. Guidance on the disposition of outliers and their accommodation in a data set by using a transformation (e.g., lognormal distribution) is discussed in Chapter 1 of this User Guide.

ProUCL 4.0 has improved graphical methods, which may be used to compare the concentrations of two or more populations such as:

- Site versus background populations,
- Surface versus subsurface concentrations,
- Concentrations of two or more AOCs, and
- Identification of mixture samples and/or potential outliers

These graphical methods include multiple quantile-quantile (Q-Q) plots, side-by-side box plots, and histograms. Whenever possible, it is desirable to supplement statistical results with useful visual displays of data sets. There is no substitute for graphical displays of a data set. For example, in addition to providing information about the data distribution, a normal Q-Q plot can also help identify outliers and multiple populations that may be present in a data set. On a Q-Q plot, observations well separated from the majority of the data may represent potential outliers, and jumps and breaks of significant magnitude may suggest the presence of observations from multiple populations in the data set. It is suggested that analytical outlier tests (e.g., Rosner test) and goodness-of-fit (G.O.F.) tests (e.g., SW test) should always be supplemented with the graphical displays such as Q-Q plot and box plot.

ProUCL 4.00.05 (and all its previous versions) serves as a companion software package for *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites* (EPA, 2002a) and *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b). ProUCL 4.00.05 is also useful to verify the attainment of cleanup standards (EPA, 1989). ProUCL 4.00.05 can also be used to perform two-sample hypotheses tests and to compute various upper limits often needed in groundwater monitoring applications (EPA, 1992 and EPA, 2004).

Acronyms and Abbreviations

% NDs	Percentage of Nondetect observations
ACL	Alternative concentration limit
A-D, AD	Anderson-Darling test
AM	Arithmetic mean
AOC	area(s) of concern
BC	Box-Cox-type transformation
BCA	bias-corrected accelerated bootstrap method
BDL	below detection limit
BTV	background threshold value
BW	Black and White (for printing)
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CL	compliance limit
CLT	central limit theorem
CMLE	Cohen's maximum likelihood estimate
COPC	contaminant(s) of potential concern
CV	Coefficient of Variation
DL	detection limit
DL/2 (t)	UCL based upon DL/2 method using Student's t-distribution cutoff value
DL/2 Estimates	estimates based upon data set with nondetects replaced by half of the respective detection limits
DQO	data quality objective
EA	exposure area
EDF	empirical distribution function
EM	expectation maximization
EPA	Environmental Protection Agency
EPC	exposure point concentration

FP-ROS (Land)	UCL based upon fully parametric ROS method using Land's H-statistic
Gamma ROS (Approx.)	UCL based upon Gamma ROS method using the gamma approximate-UCL method
Gamma ROS (BCA)	UCL based upon Gamma ROS method using the bias-corrected accelerated bootstrap method
GOF, G.O.F.	goodness-of-fit
H-UCL	UCL based upon Land's H-statistic
ID	identification code
IQR	interquartile range
K	Next K, Other K, Future K
KM (%)	UCL based upon Kaplan-Meier estimates using the percentile bootstrap method
KM (Chebyshev)	UCL based upon Kaplan-Meier estimates using the Chebyshev inequality
KM (t)	UCL based upon Kaplan-Meier estimates using the Student's t-distribution cutoff value
KM (z)	UCL based upon Kaplan-Meier estimates using standard normal distribution cutoff value
K-M, KM	Kaplan-Meier
K-S, KS	Kolmogorov-Smirnov
LN	lognormal distribution
Log-ROS Estimates	estimates based upon data set with extrapolated nondetect values obtained using robust ROS method
MAD	Median Absolute Deviation
Maximum	Maximum value
MCL	maximum concentration limit
Mean	classical average value
Median	Median value
Minimum	Minimum value
MLE	maximum likelihood estimate
MLE (t)	UCL based upon maximum likelihood estimates using Student's t-distribution cutoff value

MLE (Tiku)	UCL based upon maximum likelihood estimates using the Tiku's method
Multi Q-Q	multiple quantile-quantile plot
MVUE	minimum variance unbiased estimate
ND	nondetect or nondetects
NERL	National Exposure Research Laboratory
NumNDs	Number of Nondetects
NumObs	Number of Observations
ORD	Office of Research and Development
PRG	preliminary remediation goals
Q-Q	quantile-quantile
RBC	risk-based cleanup
RCRA	Resource Conservation and Recovery Act
ROS	regression on order statistics
RU	remediation unit
S	substantial difference
SD, *Sd, sd*	standard deviation
SSL	soil screening levels
S-W, SW	Shapiro-Wilk
UCL	upper confidence limit
UCL95, 95% UCL	95% upper confidence limit
UPL	upper prediction limit
UPL95, 95% UPL	95% upper prediction limit
USEPA	United States Environmental Protection Agency
UTL	upper tolerance limit
Variance	classical variance
WMW	Wilcoxon-Mann-Whitney
WRS	Wilcoxon Rank Sum
WSR	Wilcoxon Signed Rank

Acknowledgements

We wish to express our gratitude and thanks to all of the many people that reviewed, tested, and gave some very helpful suggestions for the development of ProUCL 4.0. We wish to especially acknowledge: Nadine Adkins, Ray Bienert, Cheri L. Butler, Lucía Casabó, John P. Christopher, Marion Edison, Evan Englund, Tim Ehli, Kuen Huang-Farmer, Dolores Gardner, Philip E. Goodrum, Scott A. Grubisich, Dennis R. Helsel, Jennifer Hubbard, Richard O. Gilbert, Sarah Levinson, Jayne Michaud, Jeff Myers, Yassine Nachti,Gareth Pearson, Marian Olsen, Kristina Rayner, Nancy Rios-Jafolla, Shahrokh Rouhani, Peter Rousseeuw, Dan Sahagun, Terry W. Schulz, Mike Schum, Tom Siard, Christopher Sibert, Bob Stewart, Rick Sugatt, J. Keith Tolson, Robert Tucker, John Warren, and Edward P. Wosika.

Table of Contents

Notice…….....iii

Changes from ProUCL 4.0 (4.00.00) to ProUCL 4.00.02.....v

Changes from ProUCL 4.00.02 to ProUCL 4.00.05.....v

Contact Information for ProUCL 4.00.02.....vi

Executive Summary.....vii

Acronyms and Abbreviations.....xiii

Acknowledgements…….....xv

Introduction.....1
The Need for ProUCL Software.....1
ProUCL 4.0 Capabilities.....2
ProUCL Applications.....4
ProUCL Methods.....5
Background versus Site Comparison Evaluations.....7
Graphical Capabilities.....11
Technical Guide.....12
Minimum Hardware Requirements.....12
Software Requirements.....12
Installation Instructions.....13
Getting Started.....13

Chapter 1 Guidance on the Use of Statistical Methods and Associated Minimum Sample Size Requirements.....15
1.1 Background Data Sets.....16
1.2 Site Data Sets.....17
1.3 Discrete Samples or Composite Samples?.....18
1.4 Upper Limits and Their Use.....18
1.5 Point-by-Point Comparison of Site Observations with BTVs, Compliance Limits, and Other Threshold Values.....21
1.6 Hypothesis Testing Approaches and Their Use.....23
1.6.1 Single Sample Hypotheses – BTVs and Not-to-Exceed Values are Known (Pre-established).....23
1.6.2 Two-sample Hypotheses – When BTVs and Not-to-Exceed Values are Unknown.....24
1.7 Minimum Sample Size Requirements.....25
1.7.1 Minimum Sample Size Requirements to Compute Decision Statistics for Data without Non-detects.....26
1.7.2 Minimum Sample Size Requirements for Datasets with Non-detects.....27
1.7.3 Minimum Sample Size for Estimation and Point-by-Point Site Observation Comparisons.....27
1.7.4 Minimum Sample Size Requirements for Hypothesis Testing Approaches.....28

1.8 Sample Sizes for Bootstrap Methods 28
1.9 Statistical Analyses by a Group ID 28
1.10 Use of Maximum Detected Values as Estimates of Upper Limits 29
1.10.1 Use of Maximum Detected Values to Estimate BTVs and Not-to-Exceed Values 29
1.10.2 Use of Maximum Detected Values to Estimate EPC Terms 29
1.10.3 Samples with Nondetect Observations 31
1.10.4 Avoid the Use of DL/2 Method to Compute UCL95 31
1.10.5 Samples with Low Frequency of Detection 31
1.11 Other Applications of Methods in ProUCL 4.0 32
1.11.1 Identification of COPCs 32
1.11.2 Identification of Non-Compliance Monitoring Wells 33
1.11.3 Verification of the Attainment of Cleanup Standards, C_s 33
1.11.4 Using BTVs (Upper Limits) to identify Hot Spots 33
1.12 Some General Issues and Recommendations Made by ProUCL 34
1.12.1 Multiple Detection Limits 34
1.12.2 Field Duplicates 34
1.12.3 ProUCL Recommendations about ROS Method and Substitution (DL/2) Method 34

Chapter 2 Entering and Manipulating Data 35
2.1 Creating a New Data Set 35
2.2 Opening an Existing Data Set 35
2.3 Input File Format 36
2.4 Number Precision 36
2.5 Entering and Changing a Header Name 37
2.6 Saving Files 38
2.7 Editing 38
2.8 Handling Nondetect Observations 39
2.9 Caution 40
2.10 Summary Statistics for Data Sets with Nondetect Observations 41
2.11 Warning Messages and Recommendations for Datasets with Insufficient Amount of Data 40
2.12 Handling Missing Values 44
2.13 User Graphic Display Modification 46
2.13.1 Graphics Tool Bar 44
2.13.2 Drop-Down Menu Graphics Tools 44

Chapter 3 Select Variables Screen 49
3.1 Select Variables Screen 49
3.1.1 Graphs by Groups 48

Chapter 4 Summary Statistics 53
4.1 Summary Statistics with Full Data Sets 53
4.2 Summary Statistics with NDs 55

Chapter 5 Estimating Nondetects Using ROS Methods 59

Chapter 6 Graphical Methods (Graph) 61

6.1 Box Plot 62
6.2 Histogram 64
6.3 Multi-QQ 66
6.3.1 Multi-QQ (Full) 63
6.3.2 Multi-QQ (with NDs) 64

Chapter 7 Simple Classical Outlier Tests 71
7.1 Outlier Test for Full Data Set 73
7.2 Outlier Test for Data Set with NDs 73

Chapter 8 Goodness-of-Fit (G.O.F.) Tests 77
8.1 ROS Estimated (Est.) NDs – Saving Extrapolated NDs 80
8.2 Goodness-of-Fit Tests with Full Data Sets 80
8.2.1 GOF Tests for Normal and Lognormal Distribution 81
8.2.2 GOF Tests for Gamma Distribution 84
8.3 Goodness-of-Fit Tests Excluding NDs 86
8.3.1 Normal and Lognormal Options 87
8.3.2 Gamma Distribution Option 90
8.4 Goodness-of-Fit Tests with Log-ROS Estimates 92
8.4.1 Normal or Lognormal Distribution (Log-ROS Estimates) 93
8.4.2 Gamma Distribution (Log-ROS Estimates) 96
8.5 Goodness-of-Fit Tests with DL/2 Estimates 98
8.5.1 Normal or Lognormal Distribution (DL/2 Estimates) 99
8.5.2 Gamma Distribution (DL/2 Estimates) 102
8.6 Goodness-of-Fit Tests Statistics 104

Chapter 9 Single Sample and Two-Sample Hypotheses Testing Approaches 107
9.1 Single Sample Hypotheses Tests 107
9.1.1 Single Sample Hypothesis Testing for Full Data without Nondetects 105
9.1.1.1 Single Sample t-Test 108
9.1.1.2 Single Sample Proportion Test 110
9.1.1.3 Single Sample Sign Test 112
9.1.1.4 Single Sample Wilcoxon Signed Rank (WSR) Test 111
9.1.2 Single Sample Hypothesis Testing for Data Sets with Nondetects 114
9.1.2.1 Single Proportion Test on Data Sets with NDs 114
9.1.2.2 Single Sample Sign Test with NDs 119
9.1.2.3 Single Sample Wilcoxon Signed Rank Test with NDs 121
9.2 Two-Sample Hypotheses Testing Approaches 124
9.2.1 Two-Sample Hypothesis Tests for Full Data 122
9.2.1.1 Two-Sample t-Test without NDs 124
9.2.1.2 Two-Sample Wilcoxon-Mann-Whitney (WMW) Test without NDs 126
9.2.1.3 Two-Sample Quantile Test for Full Data without NDs 129
9.2.2 Two-Sample Hypothesis Testing for Data Sets without Nondetects 132
9.2.2.1 Two-Sample Wilcoxon-Mann-Whitney Test with Nondetects 132
9.2.2.2 Two-Sample Gehan Test for Data Sets with Nondetects 139
9.2.2.3 Two-Sample Quantile Test for Data Sets with Nondetects 138

Chapter 10 Background Statistics 145
10.1 Background Statistics for Full Data Sets without Nondetects 146
10.1.1 Normal or Lognormal Distribution 143
10.1.2 Gamma Distribution 150
10.1.3 Nonparametric Methods 152
10.1.4 All Statistics Option 154
10.2 Background Statistics with NDs 156
10.2.1 Normal or Lognormal Distribution 156
10.2.2 Gamma Distribution 161
10.2.3 Nonparametric Methods (with NDs) 163
10.2.4 All Statistics Option 165

Chapter 11 Computing Upper Confidence Limits (UCLs) of Mean 169
11.1 UCLs for Full Data Sets 170
11.1.1 Normal Distribution (Full Data Sets without NDs) 170
11.1.2 Gamma, Lognormal, Nonparametric, All Statistics Option (Full Data without NDs) 173
11.2 UCL for Data Sets with NDs 178

Chapter 12 Sample Sizes for User Specified Performance Measures 189
12.1 Estimation of Mean 189
12.2 Sample Sizes for Single Sample Hypothesis Tests 190
12.2.1 Sample Size for Single Sample t-Test 170
12.2.2 Sample Size for Single Sample Proportion Test 1732
12.2.3 Sample Size for Single Sample Sign Test 1733
12.2.4 Sample Size for Single Sample Wilcoxon Signed Rank Test 1734
12.3 Sample Sizes for Two Sample Hypothesis Tests 196
12.3.1 Sample Size for Two Sample t-Test 196
12.3.2 Sample Size for Two Sample Wilcoxon Mann-Whitney 1737
12.4 Sample Sizes for Acceptance Sampling 17899

Chapter 13 Windows 201

Chapter 14 Help 203

Chapter 15 Handling the Output Screens and Graphs 205
Copying Graphs 205
Printing Graphs 206
Printing Non-graphical Outputs 207
Saving Output Screens as Excel Files 208

Chapter 16 Recommendations to Compute a 95% UCL (Estimate of EPC Term) of the Population Mean, μ_1, Using Symmetric and Positively Skewed Full Data Set without any Nondetects 211
16.1 Normally or Approximately Normally Distributed Data Sets 211
16.2 Gamma Distributed Skewed Data Sets 212
16.3 Lognormally Distributed Skewed Data Sets 213
16.4 Data Sets without a Discernable Skewed Distribution – Nonparametric Methods for Skewed Data Sets 215

16.5 Should the Maximum Observed Concentration be Used as an Estimate of the EPC Term? 220

Chapter 17 Recommendations to Compute a 95% UCL of the Population Mean, μ_1, Using Data Sets with Nondetects with Multiple Detection Limits 209
17.1 General Recommendations and Suggestions 221
17.2 Recommended UCL95 Methods for Normal (Approximate Normal) Distribution 223
17.3 Recommended UCL95 Methods for Gamma Distribution 224
17.4 Recommended UCL95 Methods for Lognormal Distribution 224
17.5 Recommended Nonparametric UCL Methods 225

Glossary 229

References 231

Introduction

The Need for ProUCL Software

Statistical inferences about the sampled populations and their parameters are made based upon defensible and representative data sets of appropriate sizes collected from the populations under investigation. Statistical inference, including both estimation and hypotheses testing approaches, is routinely used to:

1. Estimate environmental parameters of interest such as exposure point concentration (EPC) terms, not-to-exceed values, and background level threshold values (BTVs) for contaminants of potential concern (COPC),
2. Identify areas of concern (AOC) at a contaminated site,
3. Compare contaminant concentrations found at two or more AOCs of a contaminated site,
4. Compare contaminant concentrations found at an AOC with background or reference area contaminant concentrations,
5. Compare site concentrations with a cleanup standard to verify the attainment of cleanup standards.

Statistical inference about the sampled populations and their parameters are made based upon defensible and representative data sets of appropriate sizes collected from the populations under investigation. Environmental data sets originated from the Superfund and RCRA sites often consist of observations below one or more detection limits (DLs). In order to address the statistical issues arising in: exposure and risk assessment applications; background versus site comparison and evaluation studies; and various other environmental applications, several graphical, parametric, and nonparametric statistical methods for data sets with nondetects and without nondetects have been incorporated in ProUCL 4.0.

Exposure and risk management and cleanup decisions in support of United States Environmental Protection Agency (EPA) projects are often made based upon the mean concentrations of the COPCs. A 95% upper confidence limit (UCL95) of the unknown population (e.g., an AOC) arithmetic mean (AM), μ_1, can be used to:

- Estimate the EPC term of the AOC under investigation,
- Determine the attainment of cleanup standards,
- Compare site mean concentrations with reference area mean concentrations, and
- Estimate background level mean contaminant concentrations. The background mean contaminant concentration level may be used to compare the mean of an AOC. It should be noted that it is not appropriate to compare individual point-by-point site observations with the background mean concentration level.

It is important to compute a reliable and stable *UCL95* of the population mean using the available data. The *UCL95* should approximately provide the 95% coverage for the unknown population mean, μ_1. Based upon the available background data, it is equally important to compute reliable and stable upper percentiles, upper prediction limits (*UPLs*), or upper tolerance limits (*UTLs*). These upper limits based upon background (or reference) data are used as estimates of BTVs, compliance limits (*CL*), or not-to-exceed values. These upper limits are often used in site (point-by-point) versus background comparison evaluations.

Environmental scientists often encounter trace level concentrations of COPCs when evaluating sample

analytical results. Those low level analytical results cannot be measured accurately, and therefore are typically reported as less than one or more detection limit (*DL*) values (also called nondetects). However, practitioners often need to obtain reliable estimates of the population mean, μ_1, the population standard deviation, σ_1, and upper limits, including the upper confidence limit (*UCL)* of the population mass or mean, the *UPL*, and the *UTL* based upon data sets with nondetect (*ND*) observations. Hypotheses testing approaches are often used to verify the attainment of cleanup standards, and compare site and background concentrations of COPCs.

Background evaluation studies, BTVs, and not-to-exceed values should be estimated based upon defensible background data sets. The estimated BTVs or not-to-exceed values are then used to identify the COPCs, to identify the site AOCs or hot spots, and to compare the contaminant concentrations at a site with background concentrations. The use of appropriate statistical methods and limits for site versus background comparisons is based upon the following factors:

1. Objective of the study,
2. Environmental medium (e.g., soil, groundwater, sediment, air) of concern,
3. Quantity and quality of the available data,
4. Estimation of a not-to-exceed value or of a mean contaminant concentration,
5. Pre-established or unknown cleanup standards and BTVs, and
6. Sampling distributions (parametric or nonparametric) of the concentration data sets collected from the site and background areas under investigation.

In background versus site comparison evaluations, the environmental population parameters of interest may include:

- Preliminary remediation goals (PRGs),
- Soil screening levels (SSLs),
- Risk-based cleanup (RBC) standards,
- BTVs, not-to-exceed values, and
- Compliance limit, maximum concentration limit (MCL), or alternative concentration limit (ACL), frequently used in groundwater applications.

When the environmental parameters listed above are not known or have not been pre-established, appropriate upper statistical limits are used to estimate the parameters. The UPL, UTL, and upper percentiles are used to estimate the BTVs and not-to-exceed values. Depending upon the site data availability, point-by-point site observations are compared with the estimated (or pre-established) BTVs and not-to-exceed values. If enough site and background data are available, two-sample hypotheses testing approaches are used to compare site concentrations with background concentrations levels. These statistical methods can also be used to compare contaminant concentrations of two site AOCs, surface and subsurface contaminant concentrations, or upgradient versus monitoring well contaminant concentrations.

ProUCL 4.00.05 Capabilities

ProUCL Version 4.00.05 is an upgrade of ProUCL Version 4.0 (EPA, 2007). ProUCL 4.00.02 contains statistical methods to address various environmental issues for both full data sets without nondetects and for data sets with NDs (also known as left-censored data sets).

ProUCL 4.00.02/ProUCL 4.0 contains:

- Rigorous parametric and nonparametric (including bootstrap methods) statistical methods (instead of simple ad hoc or substitution methods) that can be used on full data sets without nondetects and on data sets with below detection limit (BDL) or nondetect (ND) observations.

- State-of-the-art parametric and nonparametric UCL, UPL, and UTL computation methods. These methods can be used on full-uncensored data sets without nondetects and also on data sets with BDL observations. Some of the methods (e.g., Kaplan-Meier method, ROS methods) are applicable on left-censored data sets having multiple detection limits. The UCL and other upper limit computation methods cover a wide range of skewed data sets with and without the BDLs.

- Single sample (e.g., Student's t-test, sign test, proportion test, Wilcoxon Singed Rank test) and two-sample (Student's t-test, Wilcoxon-Mann-Whitney test, Gehan test, quantile test) parametric and nonparametric hypotheses testing approaches for data sets with and without ND observations. These hypothesis testing approaches can be used to: verify the attainment of cleanup standards, perform site versus background comparisons, and compare two or more AOCs, monitoring wells (MWs).

- The single sample hypotheses testing approaches are used to compare site mean, site median, site proportion, or a site percentile (e.g., 95th) to a compliance limit (action level, regularity limit). The hypotheses testing approaches can handle both full-uncensored data sets without nondetects, and left-censored data sets with nondetects. Simple two-sample hypotheses testing methods to compare two populations are available in ProUCL 4.0, such as two-sample t-tests, Wilcoxon-Mann-Whitney (WMW) Rank Sum test, quantile test, Gehan's test, and dispersion test. Variations of hypothesis testing methods (e.g., Levene's method to compare dispersions, generalized WRS test) are easily available in most commercial and freely available software packages (e.g., MINITAB, R).

- ProUCL 4.0 includes graphical methods (e.g., box plots, multiple Q-Q plots, histogram) to compare two or more populations. Additionally, ProUCL 4.0 can also be used to display a box plot of one population (e.g., site data) with compliance limits or upper limits (e.g., UPL) of other population (background area) superimposed on the same graph. This kind of graph provides a useful visual comparison of site data with a compliance limit or BTVs. Graphical displays of a data set (e.g., Q-Q plot) should be used to gain insight knowledge contained in a data set that may not otherwise be clear by looking at simple test statistics such as t-test, Dixon test statistic, or Shapiro-Wilk (S-W) test statistic.

6. ProUCL 4.0 can process multiple contaminants (variables) simultaneously and has the capability of processing data by groups. A valid group column should be included in the data file.

7. ProUCL 4.0 provides a GOF test for data sets with nondetects. The user can create additional columns to store extrapolated (estimated) values for nondetects based upon normal ROS, gamma ROS, and lognormal ROS (robust ROS) methods.

ProUCL Applications

The methods incorporated in ProUCL 4.00.05 (and in all previous versions) can be used on data sets with and without BDL and ND observations. Methods and recommendations as incorporated in ProUCL 4.0 are based upon the results and findings of the extensive simulation studies as summarized in Singh and Singh (2003), and Singh, Maichle, and Lee (EPA, 2006). It is anticipated that ProUCL 4.00.05 will serve as a companion software package for the following EPA documents:

- *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites* (EPA, 2002a), and
- The revised *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b).

Methods included in ProUCL 4.00.05 can be used in various other environmental applications including the verification of cleanup standards (EPA, 1989), and computation of upper limits needed in groundwater monitoring applications (EPA, 1992 and EPA, 2004).

In 2002, EPA issued guidance for calculating the UCLs of the unknown population means for contaminant concentrations at hazardous waste sites. The ProUCL 4.0 software package (EPA, 2007) has served as a companion software package for the EPA (2002a) guidance document for calculating UCLs of mean contaminant concentrations at hazardous waste sites. ProUCL 4.0 has several parametric and nonparametric statistical methods that can be used to compute appropriate UCLs based upon full-uncensored data sets without any ND observations. ProUCL 4.0 retains the capabilities of ProUCL 3.0, including goodness-of-fit (GOF) and the UCL computation methods for data sets without any BDL observations. However, ProUCL 4.0 has the additional capability to perform GOF tests and computing UCLs and other upper limits based upon data sets with BDL observations.

ProUCL 4.0 defines log-transform (*log*) as the natural logarithm (*ln*) to the base e. ProUCL 4.0 also computes the maximum likelihood estimates (MLEs*)* and the minimum variance unbiased estimates (MVUEs) of unknown population parameters of normal, lognormal, and gamma distributions. This, of course, depends upon the underlying data distribution. ProUCL 4.0 computes the $(1 - \alpha)100\%$ UCLs of the unknown population mean, μ_1, using 5 parametric and 10 nonparametric methods. It should be pointed out that ProUCL 4.0 computes the <u>simple</u> summary statistics for detected raw and log-transformed data for full data sets without NDs, as well as for data sets with BDL observations. It is noted that estimates of mean and *sd* for data sets with NDs based upon rigorous statistical methods (e.g., MLE, ROS, K-M methods) are note provided in the summary statistics. Those estimates and the associated upper limits for data sets with NDs are provided under the menu options: Background and UCL.

It is emphasized that throughout this User Guide, and in the ProUCL 4.00.05 (and in all previous versions of ProUCL) software, it is assumed that one is dealing with a single population. If multiple populations (e.g., background and site data mixed together) are present, it is recommended to first separate them out (e.g., using appropriate statistical population partitioning techniques), and then compute appropriate respective 95% UCLs separately for each of the identified populations. Outliers, if any, should be identified and thoroughly investigated. ProUCL 4.0 provides two commonly used simple classical outlier identification procedures: 1) the Dixon test and 2) the Rosner test. Outliers distort most parametric statistics (e.g., mean, UCLs, upper prediction limits (UPLs), test statistics) of interest. Moreover, it should be <u>noted</u> that even though outliers might have minimal influence on hypotheses testing statistics based upon ranks (e.g., WMW test), outliers do distort those nonparametric statistics (including bootstrap methods), which are based upon higher order statistics such as UPLs and UTLs. Decisions about the

disposition (exclusion or inclusion) of outliers in a data set used to estimate the EPC terms or BTVs should be made by all parties involved (e.g., project team, EPA, local agency, potentially responsible party, etc.) in the decision making process.

The presence of outlying observations also distorts statistics based upon bootstrap re-samples. The use of higher order values (quantiles) of the distorted statistics for the computation of the UCLs or UPLs based upon bootstrap t and Hall's bootstrap methods may yield unstable and erratic UCL values. This is especially true for the upper limits providing higher confidence coefficients such as 95%, 97.5%, or 99%. Similar behavior of the bootstrap t UCL is observed for data sets having BDL observations. Therefore, the bootstrap t and Hall's bootstrap methods should be used with caution. It is suggested that the user should examine various other UCL results and determine if the UCLs based upon the bootstrap t and Hall's bootstrap methods represent reasonable and reliable UCL values of practical merit. If the results based upon these two bootstrap methods are much higher than the rest of methods, then this could be an indication of erratic behavior of those bootstrap UCL values, perhaps distorted by outlying observations. In case these two bootstrap methods yield erratic and inflated UCLs, the UCL of the mean should be computed using the adjusted or the approximate gamma UCL computation method for highly skewed gamma distributed data sets of small sizes. Alternatively, one may use a 97.5% or 99% Chebyshev UCL to estimate the mean of a highly skewed population. It should be noted that typically, a Chebyshev UCL may yield conservative and higher values of the UCLs than other methods available in ProUCL 4.0 This is especially true when data are moderately skewed and sample size is large. In such cases, when the sample size is large, one may want to use a 95% Chebyshev UCL or a Chebyshev UCL with lower confidence coefficient such as 92.5% or 90% as estimate of the population mean.

ProUCL Methods

ProUCL 4.00.05 provides 15 UCL computation methods for full data sets without any BDL observations; 5 are parametric and 10 are nonparametric methods. The nonparametric methods do not depend upon any assumptions about the data distributions. The five parametric UCL computation methods are:

- Student's t-UCL,
- Approximate gamma UCL using chi-square approximation,
- Adjusted gamma UCL (adjusted for level significance),
- Land's H-UCL, and
- Chebyshev inequality-based UCL (using MVUEs of parameters of a lognormal distribution).

The 10 nonparametric methods are:

1. The central limit theorem (CLT)-based UCL,
2. Modified-t statistic (adjusted for skewness)-based UCL,
3. Adjusted-CLT (adjusted for skewness)-based UCL,
4. Chebyshev inequality-based UCL (using sample mean and sample standard deviation),
5. Jackknife method-based UCL,
6. UCL based upon standard bootstrap,
7. UCL based upon percentile bootstrap,
8. UCL based upon bias-corrected accelerated (BCA) bootstrap,
9. UCL based upon bootstrap t, and
10. UCL based upon Hall's bootstrap.

Environmental scientists often encounter trace level concentrations of COPCs when evaluating sample analytical results. Those low level analytical results cannot be measured accurately, and therefore are typically reported as less than one or more DL values. However, the practitioners need to obtain reliable estimates of the population mean, μ_1, and the population standard deviation, σ_1, and upper limits including the UCL of the population mass (measure of central tendency) or mean, UPL, and UTL. Several methods are available and cited in the environmental literature (Helsel (2005), Singh and Nocerino (2002), Millard and Neerchal (2001)) that can be used to estimate the population mean and variance. However, till to date, no specific recommendations are available for the use of appropriate methods that can be used to compute upper limits (e.g., UCLs, UPLs) based upon data sets with BDL observations. Singh, Maichle, and Lee (EPA, 2006) extensively studied the performance of several parametric and nonparametric UCL computation methods for data sets with BDL observations. Based upon their results and findings, several methods to compute upper limits (UCLs, UPLs, and UTLs) needed to estimate the EPC terms and BTVs have been incorporated in ProUCL 4.0.

In 2002, EPA issued another *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b). This EPA (2002b) background guidance document is currently being revised to include statistical methods that can be used to estimate the BTVs and not-to-exceed values based upon data sets with and without the BDL observations. In background evaluation studies, BTVs, compliance limits, or not-to-exceed values often need to be estimated based upon defensible background data sets. The estimated BTVs or not-to-exceed values are then used for screening the COPCs, to identify the site AOCs or hot spots, and also to determine if the site concentrations (perhaps after a remediation activity) are comparable to background concentrations, or are approaching the background level concentrations. Individual point-by-point site observations (composite samples preferred) are sometimes compared with those not-to-exceed values or BTVs. It should be pointed out that in practice, it is preferred to use hypotheses testing approaches to compare site versus background concentrations provided enough (e.g., at least 8-10 detected observations from each of the two populations) site and background data are available. Chapter 1 provides practical guidance on the minimum sample size requirements to estimate and use the BTVs, single and two-sample hypotheses testing approaches to perform background evaluations and background versus site comparisons. Chapter 1 also briefly discusses the differences in the definitions and uses of the various upper limits as incorporated in ProUCL 4.0. Detailed discussion of the various methods to estimate the BTVs and other not-to-exceed values for full-uncensored data sets (Chapter 5) without any nondetect values and for left-censored data sets (Chapter 6) with nondetect values are given in the revised background guidance document.

ProUCL 4.0 includes statistical methods to compute UCLs of the mean, upper limits to estimate the BTVs, other not-to-exceed values, and compliance limits based upon data sets with one or more detection limits. The use of appropriate statistical methods and limits for exposure and risk assessment, and site versus background comparisons, is based upon several factors:

1. Objective of the study;
2. Environmental medium (e.g., soil, groundwater, sediment, air) of concern;
3. Quantity and quality of the available data;
4. Estimation of a not-to-exceed value or of a mean contaminant concentration;
5. Pre-established or unknown cleanup standards and BTVs; and
6. Sampling distributions (parametric or nonparametric) of the concentration data sets collected from the site and background areas under investigation.

In background versus site comparison studies, the population parameters of interest are typically represented by *upper threshold limits* (e.g., upper percentiles, upper confidence limits of an upper

percentile, upper prediction limit) of the background data distribution. It should be noted that the upper threshold values are estimated and represented by upper percentiles and other values from the upper tail of the background data distribution. These background upper threshold values do not represent measures of central tendency such as the mean, the median, or their upper confidence limits. These environmental parameters may include:

- Preliminary remediation goals (PRGs), Compliance Limits,
- Soil screening levels (SSLs),
- Risk-based cleanup (RBC) standards,
- BTVs, compliance limits, or not-to-exceed values, and
- Maximum concentration limit (MCL) or alternative concentration limit (ACL) used in Groundwater applications.

When the environmental parameters listed above are not known or pre-established, appropriate upper statistical limits are used to estimate those parameters. The UPL, UTL, and upper percentiles are typically used to estimate the BTVs, not-to-exceed values, and other parameters listed above. Depending upon the availability of site data, point-by-point site observations are compared with the estimated (or pre-established) BTVs and not-to-exceed values. If enough site and background data are available, two-sample hypotheses testing approaches (preferred method to compare two populations) are used to compare site concentrations with background concentrations levels. The hypotheses testing methods can also be used to compare contaminant concentrations of two site AOCs, surface and subsurface contaminant concentrations, or upgradient versus monitoring well contaminant concentrations.

Background versus Site Comparison Evaluations

The following statistical limits have been incorporated in ProUCL 4.0 to assist in background versus site comparison evaluations:

Parametric Limits for Full-Uncensored Data Sets without Nondetect Observations

- UPL for a single observation (Normal, Lognormal) not belonging to the original data set
- UPL for next k (k is user specified) or k future observations (Normal, Lognormal)
- UTL, an upper confidence limit of a percentile (Normal, Lognormal)
- Upper percentiles (Normal, Lognormal, and Gamma)

Nonparametric Limits for Full-Uncensored Data Sets without Nondetect Observations

Nonparametric limits are typically based upon order statistics of a data set such as a background or a reference data set. Depending upon the size of the data set, higher order statistics (maximum, second largest, third largest, and so on) are used as these upper limits (e.g., UPLs, UTLs). The details of these methods with sample size requirements can be found in Chapter 5 of the revised *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b). It should be, noted that the following statistics might get distorted by the presence of outliers (if any) in the data set under study.

- UPL for a single observation not belonging to the original data set
- UTL, an upper confidence limit of a percentile
- Upper percentiles

- Upper limit based upon interquartile range (IQR)
- Upper limits based upon bootstrap methods

For data sets with BDL observations, the following parametric and nonparametric methods to compute the upper limits were studied and evaluated by Singh, Maichle, and Lee (EPA, 2006) via Monte Carlo Simulation Experiments. Depending upon the performances of those methods, only some of the methods have been incorporated in ProUCL 4.0. Methods (e.g., Delta method, DL method, uniform (0, DL) generation method) not included in ProUCL 4.0 do not perform well in comparison with other methods.

Note: *When the percentage of NDs in a data set is high (e.g., > 40%-50%), especially when multiple detection limits might be present, it is hard to reliably perform GOF tests (to determine data distribution) on those data sets with many NDs. The uncertainty associated with those GOF tests will be high, especially when the data sets are of small sizes (< 10-20). It should also be noted that the parametric MLE methods (e.g., for normal and lognormal distributions) often yield unstable estimates of mean and sd. This is especially true when the number of nondetects exceeds 40%-50%. In such situations, it is preferable to use nonparametric (e.g., KM method) methods to compute statistics of interest such as UCLs, UPLs, and UTLs. Nonparametric methods do not require any distributional assumptions about the data sets under investigation. Singh, Maichle, and Lee (EPA, 2006) also concluded that the performance of the KM estimation method is better (in terms of coverage probabilities) than various other parametric estimation (e.g., MLE, EM, ROS) methods.*

Parametric Methods to Compute Upper Limits for Data Sets with Nondetect Observations

- Simple substitution (proxy) methods (0, DL/2, DL)
- MLE method, often known as Cohen's MLE method – single detection limit
- Restricted MLE method – single detection limit – not in ProUCL 4.0
- Expectation maximization (EM) method – single detection limit – not in ProUCL 4.0
- EPA Delta log method – single detection limit – not in ProUCL 4.0
- Regression method on detected data and using slope and intercept of the OLS regression line as estimates of standard deviation, *sd,* and mean (not a recommended method)
- Robust ROS (regression on order statistics) on log-transformed data – nondetects extrapolated (estimated) using robust ROS; mean, *sd, UCLs,* and other statistics computed using the detected and extrapolated data in original scale – multiple detection limits
- Normal ROS – nondetects extrapolated (estimated) using normal distribution, mean, *sd, UCLs,* and other statistics computed using the detected and extrapolated data – multiple detection limits.
- It is noted that the estimated NDs often become negative and even larger than the detection limits (not a recommended method)
- Gamma ROS – nondetects extrapolated (estimated) using gamma distribution, mean, *sd, UCLs,* and other statistics computed using the detected and extrapolated data – multiple detection limits

Nonparametric Methods to Compute Upper Limits for Data Sets with Nondetect Observations

- Bootstrap Methods
  - Percentile Bootstrap on robust ROS
  - Percentile Bootstrap

- BCA Bootstrap
- Bootstrap t

- Jackknife Method
  - Jackknife on robust ROS

- Kaplan-Meier (KM) Method
  - Bootstrap (percentile, BCA) using KM estimates
  - Jackknife using KM estimates
  - Chebyshev Method using KM estimates

- Winsorization Method

For uncensored full data sets without any NDs, the performance (in terms of coverage for the mean) of the various UCL computation methods was evaluated by Singh and Singh (2003). The performance of the parametric and nonparametric UCL methods based upon data sets with nondetect observations was studied by Singh, Maichle, and Lee (EPA, 2006). Several of the methods listed above have been incorporated in ProUCL 4.0 to compute the estimates of EPC terms (95% UCL), and of BTVs (UPLs, UTLs, upper percentiles). Methods that did not perform well (e.g., poor coverage or unrealistically large values, infeasible and biased estimates) are not included in ProUCL 4.0. Methods not incorporated in ProUCL 4.0 are: EPA Delta Log method, Restricted MLE method, and EM method, substitution method (0, and DL), and Regression method.

Note: *It should be noted that for data sets with NDs, the DL/2 substitution method has been incorporated in ProUCL 4.0 only for historical reasons and also for its current default use. It is well known that the DL/2 method (with NDs replaced by DL/2) does not perform well (e.g., Singh, Maichle, and Lee (EPA, 2006)) even when the percentage of NDs is only 5%-10%. It is strongly suggested to avoid the use of DL/2 method for estimation and hypothesis testing approaches used in various environmental applications. Also, when the % of NDs becomes high (e.g., > 40%-50%), it is suggested to avoid the use of parametric MLE methods. For data sets with high percentage of NDs (e.g., > 40%), the distributional assumptions needed to use parametric methods are hard to verify; and those parametric MLE methods may yield unstable results.*

It should also be noted that even though the lognormal distribution and some statistics based upon lognormal assumption (e.g., Robust ROS, DL/2 method) are available in ProUCL 4.0, ProUCL 4.0 does not compute MLEs of mean and sd based upon a lognormal distribution. The main reason is that the estimates need to be computed in the original scale via back-transformation (Shaarawi, 1989, and Singh, Maichle, and Lee (EPA, 2006)). Those back-transformed estimates often suffer from an unknown amount of significant bias. Hence, it is also suggested to avoid the use of a lognormal distribution to compute MLEs of mean and sd, and associated upper limits, especially UCLs based upon those MLEs obtained using a lognormal distribution.

ProUCL 4.0 recommends the use of an appropriate UCL to estimate the EPC terms. It is desirable that the user consults with the project team and experts familiar with the site before using those recommendations. Furthermore, there does not seem to be a general agreement about the use of an upper limit (e.g., UPL, percentile, or UTL) to estimate not-to-exceed values or BTVs to be used for screening of the COPCs and in site versus background comparison studies. ProUCL 4.0 can compute both parametric and nonparametric upper percentiles, UPLs, and UTLs for uncensored and censored data sets. However, no specific recommendations have been made regarding the use of UPLs, UTLs, or upper percentiles to

estimate the BTVs, compliance limits, and other related background or reference parameters. However, the developers of ProUCL 4.0 prefer the use of UPLs or upper percentiles to estimate the background population parameters (e.g., BTVs, not-to-exceed values) that may be needed to perform point-by-point site versus background comparisons.

The standard bootstrap and the percentile bootstrap UCL computation methods do not perform well (do not provide adequate coverage to population mean) for skewed data sets. For skewed distributions, the bootstrap t and Hall's bootstrap (meant to adjust for skewness) methods do perform better (in terms of coverage for the population mean) than the other bootstrap methods. However, it has been noted (e.g., Efron and Tibshirani (1993), and Singh, Singh, and Iaci (2002b)) that these two bootstrap methods sometimes yield erratic and inflated *UCL* values (orders of magnitude higher than the other *UCLs*). This may occur when outliers are present in a data set. Similar behavior of the bootstrap t UCL is observed based upon data sets with NDs. Therefore, whenever applicable, ProUCL 4.0 provides cautionary statements regarding the use of bootstrap methods.

ProUCL 4.0 provides several state-of-the-art parametric and nonparametric UCL, UPL, and UTL computation methods that can be used on uncensored data sets (full data sets) and on data sets with BDL observations. Some of the methods (e.g., Kaplan-Meier method, ROS methods) incorporated in ProUCL 4.0 are applicable on left-censored data sets having multiple detection limits. The UCLs and other upper limits computation methods in ProUCL 4.0 cover a wide range of skewed data distributions with and without the BDLs arising from the environmental applications.

ProUCL 4.0 also has parametric and nonparametric single and two-sample hypotheses testing approaches required to: compare site location (e.g., mean, median) to a specified cleanup standard; perform site versus background comparisons; or compare of two or more AOCs. These hypotheses testing methods can handle both full (uncensored data sets without NDs) and left-censored (with nondetects) data sets. Specifically, two-sample tests such as t-test, Wilcoxon Mann-Whitney (WMW) Rank Sum test, quantile test, and Gehan's test are available in ProUCL 4.0 to compare concentrations of two populations.

Single sample parametric (Student's t-test) and nonparametric (sign test, Wilcoxon Signed Rank (WSR) test, tests for proportions and percentiles) hypotheses testing approaches are also available in ProUCL 4.0. The single sample hypotheses tests are useful when the environmental parameters such as the clean standard, action level, or compliance limits are known, and the objective is to compare site concentrations with those known threshold values. Specifically, a t-test (or a sign test) may be used to verify the attainment of cleanup levels at an AOC after a remediation activity; and a test for proportion may be used to verify if the proportion of exceedances of an action level (or a compliance limit) by sample concentrations collected from the AOC (or a MW) exceeds a certain specified proportion (e.g., 1%, 5%, 10%). As mentioned before, ProUCL 4.0 can perform these hypotheses on data sets with and without nondetect observations.

Note: *It should be noted that as cited in the literature, some of the hypotheses testing approaches (e.g., nonparametric two-sample WMW) deal with the single detection limit scenario. If multiple detection limits are present, all NDs below the largest detection limit need to be considered as NDs (Gilbert, 1987, and Helsel, 2005). This in turn may reduce the power and increase uncertainty associated with test. As mentioned before, it is always desirable to supplement the test statistics and test conclusions with graphical displays such as the multiple Q-Q plots and side-by-side box plots. ProUCL 4.0 can graph box plots and Q-Q plots for data sets with nondetect observations. Gehan test as available in ProUCL 4.0 should be used in case multiple detection limits are present. ProUCL 4.0 can draw Q-Q plots and box plots for data sets with and without nondetect observations.*

It should be pointed out that when using two-sample hypotheses approaches (WMW test, Gehan test, and quantile test) on data sets with NDs, both samples and variables (e.g., site-As, Back-As) should be specified as having nondetects. This means, a ND column (0 = ND, and 1 = detect) should be provided for each variable (here D_site-As, and D_Back-As) to be used in this comparison. If a variable (e.g., site-As) does not have any nondetects, still a column with label D_site-As should be included in the data set with all entries = 1 (detected values).

Moreover, in single sample hypotheses tests (e.g., sign test, proportion test) used to compare site mean/median concentration level with a cleanup standard, C_s, or compliance limit (e.g., proportion test), all NDs (if any) should lie below the cleanup standard, C_s.

The differences between these tests should be noted and understood. Specifically, a t-test or a Wilcoxon Signed Rank (WSR) test are used to compare the measures of location and central tendencies (e.g., mean, median) of a site area (e.g., AOC) to a cleanup standard, C_s, or action level also representing a measure of central tendency (e.g., mean, median); whereas, a proportion test compares if the proportion of site observations from an AOC exceeding a compliance limit (*CL*) exceeds a specified proportion, P_0 (e.g., 5%, 10%). The percentile test compares a specified percentile (e.g., 95th) of the site data to a pre-specified upper threshold (e.g., reporting limit, action level). All of these tests have been incorporated in ProUCL 4.0. Most of the single sample and two-sample hypotheses tests also report associated p-values. For some of the hypotheses tests (e.g., WMW test, WSR test, proportion test), large sample approximate p-values are computed using continuity correction factors.

Graphical Capabilities

ProUCL 4.0 has useful exploratory graphical methods that may be used to visually compare the concentrations of:

1. A site area of concern (AOC) with an action level. This can be done using a box plot of site data with action level superimposed on that graph,
2. Two or more populations, including site versus background populations, surface versus subsurface concentrations, and
3. Two or more AOCs.

The graphical methods include double and multiple quantile-quantile (Q-Q) plots, side-by-side box plots, and histograms. Whenever possible, it is desirable to supplement statistical test results and statistics with visual graphical displays of data sets. There is no substitute for graphical displays of a data set as the visual displays often provide useful information about a data set, which cannot be revealed by simple test statistics such as t-test, SW test, Rosner test, WMW test. For example, in addition to providing information about the data distribution, a normal Q-Q plot can also help identify outliers and multiple populations that might be present in a data set. This kind of information cannot be revealed by simple test statistics such as a Shapiro-Wilk (SW) test or Rosner's outlier test statistic. Specifically, the SW test may lead to the conclusion that a mixture data set (representing two or more populations) can be modeled by a normal (or lognormal) distribution, whereas the occurrence of obvious breaks and jumps in the associated Q-Q plot may suggest the presence of multiple populations in the mixture data set. It is suggested that the user should use exploratory tools to gain necessary insight into a data set and the underlying assumptions (e.g., distributional, single population) that may not be revealed by simple test statistics.

Note: *On a Q-Q plot, observations well separated from the majority of the data may represent potential outliers, and obvious jumps and breaks of significant magnitude may suggest the presence of observations from multiple populations in the data set.*

The analyses of data categorized by a group ID variable such as: 1) Surface vs. Subsurface; 2) AOC1 vs. AOC2; 3) Site vs. Background; and 4) Upgradient vs. Downgradient monitoring wells are quite common in many environmental applications. ProUCL 4.0 offers this option for data sets with and without nondetects. The Group Option provides a powerful tool to perform various statistical tests and methods (including graphical displays) separately for each of the group (samples from different populations) that may be present in a data set. For an example, the same data set may consist of samples from the various groups or populations representing site, background, two or more AOCs, surface, subsurface, monitoring wells. The graphical displays (e.g., box plots, Q-Q plots) and statistics (computations of background statistics, UCLs, hypotheses testing approaches) of interest can be computed separately for each group by using this option.

Technical Guide

In addition to this User Guide, a Technical document also accompanies ProUCL 4.0, providing useful technical details of the graphical and statistical methods as incorporated in ProUCL 4.0. Most of the mathematical algorithms and formulas (with references) used in the development of ProUCL 4.0 are summarized in the Technical Guide.

Minimum Hardware Requirements

- Intel Pentium 1.0 GHz
- 45 MB of hard drive space
- 512 MB of memory (RAM)
- CD-ROM drive
- Windows 98 or newer. ProUCL was thoroughly tested on NT-4, Windows 2000, and Windows XP Operating Systems (limited testing on Windows ME).

Software Requirements

ProUCL 4.00.05 has been developed in the Microsoft .NET Framework using the C# programming language. As such, to properly run ProUCL 4.00.05, the computer using the program must have the .NET Framework pre-installed. The downloadable .NET files can be found at one of the following two Web sites:

- http://msdn.microsoft.com/netframework/downloads/updates/default.aspx
  Note: *Download .Net version 1.1*

- http://www.microsoft.com/downloads/details.aspx?FamilyId=262D25E3-F589-4842-8157-034D1E7CF3A3&displaylang=en

The first Web site lists all of the downloadable .NET Framework files, while the second Web site provides information about the specific file (s) needed to run ProUCL 4.0. Download times are estimated at 57 minutes for a dialup connection (56K), and 13 minutes on a DSL/Cable connection (256K).

Installation Instructions

- Download the file SETUP.EXE from the EPA Web site and save to a temporary location. **Note:** This download is pending release – beta-testers: See text file on CD.

- Run the SETUP.EXE program. This will create a ProUCL directory and two folders: 1) the USER GUIDE (this document), and 2) DATA (example data sets).

- To run the program, use Windows Explorer to locate the ProUCL application file, and double click on it, or use the RUN command from the start menu to locate the ProUCL.exe file, and run ProUCL.exe.

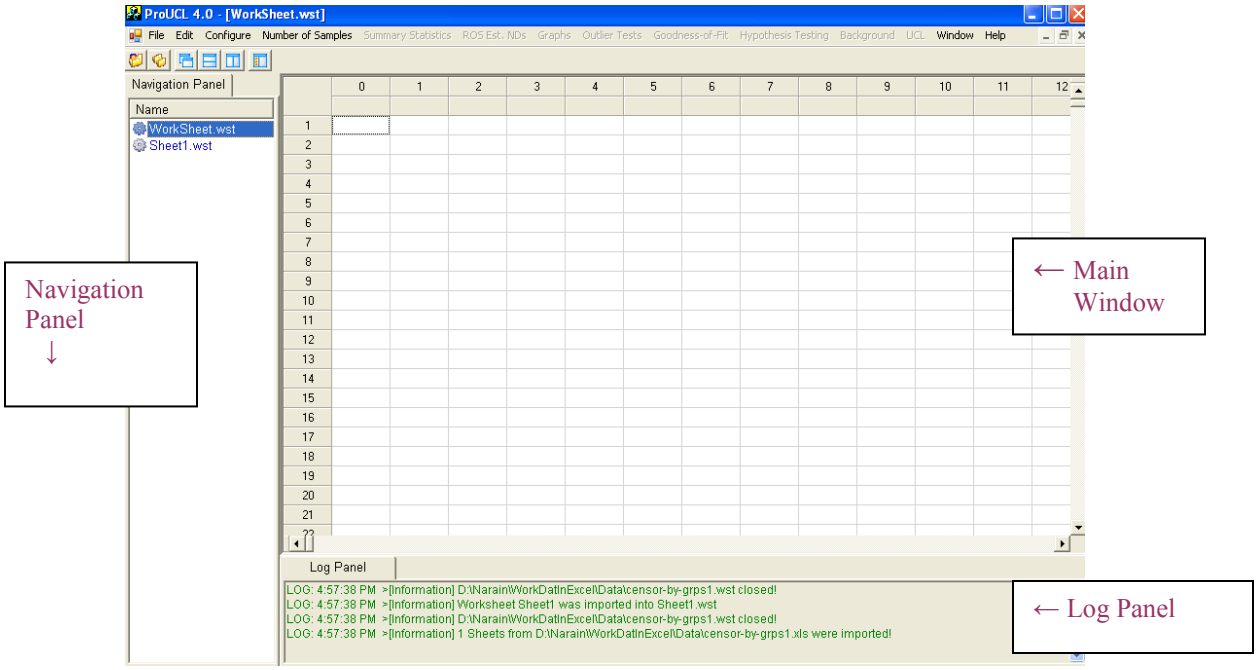
- To uninstall the program, use Windows Explorer to locate and delete the ProUCL folder.

Caution: If you have previous versions of the ProUCL, which were installed on your computer, you should remove or rename the directory in which earlier ProUCL versions are currently located.

Getting Started

The functionality and the use of the methods and options available in ProUCL 4.0 have been illustrated using Screen Shots of output screen generated by ProUCL 4.0. ProUCL 4.0 uses a pull-down menu structure, similar to a typical Windows program.

The screen below appears when the program is executed.



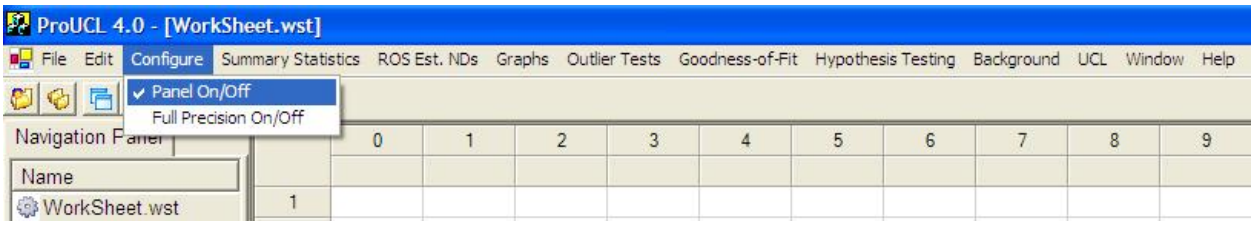
The screen consists of three main window panels:

- The **MAIN WINDOW** displays data sheets and outputs from the procedure used.

- The **NAVIGATION PANEL** displays the name of data sets and all generated outputs.
  - At present, the navigation panel can hold at most 20 outputs. In order to see more files (data files or generated output files), one can click on Widow Option.

- The **LOG PANEL** displays transactions in green, warnings in orange, and errors in red. For an example, when one attempts to run a procedure meant for censored data sets on a full-uncensored data set, ProUCL 4.0 will print out a warning message in orange in this panel.
  - Should both panels be unnecessary, you can click or choose **Configure ► Panel ON/OFF**.

The use of this option will give extra space to see and print out the statistics of interest. For an example, one may want to turn off these panels when multiple variables (e.g., multiple Q-Q plots) are analyzed and GOF statistics and other statistics may need to be captured for all of the variables.



Chapter 1

Guidance on the Use of Statistical Methods and Associated Minimum Sample Size Requirements

This chapter briefly describes the differences between the various statistical limits (e.g., UCLs, UPLs, UTLs) often used to estimate the environmental parameters of interest including exposure point concentration (EPC) terms and background threshold values (BTVs). Suggestions are provided about the minimum sample size requirements needed to use statistical inferential methods to estimate the environmental parameters: EPC terms, BTVs and not-to-exceed values, and to compare site data with background data or with some pre-established reference limits (e.g., preliminary remediation goals (PRGs), action levels, compliance limits). It is noted that several EPA guidance documents (e.g., EPA 1997, 2002a, 2006) discuss in details about data quality objectives (DQOs) and sample size determinations based upon those DQOs needed for the various statistical methods used in environmental applications.

Also, appropriate sample collection methods (e.g., instruments, sample weights, discrete or composite, analytical methods) depend upon the medium (e.g., soil, sediment, water) under consideration. For an example, Gerlach and Nocerino (EPA, 2003) describe optimal soil sample (based upon Gy theory) collection methods. Therefore, the topics of sample size determination based upon DQOs, data validation, and appropriate sample collection methods for the various environmental media are not considered in ProUCL 4.0 and its associated Technical Guide. It is assumed that data sets to be used in ProUCL are of good quality, and whenever possible have been obtained using the guidance provided in various EPA (2003, 2006) documents. It is the users' responsibility to assure that adequate amount of data have been collected, and the collected data are of good quality.

Note: *In ProUCL 4.0 and its associated guidance documents, emphasis is given on the practical applicability and appropriate use of statistical methods needed to address statistical issues arising in risk management, background versus site evaluation studies, and various other environmental applications. Specifically, guidance on minimum sample size requirements as provided in this chapter is useful when data have already been collected, or it is not possible (e.g., due to resource limitations) to collect the number of samples obtained using DQO processes as described in EPA (2006).*

Decisions based upon statistics obtained using data sets of small sizes (e.g., 4 to 6 detected observations) cannot be considered reliable enough to make a remediation decision that affects human health and the environment. For an example, a background data set of size 4 to 6 is not large enough to characterize background population, to compute BTV values, or to perform background versus site comparisons. In order to perform reliable and meaningful statistical inference (estimation and hypothesis testing), one should determine the sample sizes that need to be collected from the populations under investigation using appropriate DQO processes and decision error rates (EPA, 2006). However, in some cases, it may not be possible (e.g., resource constraints) to collect the same number of samples recommended by the DQO process. In order to address such cases, minimum sample size requirements for background and site data sets are described.

The use of an appropriate statistical method depends upon the environmental parameter(s) being estimated or compared. The measures of central tendency (e.g., means, medians, or their upper confidence limits (UCLs)) are often used to compare site mean concentrations (e.g., after remediation activity) with a

cleanup standard, C_s, representing some central tendency measure of a reference area or other known threshold representing a measure of central tendency. The upper threshold values, such as the compliance limits (e.g., alternative concentration limit (ACL), maximum concentration limit (MCL)), or not-to-exceed values, are used when individual point-by-point observations are compared with those not-to-exceed values or other compliance limit. It should be noted that depending upon whether the environmental parameters (e.g., BTVs, not-to-exceed value, EPC term, cleanup standards) are known or unknown, different statistical methods with different data requirements are needed to compare site concentrations with pre-established (known) or estimated (unknown) cleanup standards and BTVs.

ProUCL 4.0 has been developed to address issues arising in exposure assessment, risk assessment, and background versus site comparison applications. Several upper limits, and single- and two-sample hypotheses testing approaches, for both full uncensored and left-censored data sets, are available in ProUCL 4.0. The details of the statistical and graphical methods included in ProUCL 4.0 can be found in the ProUCL Technical Guidance. In order to make sure that the methods in ProUCL 4.0 are properly used, this chapter provides guidance on:

1. analysis of site and background areas and data sets,
2. collection of discrete or composite samples,
3. appropriate use of the various upper limits,
4. guidance regarding minimum sample sizes,
5. point-by-point comparison of site observations with BTVs,
6. use of hypotheses testing approaches,
7. using small data sample sets,
8. use of maximum detected value, and
9. discussion of ProUCL usage for special cases.

1.1 Background Data Sets

The project team familiar with the site should identify and chose a background area. Depending upon the site activities and the pollutants, the background area can be site-specific or a general reference area. An appropriate random sample of independent observations should be collected from the background area. A defensible background data set should represent a “single” background population (e.g., representing pristine site conditions before any of the industrial site activities) free of contaminating observations such as outliers. In a background data set, outliers may represent potentially contaminated observations from impacted site areas under study or possibly from other polluted site(s). This scenario is common when background samples are obtained from the various onsite areas (e.g., large federal facilities). Outlying observations should not be included in the estimation (or hypotheses testing procedures) of the BTVs. The presence of outliers in the background data set will yield distorted estimates of the BTVs and hypothesis testing statistics. The proper disposition of outliers to include or not include them in the data set should be decided by the project team.

Decisions based upon distorted statistics can be incorrect, misleading, and expensive. It should be noted that the objective is to compute background statistics based upon the majority of the data set representing the dominant background population, and not to accommodate a few low probability outliers that may also be present in the background data set. A couple of simple classical outlier tests (Dixon and Rosner tests) are available in ProUCL 4.0. Since these classical tests suffer from masking effects (e.g., extreme outliers may mask the occurrence of other intermediate outliers), it is suggested that these classical outlier tests should always be supplemented with graphical displays such as a box plot or a Q-Q plot. The use of

robust and resistant outlier identification procedures (Singh and Nocerino, 1995, Rousseeuw and Leroy, 1987) is recommended when multiple outliers may be present in a data set. Those methods are beyond the scope of ProUCL 4.0.

An appropriate background data set of a reasonable size (preferably computed using DQO processes) is needed to characterize a background area including computation of upper limits (e.g., estimates of BTVs, not-to-exceed values) based upon background data sets and also to compare site and background data sets using hypotheses testing approaches. As mentioned before, a small background data set of size 4 to 6 is not large enough to compute BTVs or to perform background versus site comparisons. At the minimum, a background sample should have at least 8 to 10 (more observations are preferable) detected observations to estimate BTVs or to use hypotheses testing approaches.

1.2 Site Data Sets

A defensible data set from a site population (e.g., AOC, EA, RU, group of monitoring wells) should be representative of the site area under investigation. Depending upon the site areas under investigation, different soil depths and soil types may be considered as representing different statistical populations. In such cases, background-versus-site comparisons may have to be conducted separately for each of those site sub-populations (e.g., surface and sub-surface layers of an AOC, clay and sandy site areas). These issues, such as comparing depths and soil types, should also be considered in a planning and sampling design before starting to collect samples from the various site areas under investigation. Specifically, the availability of an adequate amount of representative site data is required from each of those site sub-populations defined by sample depths, soil types, and the various other characteristics. For detailed guidance on soil sample collections, the reader is referred to Gerlach and Nocerino (EPA (2003)).

The site data collection requirements depend upon the objective of the study. Specifically, in background-versus-site comparisons, site data are needed to perform:

- Individual point-by-point site observation comparisons with pre-established or estimated BTVs, PRGs, cleanup standards, and not-to-exceed-values. Typically, this approach is used when only a small number (e.g., < 4 to 6) of detected site observations (preferably based upon composite samples) are available which need to be compared with BTVs and not-to-exceed values.

- Single sample hypotheses tests to compare site data with pre-established cleanup standards, C_s (e.g., representing a measure of central tendency); or with BTVs and not-to-exceed values (used for tests for proportions and percentiles). The hypotheses testing approaches are used when enough site data are available. Specifically, when at least 8 to 10 detected (more are desirable) site observations are available, it is preferable to use hypotheses testing approaches to compare site observations with specified threshold values. The use of hypotheses testing approaches can control the two types (Type 1 and Type 2) of error rates more efficiently than the point-by-point individual observation comparisons. This is especially true as the number of point-by-point comparisons increases. This issue is illustrated by the following table summarizing the probabilities of exceedances (false positive error rate) of the background threshold value (e.g., 95th percentile) by site observations, even when the site and background populations have comparable distributions. The probabilities of these chance exceedances increase as the sample size increases.

Sample Size	Probability of Exceedance
1	0.05
2	0.10
5	0.23
8	0.34
10	0.40
12	0.46
64	0.96

- Two-sample hypotheses testing to compare site data distribution with background data distribution to determine if the site concentrations are comparable to background concentrations. Adequate amount of data need to be made available from the site as well as the background populations. It is preferable to collect at least 8 to 10 detected observations from each of the population under comparison.

1.3 Discrete Samples or Composite Samples?

In a data set (background or site), collected samples should be either all discrete or all composite. In general, both discrete and composite site samples may be used for individual point-by-point site comparisons with a threshold value, and for single and two-sample hypotheses testing applications.

- If possible, the use of composite site samples is preferred when comparing individual point-by-point site observations from an area (e.g., area of concern (AOC), remediation unit (RU), exposure area (EA)) with pre-established or estimated BTV, compliance limit (CL), or other not-to-exceed value. This comparison approach is useful when few (< 4 to 6) detected site observations are compared with a pre-established or estimated BTV or other not-to-exceed threshold.

- When using a single sample hypothesis testing approach, site data can be obtained by collecting all discrete or all composite samples. The hypothesis testing approach is used when many (e.g., exceeding 8 to 10) detected site observations are available. Details of the single sample hypothesis approaches are widely available in EPA documents (1989, 1997, and 2006). Selected single sample hypotheses testing procedures are available in ProUCL 4.0.

- If a two-sample hypotheses testing approach is used to perform site versus background comparisons, then samples from both of the populations should be either all discrete or all composite samples. The two-sample hypothesis testing approach is used when many (e.g., exceeding 8 to 10) site, as well as background, observations are available. For better and more accurate results with higher statistical power, the availability of more observations (e.g., exceeding 10-15) from each of the two populations is desirable, perhaps based upon an appropriate DQO process, as described in an EPA guidance document (2006).

1.4 Upper Limits and Their Use

The appropriate computation and use of statistical limits depend upon their applications and the parameters (e.g., EPC term, not-to-exceed value) they are supposed to be estimating. Depending upon the objective of the study, a pre-specified cleanup standard, $C_{s,}$ or a risk-based cleanup (RBC) can be viewed

as to represent: 1) as average contaminant concentration; or 2) a not-to-exceed upper threshold value. These two threshold values, an average value, μ_0, and a not-to-exceed value, A_0, represent two significantly different parameters, and different statistical methods and limits are used to compare the site data with these two different parameters or threshold values. Statistical limits, such as an upper confidence limit (UCL) of the population mean, an upper prediction limit (UPL) for an independently obtained "single" observation, or independently obtained k observations (also called future k observations, next k observations, or k different observations), upper percentiles, and upper tolerance limits (UTLs), are often used to estimate the environmental parameters, including the EPC terms, compliance limits (e.g., ACL, MLC), BTVs, and other not-to-exceed values. Here, UTL95%-95% represents a 95% confidence limit of the 95th percentile of the distribution of the contaminant under study.

It is important to understand and note the differences between the uses and numerical values of these statistical limits so that they can be properly used. Specifically, the differences between UCLs and UPLs (or upper percentiles), and UCLs and UTLs should be clearly understood and acknowledged. A UCL with a 95% confidence limit (UCL95) of the mean represents an estimate of the population mean (measure of the central tendency of a data distribution), whereas a UPL95, a UTL95%-95%, and an upper 95th percentile represent estimates of a threshold value in the upper tail of the data distribution. Therefore, a UCL95 should represent a smaller number than an upper percentile or an upper prediction limit. Also, since a UTL 95%-95% represents a 95% UCL of the upper 95th percentile, a UTL should be ≥ the corresponding UPL95 and the 95th upper percentile. Typically, it is expected that the numerical values of these limits should follow the order given as follows:

Sample Mean ≤ UCL95 of Mean ≤ Upper 95th Percentile ≤ UPL95 of a Single Observation ≤ UTL95%-95%

It should also be pointed out that as the sample size increases, a UCL95 of the mean approaches (converges to) the population mean, and a UPL95 approaches the 95th percentile. The differences among the various upper limits are further illustrated in Example 1-1 below. It should be noted that, in some cases, these limits might not follow the natural order described above. This is especially true when the upper limits are computed based upon a lognormal distribution (Singh, Singh, and Engelhardt, 1997). It is well known that a lognormal distribution-based H-UCL95 (Land's UCL95) often yields unstable and impractically large UCL values. An H-UCL95 often becomes larger than UPL95 and even larger than a UTL 95%-95%. This is especially true when dealing with skewed data sets of smaller sizes. Moreover, it should also be noted that in some cases, a H-UCL95 becomes smaller than the sample mean, especially when the data are mildly skewed to moderately skewed and the sample size is large (e.g., > 50, 100).

Example 1-1: Consider a simple site-specific background data set associated with a Superfund site. The data set (given in Appendix 5 of the revised *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b)) has several inorganic contaminants of potential concern, including aluminum, arsenic, chromium, iron, and lead. It is noted that iron concentrations follow a normal distribution. Upper limits for the iron data set are summarized in Table 1-1. It is noted that the upper limits do follow the order as described above.

Table 1-1. Computation of Upper Limits for Iron (Normally Distributed)

Mean	Median	Min	Max	UCL95	UPL95 for a Single Observation	UPL95 for 4 Observations	UTL95/95	95% Upper Percentile
9618	9615	3060	18700	11478	18145	21618	21149	17534

A 95% UCL (UCL95) of the mean is the most commonly used limit in environmental applications. For an example, a 95% UCL of mean is used as an estimate of the EPC. A UCL95 should not be used to estimate a background threshold value (a value in the upper tail of the background data distribution) to be compared with individual site observations. There are many instances in background evaluations and background versus site comparison studies, when it is not appropriate to use a 95% UCL. Specifically, when point-by-point site observations are to be compared with a BTV, then that BTV should be estimated (or represented) by a limit from the upper tail of the reference set (background) data distribution.

A brief discussion about the differences between the applications and uses of the various statistical limits is provided below. This will assist a typical user in determining which upper limit (e.g., UCL95 or UPL95) to use to estimate the parameter of interest (e.g., EPC or BTV).

- A UCL represents an average value that should be compared with a threshold value also representing an average value (pre-established or estimated), such as a mean cleanup standard, C_s. For an example, a site 95% UCL exceeding a cleanup value, C_s, may lead to the conclusion that the cleanup level, C_s, has not been attained by the site area under investigation. It should be noted that UCLs of means are typically computed based upon the site data set.

- When site averages (and not individual site observations) are compared with a threshold value (pre-determined or estimated), such as a PRG or a RBC, or with some other cleanup standard, C_s, then that threshold should represent an average value, and not a not-to-exceed threshold value for individual observation comparisons.

- A UCL represents a "collective" measure of central tendency, and it is not appropriate to compare individual site observations with a UCL. Depending upon data availability, single or two-sample hypotheses testing approaches are used to compare site averages: with a specified or pre-established cleanup standard (single sample hypothesis), or with the background population averages (two-sample hypothesis).

- A UPL, an upper percentile, or an UTL represents an upper limit to be used for point-by-point individual site observation comparisons. UPLs and UTLs are computed based upon background data sets, and individual site observations are compared with those limits. A site observation for a contaminant exceeding a background UTL or UPL may lead to the conclusion that the contaminant is a contaminant of potential concern (COPC) to be included in further risk evaluation and risk management studies.

- When individual point-by-point site observations are compared with a threshold value (pre-determined or estimated) of a background population or some other threshold and compliance limit value, such as a PRG, MLC, or ACL, then that threshold value should represent a not-to-exceed value. Such BTVs or not-to-exceed values are often estimated by a 95% UPL, UTL 95%-95%, or by an upper percentile. ProUCL 4.0 can be used to compute any of these upper limits based upon uncensored data sets as well as data sets with nondetect values.

- As the sample size increases, a UCL approaches the sample mean, and a UPL95 approaches the corresponding 95th upper percentile.

- It is pointed out that the developers of ProUCL 4.0 prefer the use of a 95% UPL (UPL95) as an estimate of BTV or a not-to-exceed value. As mentioned before, the option of comparing individual site observations with a BTV (specified or estimated) should be used when few (< 4 to 6) detected site observations (preferably composite values) are to be compared with a BTV.

- When enough (e.g., > 8 to 10) detected site observations are available, it is preferred to use hypotheses testing approaches. Specifically, single sample hypotheses testing (comparing site to a specified threshold) approaches should be used to perform site versus a known threshold comparison; and two-sample hypotheses testing (provided enough background data are also available) approaches should be used to perform site versus background comparison. Several parametric and nonparametric single and two-sample hypotheses testing approaches are available in ProUCL 4.0.

It is re-emphasized that only averages should be compared with the averages or UCLs, and individual site observations should be compared with UPLs, upper percentiles, or UTLs. For an example, the comparison of a 95% UCL of one population (e.g., site) with a 90% or 95% upper percentile of another population (e.g., background) cannot be considered fair and reasonable as these limits (e.g., UCL and UPL) estimate and represent different parameters. It is hard to justify comparing a UCL of one population with a UPL of the other population. Conclusions (e.g., site dirty or site clean) derived by comparing UCLs and UPLs, or UCLs and upper percentiles as suggested in Wyoming DEQ, Fact Sheet #24 (2005), cannot be considered fair and reliable. Specifically, the decision error rates associated with such comparisons can be significantly different from the specified (e.g., Type I error = 0.1, Type II error = 0.1) decision errors.

1.5 Point-by-Point Comparison of Site Observations with BTVs, Compliance Limits, and Other Threshold Values

Point-by-point observation comparison method is used when a small number (e.g., 4 to 6 locations) of detected site observations are compared with pre-established or estimated BTVs, screening levels, or preliminary remediation goals (PRGs). In this case, individual point-by-point site observations (preferably based upon composite samples from various site locations) are compared with estimated or pre-established background (e.g., USGS values) values, PRGs, or some other not-to-exceed value. Typically, a single exceedance of the BTV, PRG, or of a not-to-exceed value by a site (or from a monitoring well) observation may be considered as an indication of contamination at the site area under investigation. The conclusion of an exceedance by a site value is some times confirmed by re-sampling (taking a few more collocated samples) that site location (or a monitoring well) exhibiting contaminant concentration in excess of the BTV or PRG. If all collocated (or collected during the same time period) sample observations collected from the same site location (or well) exceed the PRG (or MLC) or a not-to-exceed value, then it may be concluded that the location (well) requires further investigation (e.g., continuing treatment and monitoring) and cleanup.

When BTV contaminant concentrations are not known or pre-established, one has to collect, obtain, or extract a data set of an appropriate size that can be considered as representative of the site related background. Statistical upper limits are computed using the data set thus obtained, which are used as estimates of BTVs and not-to-exceed values. It should be noted that in order to compute reasonably reliable and accurate estimates of BTVs and not-to-exceed values based upon a background (or reference) data set, enough background observations (minimum of 8 to 10) should be collected, perhaps using an appropriate DQO process as described in EPA (2006). Typically, background samples are collected from a comparable general reference area or site-specific areas that are known to be free of contamination due

to any of the site related activities. Several statistical limits can be used to estimate the BTVs based upon a defensible data set of an adequate size. A detailed description of the computation and estimation of BTVs is given in Chapter 5 (for uncensored data sets) and in Chapter 6 for data sets with nondetects of the revised background guidance document. Once again, the use of this point-by-point comparison method is recommended when not many (e.g., < 4 to 6) site observations are to be compared with estimated BTVs or PRGs. An exceedance of the estimated BTV by a site value may be considered as an indication of the existing or continuing contamination at the site.

Note: *When BTVs are not known, it is suggested that at least 8 to 10 (more are preferable) detected representative background observations be made available to compute reasonably reliable estimates of BTVs and other not-to-exceed values.*

The point-by-point comparison method is also useful when quick turnaround comparisons are required. Specifically, when the decisions have to be made in real time by a sampling or screening crew, or when few detected site samples are available, then individual point-by-point site concentrations are compared either with pre-established PRGs, cleanup goals and standards, or with estimated BTVs and not-to-exceed values. The crew can use these comparisons to make the following informative decisions:

1. Screen and identify the COPCs,
2. Identify the polluted site AOCs,
3. Continue or stop remediation or excavation at a site AOC or a RU, or
4. Move the cleanup apparatus and crew to the next AOC or RU.

During the screening phase, an exceedance of a compliance limit, action level, a BTV, or a PRG by site values for a contaminant may declare that contaminant as a COPC. Those COPCs are then included in future site remediation and risk management studies. During the remediation phase, an exceedance of the threshold value such as a compliance limit (CL) or a BTV by sample values collected from a site area (or a monitoring well (MW)) may declare that site area as a polluted AOC, or a hot spot requiring further sampling and cleanup. This comparison method can also be used to verify if the site concentrations (e.g., from the base or side walls of an excavated site area) are approaching or meeting PRG, BTV, or a cleanup standard after excavation has been conducted at that site area.

If a larger number of detected samples (e.g., greater than 8 to10) are available from the site locations representing the site area under investigation (e.g., RU, AOC, EA), then the use of hypotheses testing approaches (both single sample and two-sample) is preferred. The use of a hypothesis testing approach will control the error rates more tightly and efficiently than the individual point-by-point site observations versus BTV comparisons, especially when many site observations are compared with a BTV or a not-to-exceed value.

Note: *In background versus site comparison evaluations, scientists usually prefer the use of hypotheses testing approaches to point-by-point site observation comparisons with BTVs or not-to-exceed values. Hypotheses testing approaches require the availability of larger data sets from the populations under investigation. Both single sample (used when BTVs, not-to-exceed values, compliance limits, or cleanup standards are known and pre-established) and two-sample (used when BTVs and compliance limits are unknown) hypotheses testing approaches are available in ProUCL 4.0.*

1.6 Hypothesis Testing Approaches and Their Use

Both single sample and two-sample hypotheses testing approaches are used to make cleanup decisions at polluted sites, and also to compare contaminant concentrations of two (e.g., site versus background) or more (several monitoring wells (MWs)) populations. The uses of hypotheses testing approaches in those environmental applications are described as follows.

1.6.1 Single Sample Hypotheses – BTVs and Not-to-Exceed Values are Known (Pre-established)

When pre-established BTVs and not-to-exceed values are used, such as the USGS background values (Shacklette and Boerngen (1984)), thresholds obtained from similar sites, or pre-established not-to-exceed values, PRGs, or RBCs, there is no need to extract, establish, or collect a background or reference data set. When the BTVs and cleanup standards are known, one-sample hypotheses are used to compare site data (provided enough site data are available) with known and pre-established threshold values. It is suggested that the project team determine (e.g., using DQO) or decide (depending upon resources) about the number of site observations that should be collected and compared with the "pre-established" standards before coming to a conclusion about the status (clean or polluted) of the site area (e.g., RU, AOC) under investigation. When the number of available detected site samples is less than 4 to 6, one might perform point-by-point site observation comparisons with a BTV; and when enough detected site observations (> 8 to 10, more are preferable) are available, it is desirable to use single sample hypothesis testing approaches.

Depending upon the parameter (e.g., the average value, μ_0, or a not-to-exceed value, A_0), represented by the known threshold value, one can use single sample hypothesis tests for population mean (t-test, sign test) or single sample tests for proportions and percentiles. The details of the single sample hypotheses testing approaches can be found in EPA (2006) and the Technical Guide for ProUCL 4.0. Several single sample tests listed as follows are available in ProUCL 4.0.

One-Sample t-Test: This test is used to compare the site mean, μ, with a specified cleanup standard, C_s, where the cleanup standard, C_s, represents an average threshold value, μ_0. The Student's t-test (or a UCL of mean) is often used (assuming normality of site data or when site sample size is large such as larger than 30, 50) to determine the attainment of cleanup levels at a polluted site after some remediation activities.

One-Sample Sign Test or Wilcoxon Signed Rank (WSR) Test: These tests are nonparametric tests and can also handle nondetect observations provided all nondetects (e.g., associated detection limits) fall below the <u>specified threshold value, C_s.</u> These tests are used to compare the site location (e.g., median, mean) with a specified cleanup standard, C_s, representing a similar location measure.

One-Sample Proportion Test or Percentile Test: When a specified cleanup standard, A_0, such as a PRG or a BTV represents an upper threshold value of a contaminant concentration distribution (e.g., not-to-exceed value, compliance limit) rather than the mean threshold value, μ_0, of the contaminant concentration distribution, then a test for proportion or a test for percentile (or equivalently a UTL 95%-95%) can be used to compare site proportion or site percentile with the specified threshold or action level, A_0. This test can also handle ND observations provided all NDs are below the compliance limit.

In order to obtain reasonably reliable estimates and test statistics, an adequate amount of representative site data (8 to 10 detected observations) is needed to perform the hypotheses tests. As mentioned before, in case only a few (e.g., < 4 to 6) detected site observations are available, then point-by-point site concentrations may be compared with the specified action level, A_0.

1.6.2 Two-Sample Hypotheses – When BTVs and Not-to-Exceed Values are Unknown

When BTVs, not-to-exceed values, and other cleanup standards are not available, then site data are compared directly with the background data. In such cases, a two-sample hypothesis testing approach can be used to perform site versus background comparisons. Note that this approach can be used to compare concentrations of any two populations including two different site areas or two different monitoring wells (MWs). In order to use and perform a two-sample hypothesis testing approach, enough data should be available (collected) from each of the two populations under investigation. Site and background data requirements (e.g., based upon DQOs) to perform two-sample hypothesis test approaches are described in EPA (1989b, 2006), Breckenridge and Crockett (1995), and the VSP (2005) software package. While collecting site and background data, for better representation of populations under investigation, one may also want to account for the size of the background area (and site area for site samples) into sample size determination. That is, a larger number (> 10 to 15) of representative background (or site) samples should be collected from larger background (or site) areas. As mentioned before, every effort should be made to collect as many samples as determined using DQO processes as described in EPA documents (2006).

The two-sample (or more) hypotheses approaches are used when the site parameters (e.g., mean, shape, distribution) are being compared with the background parameters (e.g., mean, shape, distribution). The two-sample hypotheses testing approach is also used when the cleanup standards or screening levels are not known *a priori*, and they need to be estimated based upon a data set from a background or reference population. Specifically, two-sample hypotheses testing approaches are used to compare 1) the average contaminant concentrations of two or more populations such as the background population and the potentially contaminated site areas, or 2) the proportions of site and background observations exceeding a pre-established compliance limit, A_0. In order to derive reliable conclusions with higher statistical power based upon hypothesis testing approaches, enough data (e.g., minimum of 8 to 10 samples) should be available from all of the populations under investigation. It is also desirable to supplement statistical methods with graphical displays, such as the double Q-Q plots, or side-by-side multiple box plots, as available in ProUCL 4.0. Several parametric and nonparametric two-sample hypotheses testing approaches, including Student's t-test, the Wilcoxon-Mann-Whitney (WMW) test, Gehan's test, and quantile test are included in ProUCL 4.0. Details of those methods are described in the ProUCL 4.0 Technical Guide. It should be noted that the WMW, Gehan, and quantile tests are also available for data sets with NDs. Gehan's test is specifically meant to be used on data sets with multiple detection limits. It is also suggested that for best and reliable conclusions, both the WMW and quantile tests should be used on the same data set. The details of these two tests with examples are given in EPA (1994, 2006).

The samples collected from the two (or more) populations should all be of the same type obtained using similar analytical methods and apparatus. In other words, the collected site and background samples should be all discrete or all composite (obtained using the same design and pattern), and be collected from the same medium (soil) at similar depths (e.g., all surface samples or all subsurface samples) and time (e.g., during the same quarter in groundwater applications) using comparable (preferably same) analytical methods. Good sample collection methods and sampling strategies are given in EPA (1996, 2003) guidance documents.

1.7 Minimum Sample Size Requirements

Due to resource limitations, it may not be possible (nor needed) to sample the entire population (e.g., background area, site area, areas of concern, exposure areas) under study. Statistics is used to draw inference(s) about the populations (clean, dirty) and their known or unknown parameters (e.g., comparability of population means, not-to-exceed values, upper percentiles, and spreads) based upon much smaller data sets (samples) collected from those populations under study. In order to determine and establish BTVs, not-to-exceed values, or site-specific screening levels, defensible data set(s) of appropriate size(s) needs to be collected from background areas (e.g., site-specific, general reference or pristine area, or historical data). The project team and site experts should decide what represents a site population and what represents a background population. The project team should determine the population size and boundaries based upon all current and future objectives for the data collection. The size and area of the population (e.g., a remediation unit, area of concern, or an exposure unit) may be determined based upon the potential land use, and other exposure and risk management objectives and decisions. Moreover, appropriate effort should be made to properly collect soil samples (e.g., methods based upon Gy sampling theory), as described in Gerlach and Nocerino (2003).

Using the collected site and background data sets, statistical methods supplemented with graphical displays are used to perform site versus background comparisons. The test results and statistics obtained by performing such site versus background comparisons are used to determine if the site and background level contaminant concentration are comparable; or if the site concentrations exceed the background threshold concentration level; or if an adequate amount of cleanup and remediation approaching the BTV or a cleanup level have been performed at polluted areas (e.g., AOC, RU) of the site under study.

In order to perform statistical inference (estimation and hypothesis testing), one needs to determine the sample sizes that need to be collected from the populations (e.g., site and background) under investigation using appropriate DQO processes (EPA 2006). However, in some cases, it may not be possible to collect the same number of samples as determined by using a DQO process. For example, the data might have already been collected (often is the case in practice) without using a DQO process, or due to resource constraints, it may not be possible to collect as many samples as determined by using a DQO-based sample size formula. It is observed that, in practice, the project team and the decision makers may not collect enough background samples, perhaps due to various resource constraints. However, every effort should be made to collect at least 8 to 10 (more are desirable) background observations before using methods as incorporated in ProUCL 4.0. The minimum sample size recommendations as described here are useful when resources are limited (as often is the case), and it may not be possible to collect as many background and site (e.g., AOC, EU) samples as computed using DQOs and the sample size determination formulae given in the EPA (2006). Some minimum sample size requirements are also given in Fact Sheet #24, prepared by Wyoming Department of Environmental Quality (June 2005).

As mentioned before, the topics of DQO processes and the sample size determination are described in detail in the EPA (2006) guidance document. Therefore, the sample size determination formulae based upon DQO processes are not included in ProUCL 4.0 and its Technical Guide. It should be noted that DQO-based sample size determination routines are available in DataQUEST (EPA, 1997) and VSP (2005) software packages. Guidance and suggestions are provided on the minimum number of background and site samples needed to be able to use statistical methods for the computation of upper limits, and to perform single sample tests, two-sample tests such as t-test, and the Wilcoxon-Mann-Whitney (WMW) test. The minimum sample size recommendations (requirements) as described here are

made so that reasonably reliable estimates of EPC terms and BTVs, and defensible values of test statistics for single or two-sample hypotheses tests (e.g., t-test, WMW test), can be computed.

Finally, it is also important to note and understand the differences between the following two minimum sample size concepts:

1. Minimum sample needed to compute a statistic based upon theoretical formulae; and
2. Minimum sample size needed to compute a reliable and usable decision statistics.

Even though, most of the decision statistics such as upper confidence limits (UCLs), upper prediction limits (UPLs), and upper tolerance limits (UTLs) can be computed based upon very small samples of sizes 2, 3, and 4, those decision statistics are not reliable and representative enough to make defensible and correct cleanup and remediation decisions. Use of such statistics should be avoided.

Specific recommendations regarding the minimum sample size requirement (when data sets of DQOs based sample sizes cannot be collected) needed to compute reliable and usable decision statistics have also been described in this chapter. It should be pointed out that those minimum sample size recommendation (at least 8-10 observations) should be used only when samples of size determined by a DQO process (EPA, 2006) cannot be collected (e.g., due to resource constraints). The intention of the developers of ProUCL 4.0 is to provide statistically rigorous and defensible methods and decision statistics. Success of the applicability of a statistical method depends upon the quality and quantity of the available data set. It is always desirable to collect appropriate number of samples based upon data quality objectives (DQOs) so that reliable decision statistics (e.g., UPLs, UCLs, and hypotheses test statistics) can be computed to make appropriate decisions.

1.7.1 Minimum Sample Size Requirements to Compute Decision Statistics for Data without NDs

It was noted by the developers of ProUCL software that some users of earlier versions of ProUCL (e.g., ProUCL 3.0 and ProUCL 4.0) were computing decisions statistics (e.g., UCLs, UPLs) based upon small data sets of sizes 2, 3 etc. As a result, in later versions of ProUCL such as ProUCL 4.00.02, the developers restricted the use of ProUCL for samples of size at least 5. ProUCL 4.00.02 and higher versions will not compute decision statistics (e.g., UCLs, UPLs, UTLs) based upon samples of sizes less than 5. Some users did complain about being not able to compute decision statistics based upon samples of size 3 or 4; but that is fine as one should not be computing decision statistics based upon such small samples. It is desirable that the ProUCL users follow the sample size requirements as described in this chapter.

At present, ProUCL 4.00.02 and higher versions will not compute any decision statistics such as UCLs and UPLs, UTLs for data sets of size less than 5 (without NDs). It may compute other summary statistics and graphs but will not compute decision making statistics. For such small data sets of size less than 5, ProUCL 4.00.02 provides warning messages informing the user about the potential deficiencies present in his data set submitted to ProUCL.

1.7.2 Minimum Sample Size Requirements for Data Sets with NDs

ProUCL 4.00.02 and higher versions will not compute any decision statistics based upon data sets of sizes less than 5 consisting of NDs. Moreover, for data sets of size at least 5, no decision statistics will be computed when not more than one detected observation is present in the data set. For such extreme data situations, ProUCL 4.00.02 provides some warning messages regarding the lack of appropriate amount of data. For data sets of size 5 or larger with only two detected values, ProUCL 4.00.02 will produce only Kaplan –Meier method based decision statistics (UCLs, UPLs, UTLs); and for data sets of size 5 or larger with 3 detected values, most nonparametric and parametric (except for gamma distribution based) decision statistics will be computed and printed. For data sets of size 5 or higher with 4 or more detected observations, ProUCL 4.00.02 will produce values for all parametric and nonparametric decision statistics.

For all small data sets (e.g., size <8-10), ProUCL 4.00.02 informs the user by providing appropriate warning messages about the potential deficiencies present in his data set submitted to ProUCL. Some recommendations about how to determine a value of a decision statistic based upon data sets consisting of all or only a few (1 or 2) detected values are also provided in this chapter. It is suggested that the project team and experts associated with the site should come to an agreement about an appropriate value that may be used for the decision statistic under consideration.

1.7.3 Minimum Sample Size for Estimation and Point-by-Point Site Observation Comparisons

- Point-by-point observation comparison method is used when a small number (e.g., 4 to 6 locations) of detected site observations are compared with pre-established or estimated BTVs, screening levels, or PRGs. In this case, individual point-by-point site observations (preferably based upon composite samples from various site locations) are compared with estimated or pre-established background (e.g., USGS values) values, PRGs, or some other not-to-exceed value.

- When BTV contaminant concentrations are not known or pre-established, one has to collect, obtain, or extract a data set of an appropriate size that can be considered as representative of the site related background. Statistical upper limits are computed using the data set thus obtained; which are used as estimates of BTVs and not-to-exceed values. It should be noted that in order to compute reasonably reliable and accurate estimates of BTVs and not-to-exceed values based upon a background (or reference) data set, enough background observations (minimum of 8 to 10) should be collected perhaps using an appropriate DQO process as described in EPA (2006). Typically, background samples are collected from a comparable general reference area or a site-specific area.

- When enough (e.g., > 8 to 10) detected site observations are available, it is preferred to use hypotheses testing approaches. Specifically, single sample hypotheses testing (comparing site to a specified threshold) approaches should be used to perform site versus a known threshold comparison and two-sample hypotheses testing (provided enough background data are also available) approaches should be used to perform site versus background comparison.

1.7.4 Minimum Sample Size Requirements for Hypothesis Testing Approaches

Statistical methods (as in ProUCL 4.0) used to estimate EPC terms, BTVs, PRGs, or to compare the site contaminant concentration data distribution with the background data distribution can be computed based upon small site and background data sets (e.g., of sizes 3, 4, 5, or 6). However, those statistics cannot be considered representative and reliable enough to make important cleanup and remediation decisions. It is recommended not to use those statistics to draw cleanup and remediation decisions potentially impacting the human health and the environment. It is suggested that the estimation and hypothesis testing methods as incorporated in ProUCL 4.0 may not be used on background data sets with fewer than 8 to 10 detected observations. Also, when using hypotheses testing approaches, it is suggested that the site and background data be obtained using an appropriate DQO process as described in EPA (2006). In case that is not possible, it is suggested that the project team at least collect 8 to 10 observations from each of the populations (e.g., site area, MWs, background area) under investigation.

Site versus background comparisons and computation of the BTVs depend upon many factors, some of which cannot be controlled. These factors include the site conditions, lack of historical information, site medium, lack of adequate resources, measurement and analytical errors, and accessibility of the site areas. Therefore, whenever possible, it is desirable to use more than one statistical method to perform site versus background comparison. The use of statistical methods should always be supplemented with appropriate graphical displays.

1.8 Sample Sizes for Bootstrap Methods

Several parametric and nonparametric (including bootstrap methods) UCL, UPL, and other limits computation methods for both full-uncensored data sets (without nondetects) and left-censored data sets with nondetects are available in ProUCL 4.0. It should be noted that bootstrap resampling methods are useful when not too few (e.g., < 10-15) and not too many (e.g., > 500-1000) detected observations are available. For bootstrap methods (e.g., percentile method, BCA bootstrap method, bootstrap t method), a large number (e.g., 1000, 2000) of bootstrap resamples (with replacement) are drawn from the same data set. Therefore, in order to obtain bootstrap resamples with some distinct values (so that statistics can be computed from each resample), it is suggested that a bootstrap method should not be used when dealing with small data sets of sizes less than 10-15. Also, it is not required to bootstrap a large data set of size greater than 500 or 1000; that is when a data set of a large size (e.g., > 1000) is available, there is no need to obtain bootstrap resamples to compute statistics of interest (e.g., UCLs). One can simply use a statistical method on the original large data set. Moreover, bootstrapping a large data set of size greater than 500 or 1000 will be time consuming.

1.9 Statistical Analyses by a Group ID

The analyses of data categorized by a group ID variable such as: 1) Surface vs. Subsurface; 2) AOC1 vs. AOC2; 3) Site vs. Background; and 4) Upgradient vs Downgradient monitoring wells are quite common in many environmental applications. ProUCL 4.0 offers this option for data sets with and without nondetects. The Group Option provides a powerful tool to perform various statistical tests and methods (including graphical displays) separately for each of the group (samples from different populations) that may be present in a data set. For an example, the same data set may consist of samples from the various groups or populations representing site, background, two or more AOCs, surface, subsurface, monitoring wells. The graphical displays (e.g., box plots, Q-Q plots) and statistics

(computations of background statistics, UCLs, hypotheses testing approaches) of interest can be computed separately for each group by using this option.

It should be pointed out that it is the users' responsibility to provide adequate amount of detected data to perform the group operations. For an example, if the user desires to produce a graphical Q-Q plot (using only detected data) with regression lines displayed, then there should be at least two detected points (to compute slope, intercept, *sd*) in the data set. Similarly if the graphs are desired for each of the group specified by the group ID variable, there should be at least two detected observations in each group specified by the group variable. ProUCL 4.0 generates a warning message (in orange color) in the lower panel of the ProUCL 4.0 screen. Specifically, the user should make sure that a variable with nondetects and categorized by a group variable should have enough detected data in each group to perform the various methods (e.g., GOF tests, Q-Q plots with regression lines) as incorporated in ProUCL 4.0.

1.10 Use of Maximum Detected Values as Estimates of Upper Limits

Some practitioners tend to use the maximum detected value as an estimate of the EPC term. This is especially true when the sample size is small such as ≤ 5, or when a UCL95 exceeds the maximum detected values (EPA, 1992b). Also, many times in practice, the BTVs and not-to-exceed values are estimated by the maximum detected value. This section discusses the appropriateness of using the maximum detected value as estimates of the EPC term, BTVs, or other nor-to-exceed values.

1.10.1 Use of Maximum Detected Values to Estimate BTVs and Not-to-Exceed Values

It is noted that BTVs and not-to-exceed values represent upper threshold values in the upper tail of a data distribution; therefore, depending upon the data distribution and sample size, the BTVs and other not-to-exceed values may be estimated by the maximum detected value. As described earlier, upper limits, such as UPLs, UTLs, and upper percentiles, are used to estimate the BTVs and not-to-exceed values. It is noted that a nonparametric UPL or UTL is often estimated by higher order statistics such as the maximum value or the second largest value (EPA 1992a, RCRA Guidance Addendum). The use of higher order statistics to estimate the UTLs depends upon the sample size. For an example: 1) 59 to 92 samples, a nonparametric UTL95%-95 is given by the maximum detected value; 2) 93 to 123 samples, a nonparametric UTL95%-95 is given by the second largest maximum detected value; and 3) 124 to 152 samples, a UTL95%-95 is given by the third largest detected value in the sample.

Note: *Therefore, when a data set does not follow a discernable distribution, the maximum observed value (or other high order statistics) may be used as an estimate of BTV or a not-to-exceed value, provided the maximum value does not represent an outlier or a contaminating observation perhaps representing a hot location.*

1.10.2 Use of Maximum Detected Values to Estimate EPC Terms

Some practitioners tend to use the maximum detected value as an estimate of the EPC term. This is especially true when the sample size is small such as ≤ 5, or when a UCL95 exceeds the maximum detected values (EPA, 1992b). Specifically, a RAGS document (EPA, 1992) suggests the use of the maximum detected value as a default value to estimate the EPC term when a 95% *UCL* (e.g., the *H-UCL*) exceeded the maximum value. ProUCL 4.0 can compute a 95% *UCL* of mean using several methods based upon normal, Gamma, lognormal, and non-discernable distributions. In past (e.g., EPA, 1992b),

only two methods were used to estimate the EPC term based upon: 1) Student's t-statistic and a normal distribution, and 2) Land's H-statistic (1975) and a lognormal model. The use of H-statistic often yields unstable and impractically large UCL95 of the mean (Singh, Singh, and Iaci, 2002). For skewed data sets of smaller sizes (e.g., < 30, < 50), H-UCL often exceeds the maximum detected value. This is especially true when extreme high outliers may be present in the data set. Since the use of a lognormal distribution has been quite common (e.g., suggested as a default model in a RAGS document (EPA, 1992)), the exceedance of the maximum detected value by H-UCL95 is frequent for many skewed data sets of smaller sizes (e.g., < 30, < 50). It is also be noted that for highly skewed data sets, the sample mean indeed can even exceed the upper 90%, 95%, etc., percentiles, and consequently, a 95% *UCL* of mean can exceed the maximum observed value of a data set.

All of these occurrences result in the possibility of using the maximum detected value as an estimate of the EPC term. It should be pointed out that in some cases, the maximum observed value actually might represent a highly polluted outlying observation. Obviously, it is not desirable to use a highly polluted value as an estimate of average exposure (EPC term) for an exposure area. This is especially true when one is dealing with lognormally distributed data sets of small sizes. As mentioned before, for such highly skewed data sets that cannot be modeled by a gamma distribution, a 95% *UCL* of the mean should be computed using an appropriate distribution-free nonparametric method.

It should be pointed out that the EPC term represents the average exposure contracted by an individual over an exposure area (EA) during a long period of time; therefore, the EPC term should be estimated by using an average value (such as an appropriate 95% *UCL* of the mean) and not by the maximum observed concentration. One needs to compute an average exposure and not the maximum exposure. It is unlikely that an individual will visit the location (e.g., in an EA) of the maximum detected value all of the time. One can argue that the use of this practice results in a conservative (higher) estimate of the EPC term. The objective is to compute an accurate estimate of the EPC term. Several other methods (instead of H-UCL) as described in EPA (2002), and included in ProUCL 4.0 (EPA 2007), are available to estimate the EPC terms. It is unlikely (but possible with outliers) that the UCLs based upon those methods will exceed the maximum detected value, unless some outliers are present in the data set. ProUCL 4.0 displays a warning message when the recommended 95% *UCL* (e.g., Hall's or bootstrap t *UCL* with outliers) of the mean exceeds the observed maximum concentration. When a 95% *UCL* does exceed the maximum observed value, ProUCL4.0 recommends the use of an alternative *UCL* computation method based upon the Chebyshev inequality. The detailed recommendations (as functions of sample size and skewness) for the use of those UCLs are summarized in ProUCL 3.0 User Guide (EPA, 2004).

Singh and Singh (2003) studied the performance of the max test (using the maximum observed value as an estimate of the EPC term) via Monte Carlo simulation experiments. They noted that for skewed data sets of small sizes (e.g., < 10-20), the max test does not provide the specified 95% coverage to the population mean, and for larger data sets, it overestimates the EPC term, which may require unnecessary further remediation. The use of the maximum value as an estimate of the EPC term also ignores most (except for maximum value) of the information contained in the data set. With the availability of so many *UCL* computation methods (15 of them), the developers of ProUCL 4.0 do not recommend using the maximum observed value as an estimate of the EPC term representing an average exposure by an individual over an EA. Also, for the distributions considered, the maximum value is not a sufficient statistic for the unknown population mean.

Note: *It is recommended that the maximum observed value NOT be used as an estimate of the EPC term representing average exposure contracted by an individual over an EA. For the sake of interested users, ProUCL displays a warning message when the recommended 95% UCL (e.g., Hall's bootstrap UCL etc.)*

of the mean exceeds the observed maximum concentration. For such scenarios (when a 95% UCL does exceed the maximum observed value), an alternative 95% UCL computation method is recommended by ProUCL 4.0.

1.10.3 Samples with Nondetect Observations

Nondetect observations (or less than obvious values) are inevitable in most environmental data sets. Singh, Maichle, and Lee (EPA, 2006) studied the performances (in terms of coverages) of the various UCL95 computation methods including the simple substitution methods (such as the DL/2 and DL methods) for data sets with nondetect observations. They concluded that the UCLs obtained using the substitution methods, including the replacement of nondetects by respective DL/2, do not perform well even when the percentage of nondetect observations is low, such as 5%-10%. They recommended avoiding the use of substitution methods to compute UCL95 based upon data sets with nondetect observations.

1.10.4 Avoid the Use of DL/2 Method to Compute UCL95

Based upon the results of the report by Singh, Maichle, and Lee (EPA, 2006), it is strongly recommended to avoid the use of the DL/2 method to perform GOF test, and to compute the summary statistics and various other limits (e.g., UCL, UPL) often used to estimate the EPC terms and BTVs. Until recently, the DL/2 method has been the most commonly used method to compute the various statistics of interest for data sets with BDL observations. The main reason of its common use has been the lack of the availability of other defensible methods and associated programs that can be used to estimate the various environmental parameters of interest. Today, several other methods (e.g., KM method, bootstrap methods) with better performances are available that can be used to compute the various upper limits of interest. Some of those parametric and nonparametric methods are available in ProUCL 4.0. *Even though the DL/2 method (to compute UCLs, UPLs, and for goodness-of-fit test) has also been incorporated in ProUCL 4.0, its use is not recommended due to its poor performance.* The DL/2 method is included in ProUCL 4.0 only for historical reasons as it had been the most commonly used and recommended method until recently (EPA, 2006). Some of the reviewers of ProUCL 4.0 suggested and requested the inclusion of DL/2 method in ProUCL for comparison purposes.

Note: *The DL/2 method has been incorporated in ProUCL 4.0 for historical reasons only. NERL-EPA, Las Vegas strongly recommends avoiding the use of DL/2 method even when the percentage (%) of NDs is as low as 5%-10%. There are other methods available in ProUCL 4.0 that should be used to compute the various summary statistics and upper limits based upon data sets with multiple detection limits.*

1.10.5 Samples with Low Frequency of Detection

When all of the sampled data values are reported as nondetects, the EPC term should also be reported as a nondetect value, perhaps by the maximum reporting limit (RL) or maximum RL/2. Statistics (e.g., UCL95) computed based upon only a few detected values (e.g., < 4 to 6) cannot be considered reliable enough to estimate the EPC terms having potential impact on the human heath and the environment. When the number of detected data is small, it is preferable to use simple ad hoc methods rather than using statistical methods to compute the EPC terms and other upper limits. Specifically, it is suggested that in cases when the detection frequency is low (e.g., < 4%-5%) and the number of detected observations is low, the project team and the decision makers together should make a decision on site-specific basis on how to estimate the average exposure (EPC term) for the contaminant and area under consideration. For such data sets with low detection frequencies, other measures such as the median or mode represent better estimates (with lesser uncertainty) of the population measure of central tendency.

Additionally, it is also suggested that when most (e.g., > %95) of the observations for a contaminant lie below the detection limit(s) or reporting limits (RLs), the sample median or the sample mode (rather than the sample average which cannot be computed accurately) may be used as an estimate the EPC term. Note that when the majority of the data are nondetects, the median and the mode will also be a nondetect. The uncertainty associated with such estimates will be high. It is noted that the statistical properties, such as the bias, accuracy, and precision of such estimates, would remain unknown. In order to be able to compute defensible estimates, it is always desirable to collect more samples.

Note: *In case the number of available detected samples is small (< 5), it is suggested that the project team decide about the estimation of the EPC term on site-specific basis. For such small data sets with very few detected values (< 5), the final decision ("policy decision") on how to estimate the EPC term should be determined by the project team and decision makers.*

1.11 Other Applications of Methods in ProUCL 4.0

In addition to performing background versus site comparisons for CERCLA and RCRA sites, and estimating the EPC terms in exposure and risk evaluation studies, the statistical methods as incorporated in ProUCL 4.0 can be used to address other issues dealing with environmental investigations that are conducted at Superfund or RCRA sites.

1.11.1 Identification of COPCs

Risk assessors and RPMs often use screening levels or BTVs to identify the COPCs during the screening phase of a cleanup project to be conducted at a contaminated site. The screening for the COPCs is performed prior to any characterization and remediation activities that may have to be conducted at the site under investigation. This comparison is performed to screen out those contaminants that may be present in the site medium of interest at low levels (e.g., at or below the background levels or pre-established screening levels) and may not pose any threat and concern to human health and the environment. Those contaminants may be eliminated from all future site investigations, and risk assessment and risk management studies.

In order to identify the COPCs, point-by-point site observations (preferably composite samples) are compared with pre-established screening levels, SSL, or estimated BTVs. This is especially true when the comparisons of site concentrations with screening levels or BTVs are conducted in real time by the sampling or cleanup crew right there in the site field. The project team should decide about the type of site samples (discrete or composite) and the number of detected site observations (not more than 4 to 6) that should be collected and compared with the screening levels or the BTVs. In case BTVs, screening levels, or not-to-exceed values are not known, the availability of a defensible background or reference data set of reasonable size (e.g., > 8 to 10, more are preferable) is required to obtain reliable estimates of BTVs and screening levels. When a reasonable number of detected site observations are available, it is preferable to use hypotheses testing approaches. The contaminants with concentrations exceeding the respective screening values or BTVs may be considered as COPCs, whereas contaminants with concentrations (in all collected samples) lower than the screening value, PRG, or an estimated BTV may be omitted from all future evaluations including the risk assessment and risk management investigations.

1.11.2 Identification of Non-Compliance Monitoring Wells

In monitoring well (MW) compliance assessment applications, individual (often discrete) contaminant concentrations from a MW are compared with pre-established ACL, MCL, or an estimated compliance limit (CL) based upon a group of upgradient wells representing the background population. An exceedance of the MCL or the BTV by a MW concentration may be considered as an indication of contamination in that MW. In such individual concentration comparisons, the presence of contamination (determined by an exceedance) may have to be confirmed by re-sampling from that MW. If concentrations of contaminants in both the original sample and the re-sample(s) exceed the MCL or BTV, then that MW may require closer scrutiny, perhaps triggering the remediation remedies as determined by the project team. If the concentration data from a MW for about 4 to 5 continuous quarters (or another designated time period determined by the project team) are below the MCL or BTV level, then that MW may be considered as complying with (achieving) the pre-established or estimated standards. Statistical methods as described in Chapters 5 and 6 of the revised background guidance document (EPA, 2002b) can be used to estimate the not-to-exceed values or BTVs based upon background or upgradient wells in case the ACLs or MCLs are not pre-determined.

1.11.3 Verification of the Attainment of Cleanup Standards, C_s

Hypothesis testing approaches may be used to verify the attainment of the cleanup standard, C_s, at polluted site areas of concern after conducting remediation and cleanup at the site AOC (EPA, 2006). In order to properly address this scenario, a site data set of adequate size (minimum of 8 to 10 detected site observations) needs to be made available from the remediated or excavated areas of the site under investigation. The sample size should also account for the size of the remediated site area; meaning that larger site areas should be sampled more (with more observations) to obtain a representative sample of the site under investigation.

Typically, the null hypothesis of interest is H_0: Site Mean, $\mu_s >= C_s$ versus the alternative hypothesis, H_1: Site Mean, $\mu_s < C_s$, where the cleanup standard, C_s, is known *a priori*. The sample size needed to perform such single sample hypotheses tests can be obtained using the DQO process-based sample size formula as given in the EPA (2006) documents. In any case, in order to use this test, a minimum of 8 to 10 detected site samples should be collected. The details of statistical methods used to perform single sample hypothesis as described above can be found in EPA (2006).

1.11.4 Using BTVs (Upper Limits) to Identify Hot Spots

The use of upper limits (e.g., UTLs) to identify hot spot(s) has also been mentioned in the *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b). Point-by-point site observations (preferably using composite samples representing a site location) are compared with a pre-established or estimated BTV. Exceedances of the BTV by site observations may be considered as representing locations with elevated concentrations (hot spots). Chapters 5 and 6 of the revised background guidance document (EPA, 2002b) describe several methods to estimate the BTVs based upon full data sets without nondetects and left-censored data sets with nondetect observations.

The rest of the chapters of this User Guide illustrate the use of the various procedures as incorporated in ProUCL 4.0. Those methods are useful to analyze environmental data sets with and without the nondetect observations. It is noted that ProUCL 4.0 is the first software package equipped with single sample and two-sample hypotheses testing approaches that can be used on data sets with nondetect observations.

1.12 Some General Issues and Recommendations made by ProUCL

Some general issues regarding the handling of multiple detection limits and field duplicates by ProUCL and recommendations made about various substitution and ROS methods for data sets with NDs are described in the following sections.

1.12.1 Multiple Detection Limits

ProUCL 4.0 and its later versions do not make distinctions between method detection limits (MDLs), adjusted MDLs, sample quantitation limits (SQLs), or detection limits (DLs). Multiple detection limits in ProUCL mean different values of the detection limit. All these nondetect (ND) observations in ProUCL are indentified by the value '0' of the indicator variable used in ProUCL 4.0 to distinguish between detected and nondetected observations. It is users' responsibility to supply correct numerical values (should be entered as the reported detection limit value) for ND observations in the data set, and to create an indicator variable column associated with each variable/analyte consisting of ND values. It should be noted that some of the methods (e.g., Kaplan-Meier - KM Method) included in ProUCL 4.0 can handle data sets with detection limits greater than the detected values.

1.12.2 Field Duplicates

Since, collection of analytical data are not cheap, field duplicates collected from the same site area under investigation may be used in the statistical analysis to compute various decision statistics of interest. Duplicates should be considered just like any other discrete samples collected from the site area under study. Alternatively, the project team may come to an agreement regarding the use of duplicates in statistical analyses (e.g., computing decision statistics) of data collected from the site.

1.12.3 ProUCL Recommendation about ROS Method and Substitution (DL/2) Method

In order to estimate EPC (computing UCL95) for data sets with NDs, ProUCL 4.0 and higher versions have several methods including substitution methods (e.g., DL/2, DL), ROS methods (normal, lognormal, and gamma), and Kaplan-Meier Method. Extensive simulation study conducted by Singh, Maichle and Lee (2006) demonstrated that statistically rigorous K-M method yields more precise and accurate estimates of EPC terms than those based upon substitution and ROS methods. Even though several of the substitution and ROS methods have been incorporated in ProUCL (for historical reasons and comparison purposes), those methods are not recommended by ProUCL to estimate the EPC terms or to compute other decision statistics.

Chapter 2

Entering and Manipulating Data

2.1 Creating a New Data Set

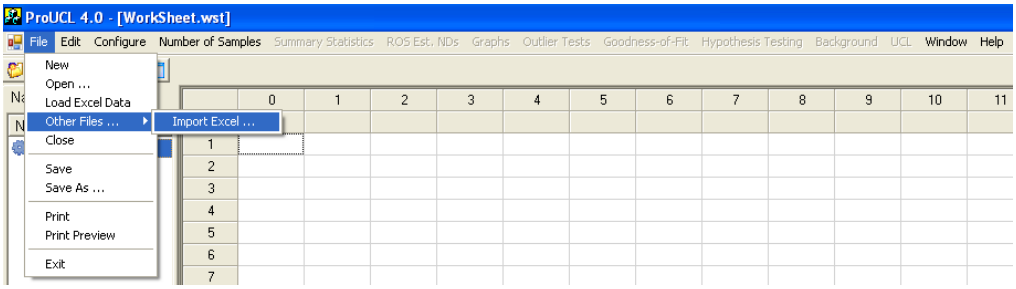
By simply executing the ProUCL 4.0, a new worksheet is generated and displayed for the user to enter data.

To create a new worksheet: click or choose **File ► New**

2.2 Opening an Existing Data Set

If your data sets are stored in the ProUCL data format (*.wst), then click or choose **File ► Open**

- If your data sets are stored in the Microsoft Excel format (*.xls), then choose **File ► Other Files… ►Import Excel…** OR **File ► Load Excel Data**



- Possible Error Messages:
  - When you import an Excel file, make sure that you have an empty worksheet. If there is no empty worksheet, then you must create a new worksheet before importing an Excel file. Otherwise, there will be an error message in the Log Panel: "[Error] Worksheet must be empty."
  - First open a new worksheet and then import the Excel file.
  - Make sure that the file you trying to import is not currently open. Otherwise, there will be the following warning message in the Log panel:
  - "[Information] Unable to open C:***.xls." Check the validity of this file.

2.3 Input File Format

- Note that ProUCL 4.0 does not require that data in each column must end with a nonzero value (ProUCL 3.0 requires this). Therefore, all zero values (in the beginning, middle, or end of data columns) are treated as valid zero values as part of the data set.

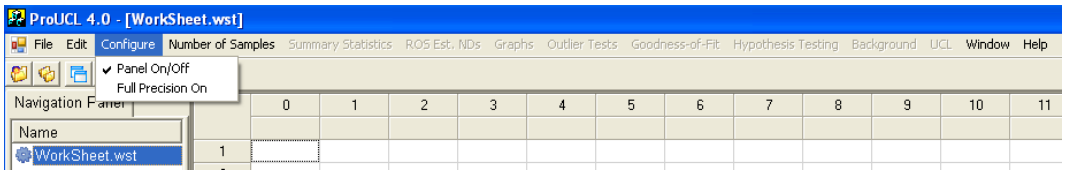
- The program can read Excel files. The user can perform typical Cut, Paste, and Copy operations.

- The first row in all input data files consist of alphanumeric (strings of numbers and characters) names representing the header row. Those header names may represent meaningful variable names such as Arsenic, Chromium, Lead, Group-ID, and so on.

  - The Group-ID column has the labels for the groups (e.g., Background, AOC1, AOC2, 1, 2, 3, a, b, c, Site 1, Site 2, and so on) that might be present in the data set. The alphanumeric strings (e.g., Surface, Sub-surface) can be used to label the various groups.
  - The data file can have multiple variables (columns) with unequal number of observations.
  - Except for the header row and columns representing the group labels, only numerical values should appear in all other columns.
  - All alphanumeric strings and characters (e.g., blank, other characters, and strings), and all other values (that do not meet the requirements above) in the data file are treated as missing values.
  - Also, a large value denoted by 1E31 (= $1x10^{31}$) can be used to represent missing data values. All entries with this value are ignored from the computations. These values are counted under number of missing values.

2.4 Number Precision

- You may turn "Full Precision" on or off by choosing **Configure ► Full Precision On/OFF**



- By leaving "Full Precision" turned off, ProUCL will display numerical values using an appropriate (default) decimal digit option. However, by turning "Full Precision" off, all decimal values will be rounded to the nearest thousandths place.

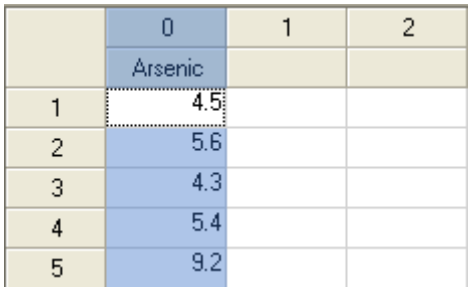
- "Full Precision" on option is specifically useful when one is dealing with data sets consisting of small numerical values (e.g., < 1) resulting in small values of the various estimates and test statistics. These values may become so small with several leading zeros

(e.g., 0.00007332) after the decimal. In such situations, on may want to use "Full Precision" option to see nonzero values after the decimal.

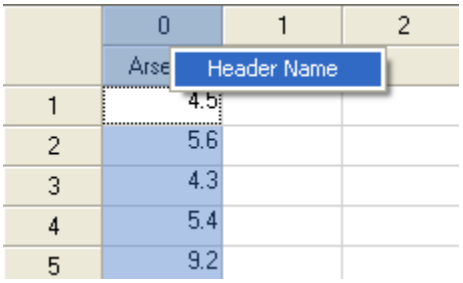
Note: *For the purpose of this User Guide, unless noted otherwise, all examples have been described using the "Full Precision" off option. This option prints out results up to 3 significant digits after the decimal.*

2.5 Entering and Changing a Header Name

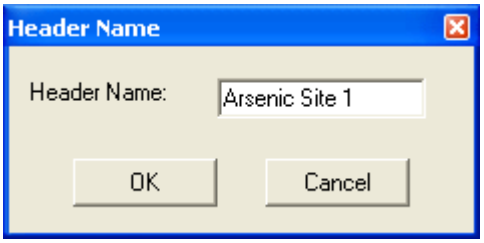
1. Highlight the column whose header name (variable name) you want to change by clicking either the column number or the header as shown below.



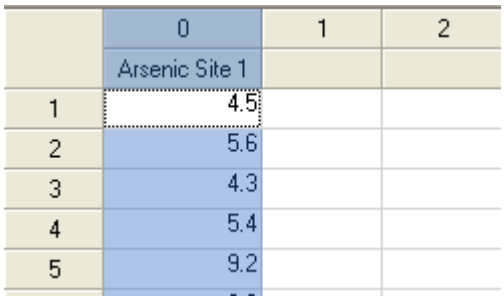
2. Right-click and then click "Header Name"



3. Change the Header Name.

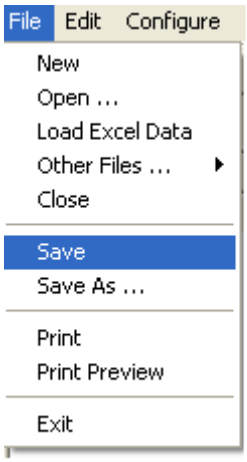


4. Click the OK button to get the following output with the changed variable name.



	0	1	2
	Arsenic Site 1		
1	4.5		
2	5.6		
3	4.3		
4	5.4		
5	9.2		

2.6 Saving Files

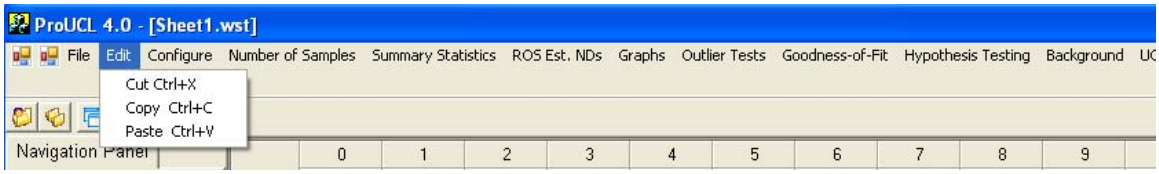


- Save option allows the user to save the active window.

- Save As option allows the user to save the active window. This option follows typical Windows standards, and saves the active window to a file in Excel 95 (or higher) format. All modified/edited data files, and output screens (excluding graphical displays) generated by the software can be saved as Excel 95 (or higher) spreadsheet.

2.7 Editing

Click on the Edit menu item to reveal the following drop-down options.



The following Edit drop-down menu options are available:

- Cut option: similar to a standard Windows Edit option, such as in Excel. It performs standard edit functions on selected highlighted data (similar to a buffer).

- Copy option: similar to a standard Windows Edit option, such as in Excel. It performs

typical edit functions on selected highlighted data (similar to a buffer).

- Paste option: similar to a standard Windows Edit option, such as in Excel. It performs typical edit functions of pasting the selected (highlighted) data to the designated spreadsheet cells or area.

- It should be noted that the Edit option could also be used to Copy Graphs. This topic (copying and pasting graphs) is illustrated in detail in Chapter 13.

2.8 Handling Nondetect Observations

- ProUCL 4.0 can handle data sets with single and multiple detection limits.

- For a variable with nondetect observations (e.g., arsenic), the detected values, and the numerical values of the associated detection limits (for less than values) are entered in the appropriate column associated with that variable.

- Specifically, the data for variables with nondetect values are provided in two columns. One column consists of the detected numerical values with less than ($< DL_i$) values entered as the corresponding detection limits (or reporting limits), and the second column represents their detection status consisting of only 0 (for less than values) and 1 (for detected values) values. The name of the corresponding variable representing the detection status should start with d_, or D_ (not case sensitive) and the variable name. The detection status column with variable name starting with a D_ (or a d_) should have only two values: 0 for nondetect values, and 1 for detected observations.

- For an example, the header name, D_Arsenic is used for the variable, Arsenic having nondetect observations. The variable D_Arsenic contains a 1 if the corresponding Arsenic value represents a detected entry, and contains a 0 if the corresponding entry for variable, Arsenic, represents a nondetect. An example data set illustrating these points is given as follows.

D:\example.wst

	0	1	2	3	4	5	6
	Arsenic	D_Arsenic	Mercury	D_Mercury	Vanadium	Zinc	Group
1	4.5	0	0.07	1	16.4	89.3	Surface
2	5.6	1	0.07	1	16.8	90.7	Surface
3	4.3	0	0.11	0	17.2	95.5	Surface
4	5.4	1	0.2	0	19.4	113	Surface
5	9.2	1	0.61	1	15.3	266	Surface
6	6.2	1	0.12	1	30.8	80.9	Surface
7	6.7	1	0.04	1	29.4	80.4	Surface
8	5.8	1	0.06	1	13.8	89.2	Surface
9	8.5	1	0.99	1	18.9	182	Surface
10	5.65	1	0.125	1	17.25	80.4	Surface
11	5.4	1	0.18	1	17.2	91.9	Subsurface
12	5.5	1	0.21	1	16.3	112	Subsurface
13	5.9	1	0.29	1	16.8	172	Subsurface
14	5.1	1	0.44	1	17.1	99	Subsurface
15	5.2	1	0.12	1	10.3	90.7	Subsurface
16	4.5	0	0.055	1	15.1	66.3	Subsurface
17	6.1	1	0.055	1	24.3	75	Subsurface
18	6.1	1	0.21	1	18	185	Subsurface
19	6.8	1	0.67	1	16.9	184	Subsurface
20	5	1	0.1	1	12	68.4	Subsurface
21			0.8	1			
22			0.26	1			
23			0.97	1			
24			0.05	1			
25			0.26	1			

2.9 Caution

- Care should be taken to avoid any misrepresentation of detected and nondetected values. Specifically, it is advised not to have any missing values (blanks, characters) in the D_column (detection status column). If a missing value is located in the D_ column (and not in the associated variable column), the corresponding value in the variable column is treated as a nondetect, even if this might not have been the intention of the user.

- It is mandatory that the user makes sure that only a 1 or a 0 are entered in the detection status D_column. If a value other than a 0 or a 1 is entered in the D_ column (the detection column), results may become unreliable, as the software defaults to any number other than 0 or 1 as a nondetect value.

- When computing statistics for full data sets without any nondetect values, the user should select only those variables (from the list of available variables) that contain no nondetect observations. Specifically, nondetect values found in a column chosen for the summary statistics (full-uncensored data set) will be treated as a detected value; whatever value (e.g., detection limit) is entered in that column will be used to compute summary statistics for a full-uncensored data set without any nondetect values.

- **Two-Sample Hypotheses:** It should be noted that at present, when using two-sample hypotheses approaches (WMW test, Gehan test, and quantile test) on data sets with NDs,

both samples or variables (e.g., site-As, Back-As) should be specified as having nondetects, even though one of the variables may not have any ND observations. This means, a ND column (with 0 = ND, and 1 = detect) should be provided for each variable (here D_site-As, and D_Back-As) to be used in this comparison. *If a variable (e.g., site-As) does not have any nondetects, still a column with label D_site-As should be included in the data set with all entries = 1 (detected values).*

- The following sample (not from a Superfund site) data set given on the next page illustrates points related to this option and issues listed above. The data set considered contains some nondetect measurements for Arsenic and Mercury. It should also be noted that the Mercury concentrations are used to illustrate the points related to nondetect observations only. Arsenic and Zinc concentrations are used to illustrate the use of the group variable, Group (Surface, Subsurface).

- If for mercury, one computes summary statistics (assuming no nondetect values) using "Full" data set option, then all nondetect values (with "0" entries in D_Mercury column) will be treated as detected values, and summary statistics will be computed accordingly.

2.10 Summary Statistics for Data Sets with Nondetect Observations

- In order to compute simple summary statistics or to compute other statistics of interest (e.g., background statistics, GOF test, UCLs, outliers) for variables with nondetect values, one should choose the nondetect option, "With NDs" from the various available menu options such as Outliers, Background Statistics, UCLs, Goodness-of-Fit test, Q-Q plot, Box Plot.

- It should be noted that "summary" statistics for a data set with nondetect observations represent (at least in ProUCL 4.0) <u>simple summary statistics</u> based upon the data set without using nondetect observations. All other parametric and nonparametric statistics and estimates of population mean, variance, percentiles (e.g., MLEs, KM, and ROS estimates) for variables with nondetect observations are given in other menu options such as background statistics and UCL. The simple "Summary Statistics/With NDs" option only provides simple statistics (e.g., % NDs, max ND, Min ND, and Mean of detected values) based upon detected values. These statistics (e.g., *sd* of log-transformed detected values) may help a user to determine the degree of skewness (e.g., mild, moderate, high) of the data set consisting of detected values. These statistics may also help the user to choose the most appropriate method (e.g., KM (BCA) UCL or KM (t) UCL) to compute UCLs, UPLs, and other limits.

- As mentioned before, various other statistics and estimates of interest (e.g., mean, *sd*, UCLs, UTLs, MLEs, and KM estimates) for data sets with nondetect observations are computed in other the menu options (UCLs, Outliers, Background Statistics, GOF tests) available in ProUCL 4.0.

2.11 Warning Messages and Recommendations for Datasets with Insufficient Amount of Data

- ProUCL 4.0 provides some warning messages and recommendations for datasets with insufficient amount of data to calculate meaningful estimates and statistics of interest. For an example, it is not desirable to compute an estimate the EPC term based upon a data set of size less than 5, especially when nondetects may be present in the data set. In such cases, it is suggested to use site specific values (perhaps determined by the Project Team) to estimate environmental parameters (e.g., EPC term, not-to-exceed background value) of interest.

- Some examples of datasets with insufficient amount of data include datasets with less than 4 distinct observations, datasets with only one detected observation, and datasets consisting of all nondetects.

- Some of the warning messages given by ProUCL 4.0 are shown below.

General Background Statistics for Data Sets with Non-Detects

User Selected Options	
From File	C:\Documents and Settings\narmbya\Desktop\ProUCL-Test\EVILDATA3.wst
Full Precision	OFF
Confidence Coefficient	95%
Coverage	90%
Different or Future K Values	1
Number of Bootstrap Operations	2000

Var_x7

General Statistics

Total Number of Observations	15	Number of Distinct Observations	3

Raw Statistics		Log-Transformed Statistics	
Minimum	1	Minimum	0
Maximum	3	Maximum	1.099
Second Largest	3	Second Largest	1.099
First Quartile	1	First Quartile	0
Median	2	Median	0.693
Third Quartile	3	Third Quartile	1.099
Mean	2	Mean	0.597
SD	0.845	SD	0.47
Coefficient of Variation	0.423		
Skewness	0		

Warning: There are only 3 Distinct Values in this data

There are insufficient Distinct Values to perform some GOF tests and bootstrap methods.

Those methods will return a 'N/A' value on your output display!

It is necessary to have 4 or more Distinct Values to compute bootstrap methods.

It is recommended to have 10-15 or more observations for accurate and meaningful bootstrap results.

General UCL Statistics for Data Sets with Non-Detects

User Selected Options	
From File	C:\Documents and Settings\narmbya\Desktop\ProUCL-Test\EVILDATA3.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000

Var_x3

General Statistics

Number of Valid Data	15	Number of Detected Data	1
Number of Distinct Detected Data	1	Number of Non-Detect Data	14
		Percent Non-Detects	93.33%

Warning: Only one distinct data value was detected! ProUCL (or any other software) should not be used on such a data set!
It is suggested to use alternative site specific values determined by the Project Team to estimate environmental parameters (e.g., EPC, BTV).

The data set for variable Var_x3 was not processed!

General Background Statistics for Data Sets with Non-Detects

User Selected Options	
From File	C:\Documents and Settings\narmbya\Desktop\ProUCL-Test\EVILDATA3.wst
Full Precision	OFF
Confidence Coefficient	95%
Coverage	90%
Different or Future K Values	1
Number of Bootstrap Operations	2000

Var_x10

General Statistics

Number of Valid Data	15	Number of Detected Data	0
Number of Distinct Detected Data	0	Number of Non-Detect Data	15

Warning: All observations are Non-Detects (NDs), therefore all statistics and estimates should also be NDs!
Specifically, sample mean, UCLs, UPLs, and other statistics are also NDs lying below the largest detection limit!
The Project Team may decide to use alternative site specific values to estimate environmental parameters (e.g., EPC, BTV).

The data set for variable Var_x10 was not processed!

v

General Statistics			
Number of Valid Data	15	Number of Detected Data	6
Number of Distinct Detected Data	2	Number of Non-Detect Data	9
		Percent Non-Detects	60.00%

Raw Statistics		Log-transformed Statistics	
Minimum Detected	1	Minimum Detected	0
Maximum Detected	4	Maximum Detected	1.386
Mean of Detected	2.5	Mean of Detected	0.693
SD of Detected	1.643	SD of Detected	0.759
Minimum Non-Detect	2	Minimum Non-Detect	0.693
Maximum Non-Detect	5	Maximum Non-Detect	1.609

Note: Data have multiple DLs - Use of KM Method is recommended	Number treated as Non-Detect	15
For all methods (except KM, DL/2, and ROS Methods),	Number treated as Detected	0
Observations < Largest ND are treated as NDs	Single DL Non-Detect Percentage	100.00%

Warning: Data set has only 2 Distinct Detected Values.
This may not be adequate enough to compute meaningful and reliable test statistics and estimates.
The Project Team may decide to use alternative site specific values to estimate environmental parameters (e.g., EPC, BTV).

Unless Data Quality Objectives (DQOs) have been met, it is suggested to collect additional observations.

The number of detected data may not be adequate enough to perform GOF tests, bootstrap, and ROS methods.
Those methods will return a 'N/A' value on your output display!

It is necessary to have 4 or more Distinct Values for bootstrap methods.
It is recommended to have 10 to 15 or more observations for accurate and meaningful results and estimates.

2.12 Handling Missing Values

- ProUCL 4.0 can handle missing values within a data set.
- All blanks, alphanumeric strings (except for group variables), or the specific large number value 1e31 are considered as missing values.
- A group variable (representing two or more groups, populations, AOCs, MWs) can have alphanumeric values (e.g., MW1, MW2, ...).
- ProUCL 4.0 ignores all missing values in all mathematical operations it performs. Missing values are therefore not treated as being part of a data set.
- Number of Valid Samples or Number of Valid Observations represents the Total Number of Observations less the Number of Missing Values. If there are missing values, then number of valid samples = total number of observations.

  Valid Samples = Total Number of Observations – Missing Values.

- It is important to note, however, that if a missing value not meant (e.g., a blank, or 1e31) to represent a group category is present in a "Group" variable, ProUCL 4.0 will treat that blank value (or 1e31 value) as a new group. Any variable that corresponds to this missing value will be treated as part of a new group and not with any existing groups. It is therefore very important to check the consistency and validity of all data sets before performing complex mathematical operations.

- ProUCL 4.0 prints out the number of missing values (if any) associated with each variable in the data sheet. This information is provided in several output sheets (e.g., summary statistics, background statistics, UCLs) generated by ProUCL 4.0.

For further clarification of labeling of missing values, the following example illustrates the terminology used for the number of valid samples, number of unique and distinct samples on the various output sheets generated by ProUCL 4.0.

Example: The following example illustrates the notion of Valid Samples, Unique or Distinct Samples, and Missing Values. The data set also has nondetect values. ProUCL 4.0 computes these numbers and prints them on the UCLs and background statistics output.

x	D_x
2	1
4	1
2.3	1
1.2	0
w34	0
1.0E+031	0
	0
anm	0
34	1
23	1
0.5	0
0.5	0
2.3	1
2.3	1
2.3	1
34	1
73	1

Valid Samples: Represents the total number of observations (censored and uncensored) excluding the missing values. If a data set has no missing value, then the total number of data points equals number of valid samples.

Missing Values: All values not representing a real numerical number are considered as missing values. Specifically, all alphanumeric values including blanks are considered as missing values. Also unrealistic big numbers such as 1.0e31 are also considered as missing values and are considered as not valid observations.

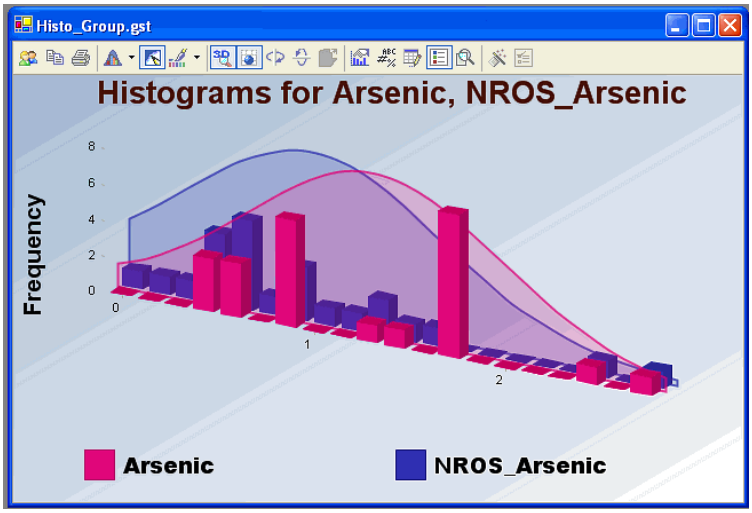
Unique or Distinct Samples: The number of unique samples or number of distinct samples represents all unique (or distinct) detected values. Number of unique or distinct values is computed for detected values

only. This number is especially useful when using bootstrap methods. As well known, it is not desirable and advisable to use bootstrap methods, when the number of unique samples is less than 4-5.

2.13 User Graphic Display Modification

Advanced users are provided two sets of tools to modify graphics displays. A graphics tool bar is available above the graphics display and the user can right-click on the desired object within the graphics display, and a drop-down menu will appear. The user can select an item from the drop-down menu list by clicking on that item. This will allow the user to make desired modifications as available for the selected menu item. An illustration is given as follows.

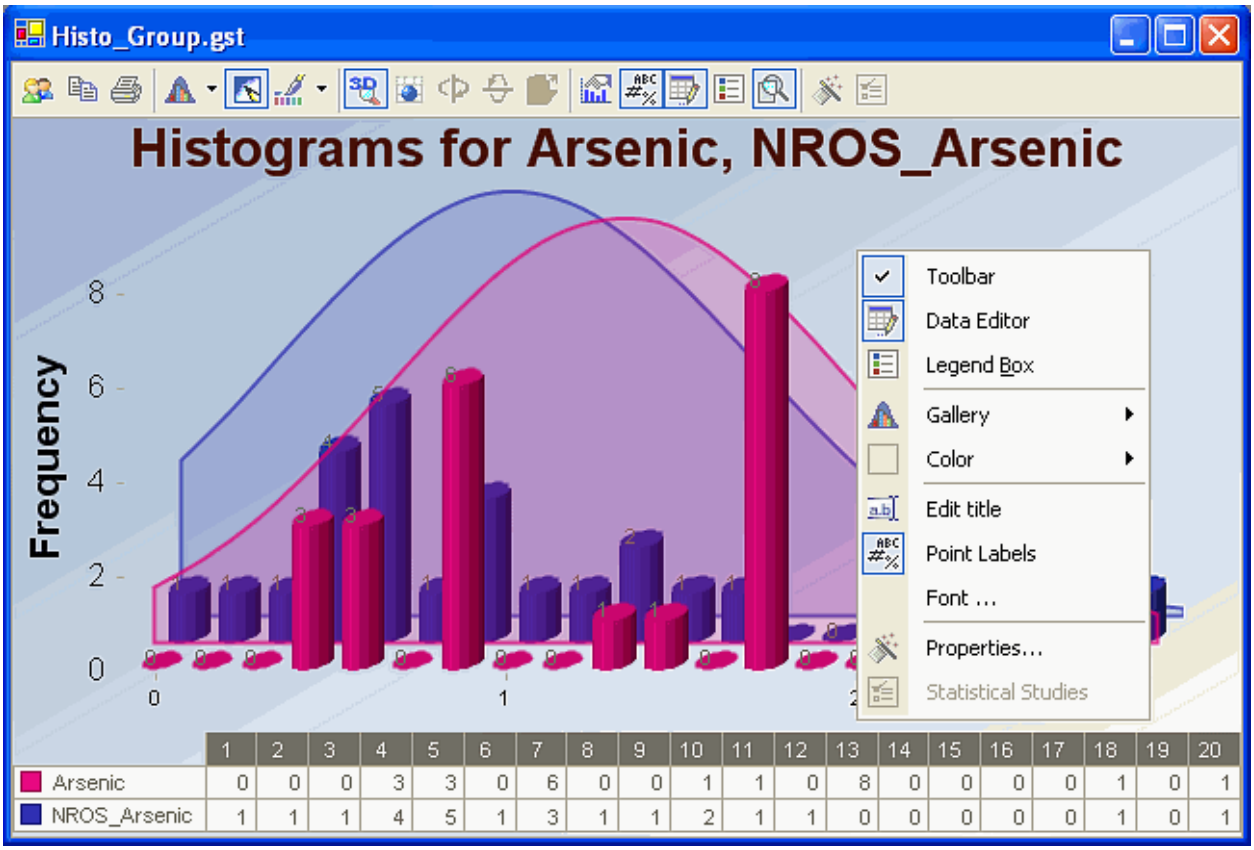
2.13.1 Graphics Tool Bar



The user can change fonts, font sizes, vertical and horizontal axis's, select new colors for the various features and text. All these actions are generally used to modify the appearance of the graphic display. The user is cautioned that these tools can be unforgiving and may put the user in a situation where the user cannot go back to the original display. Users are on their own in exploring the robustness of these tools. Therefore, less experienced users may want to stay away from using these drop-down menu graphic tools.

2.13.2 Drop-Down Menu Graphics Tools

These tools allow the user to move the mouse to a specific graphic item like an Axis Label or a display feature. The user then right-clicks their mouse and a drop-down menu will appear. This menu presents the user with available options for that particular control or graphic object. There is less of chance of making an unrecoverable error but that risk is always present. As a cautionary note, the user can always delete the graphics window and redraw the graphical displays by repeating their operations from the datasheet and menu options available in ProUCL 4.0. An example of a drop-down menu obtained by right-clicking the mouse on the background area of the graphics display is given as follows.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Arsenic	0	0	0	3	3	0	6	0	0	1	1	0	8	0	0	0	0	1	0	1
NROS_Arsenic	1	1	1	4	5	1	3	1	1	2	1	1	0	0	0	0	0	1	0	1

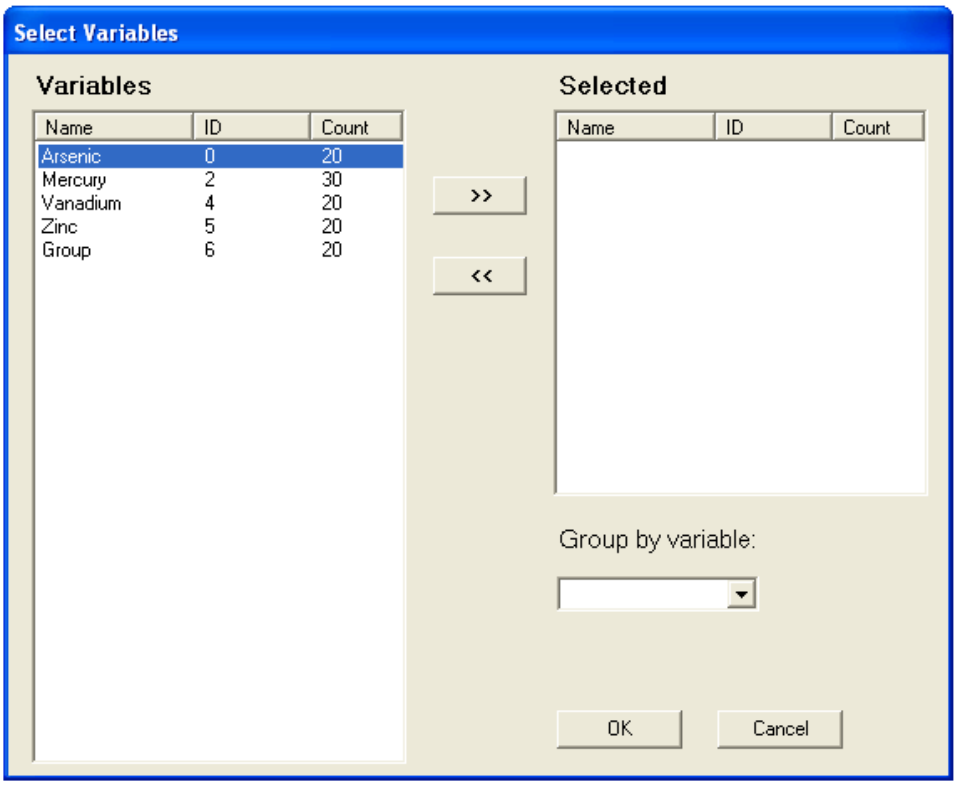
Chapter 3

Select Variables Screen

3.1 Select Variables Screen

- Variables need to be selected to perform statistical analyses.

- When the user clicks on a drop-down menu for a statistical procedure, the following window will appear.



- The Options button is available in certain menus. The use of this option leads to a different pop-up window.

- Multiple variables can be processed simultaneously in ProUCL 4.0. Note that this option was not available in ProUCL 3.0. ProUCL 4.0 can generate graphs, compute UCLs, and background statistics simultaneously for all selected variables.

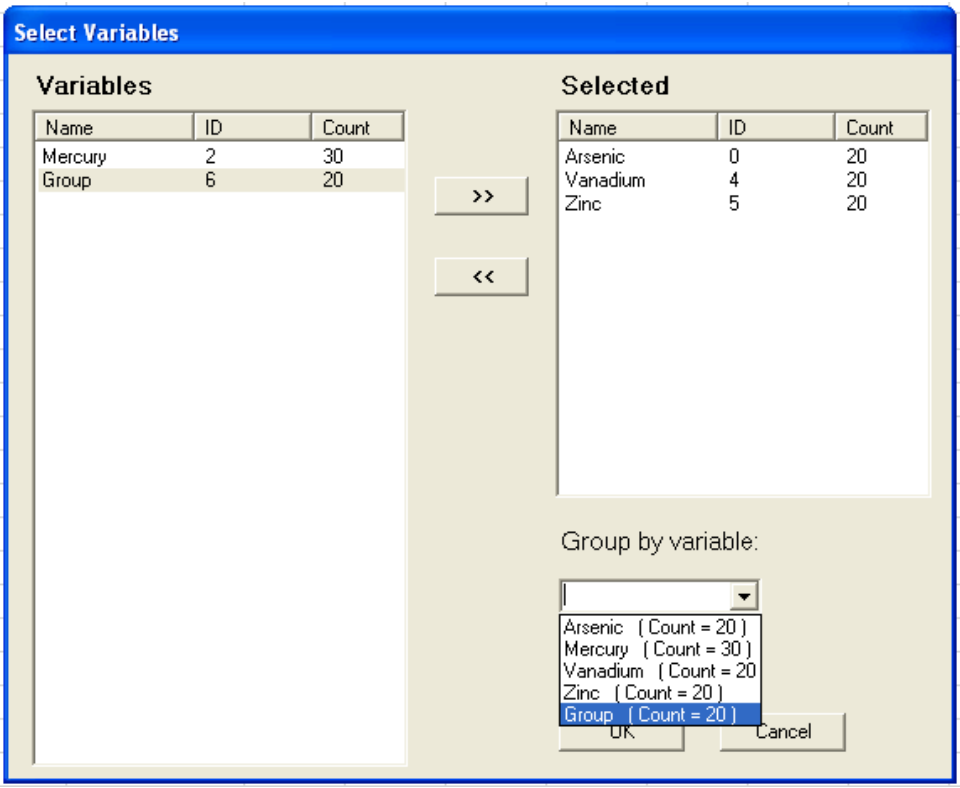
- Moreover, if the user wants to perform statistical analysis on a variable (e.g., contaminant) by a Group variable, click the arrow below the **Group by variable** to get a drop-down list of available variables to select an appropriate group variable. For an example, a group variable (e.g., Site Area) can have alphanumeric values such as AOC1,

AOC2, AOC3, and Background. Thus in this example, the group variable name, Site Area, takes four values such as AOC1, AOC2, AOC3, and Background.

- The Group variable is particularly useful when data from two or more samples need to be compared.

- Any variable can be a group variable. However, for meaningful results, only a variable, that really represents a group variable (categories) should be selected as a group variable.

- The number of observations in the group variable and the number observations in the selected variables (to be used in a statistical procedure) should be the same. In the example below, the variable "Mercury" is not selected because the number of observations for Mercury is 30; in other words mercury values have not been grouped. The group variable and each of the selected variables have 20 data values.



- It is recommended not to assign any missing value such as a "Blank" for the group variable. If there is a missing value (represented by blanks, strings or 1E31) for a group variable, ProUCL 4.0 will treat those missing values as a new group. As such, data values corresponding to the missing Group will be assigned to a new group.

Caution: Once again, care should be taken to avoid misrepresentation and improper use of group variables. It is recommended not to assign any missing value for the group variable.

More on Group Option

- The Group Option provides a powerful tool to perform various statistical tests and methods (including graphical displays) separately for each of the group (samples from different populations) that may be present in a data set. For an example, the same data set may consist of samples from the various groups (populations). The graphical displays (e.g., box plots, Q-Q plots) and statistics of interest can be computed separately for each group by using this option.

- In order to use this option, at least one group variable (with alphanumeric values) should be included in the data set. The various values of the group variable represent different group categories that may be present in the data set. This can be seen in the example.wst data set used earlier in Chapter 2.

- At this time, the number of values (representing group membership) in a Group variable should equal to the number of values in the variable (e.g., Arsenic) of interest that needs to be partitioned into various groups (e.g., monitoring wells).

- Typically, the data for the various groups (categorized by the group variable) represent data from the various site areas (e.g., background, AOC1, AOC2, ...), or from monitoring wells (e.g., MW1, MW2, …).

3.1.1 Graphs by Groups

- Individual or multiple graphs (Q-Q plots, box plots, and histograms) can be displayed on a graph by selecting the "Graphs by Groups" option.

- Individual graph for each group (specified by the selected group variable) is produced by selecting the "Individual Graph" option.

- Multiple graphs (e.g., side-by-side box plots, multiple Q-Q plots on the same graph) are produced by selecting the "Group Graph" option for a variable categorized by a group variable. Using this "Group Graph" option, multiple graphs can be displayed for all sub-groups included in the Group variable. This option is useful when data to be compared are given in the same column and are classified by the group variable.

- Multiple graphs (e.g., side-by-side box plots, multiple Q-Q plots) for selected variables are produced by selecting the "Group Graph" option. Using the "Group Graph" option, multiple graphs can be displayed for all selected variables. This option is useful when data (e.g., lead) to be compared are given in different columns, perhaps representing different populations.

Note: *It should be noted that it is the users' responsibility to provide adequate amount of detected data to perform the group operations. For an example, if the user desires to produce a graphical Q-Q plot (using only detected data) with regression lines displayed, then there should be at least two detected points (to compute slope, intercept, sd) in the data set. Similarly if the graphs are desired for each of the group specified by the group ID variable, there should be at least two detected observations in each group specified by the group variable. ProUCL 4.0 generates a warning message (in orange color) in the lower panel of the ProUCL 4.0 screen. Specifically, the user should make sure that a variable with nondetects*

and categorized by a group variable should have enough detected data in each group to perform the various methods (e.g., GOF tests, Q-Q plots with regression lines) as incorporated in ProUCL 4.0.

As mentioned before, the analyses of data categorized by a group ID variable such as: 1) Surface vs. Subsurface; 2) AOC1 vs. AOC2; 3) Site vs. Background; and 4) Upgradient vs. Downgradient monitoring wells are quite common in many environmental applications.

The usefulness of the group option is illustrated throughout the User Guide using various methods as incorporated in ProUCL 4.0.

Chapter 4

Summary Statistics

This option is used to compute general summary statistics for all variables in the data file. Summary statistics can be generated for full data sets without nondetect observations, and for data sets with nondetect observations. Two Menu options: Full and With NDs are available.

- Full – This option computes summary statistics for all variables in a data set without any nondetect values.

- With NDs – This option computes simple summary statistics for all variables in a data set that have nondetect (ND) observations. For variables with ND observations, only simple summary statistics are computed based upon detected observations only.
  - For this option, no attempt is made to compute estimates of population parameters (e.g., mean, *sd*, SE) using parametric (e.g., MLE) or nonparametric (e.g., KM, bootstrap) estimation methods. Those statistics are generated in other estimation modules (e.g., Background and UCL) of ProUCL 4.0.

Each menu option (Full and With NDs) has two sub-menu options:

- Raw Statistics

- Log-Transformed

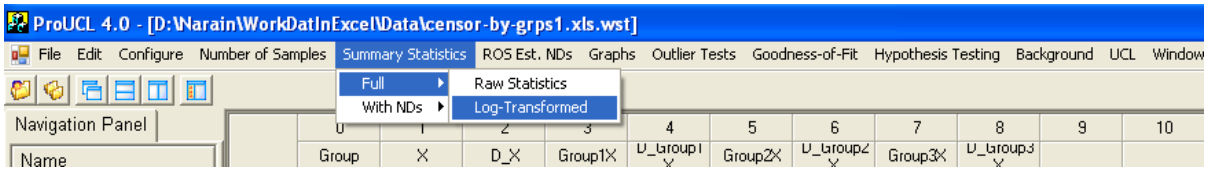
- In ProUCL, log-transformation means natural logarithm (ln)

- When computing summary statistics for raw data, a message will be displayed for each variable that contains non-numeric values.

- The Summary Statistics option computes log-transformed data only if all of the data values for the selected variable(s) are positive real numbers. A message will be displayed if non-numeric characters, zero, or negative values are found in the column corresponding to the selected variable.

4.1 Summary Statistics with Full Data Sets

1. Click **Summary Statistics ► Full**



2. Select either Log-Transformed or Raw Statistics option.

3. The **Select Variables Screen** (see Chapter 3) will appear.

- Select one or more variables from the Select Variables screen.
- If summary statistics are to be computed by a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in drop-down list of available variables, and select the proper group variable.
- Click on the OK button to continue or on the Cancel button to cancel the Summary Statistic option.

Raw Statistics

Full_Raw_Stats.ost

From File: D:\example.wst

Summary Statistics for Raw Full Data Sets

Variable	NumObs	Minimum	Maximum	Mean	Median	Variance	SD	MAD/0.675	Skewness	Kurtosis	CV
Arsenic (subsurface)	10	4.5	6.8	5.56	5.45	0.449	0.67	0.667	0.344	-0.047	0.121
Arsenic (surface)	10	4.3	9.2	6.185	5.725	2.5	1.581	1.075	0.969	0.295	0.256
Vanadium (subsurface)	10	10.3	24.3	16.4	16.85	13.86	3.723	1.26	0.495	2.092	0.227
Vanadium (surface)	10	13.8	30.8	19.53	17.23	33.72	5.807	2.669	1.46	0.896	0.297
Zinc (subsurface)	10	66.3	185	114.4	95.45	2271	47.66	35.21	0.745	-1.288	0.416
Zinc (surface)	10	80.4	266	116.7	90	3680	60.66	13.86	2.111	4.065	0.52

Log-Transformed Statistics

Full_Log_Stats.ost

From File: D:\example.wst

Summary Statistics for Log-Transformed Full Data Sets

Variable	NumObs	Minimum	Maximum	Mean	Median	Variance	SD	MAD/0.675	Skewness	Kurtosis	CV
Arsenic (subsurface)	10	1.504	1.917	1.709	1.696	0.0144	0.12	0.123	0.0692	-0.14	0.0702
Arsenic (surface)	10	1.459	2.219	1.795	1.745	0.0591	0.243	0.176	0.527	-0.153	0.135
Vanadium (subsurface)	10	2.332	3.19	2.774	2.824	0.0538	0.232	0.0735	-0.394	1.351	0.0836
Vanadium (surface)	10	2.625	3.428	2.938	2.846	0.0699	0.264	0.157	1.194	0.504	0.09
Zinc (subsurface)	10	4.194	5.22	4.666	4.558	0.159	0.399	0.425	0.452	-1.394	0.0855
Zinc (surface)	10	4.387	5.583	4.673	4.5	0.162	0.403	0.163	1.734	2.184	0.0862

4. The resulting Summary Statistics screen as shown above can be saved as an Excel file. Click Save from the file menu.

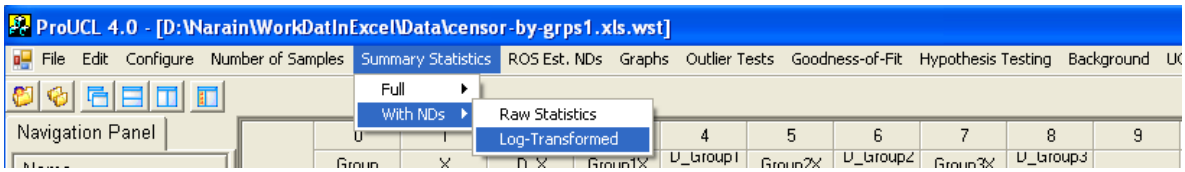
5. On the output screen, the following summary statistics are displayed for each selected variable in the data file.

NumObs = Number of Observations
Minimum = Minimum value
Maximum = Maximum value
Mean = Sample average value
Median = Median value
Variance = Classical sample variance
SD = Classical sample standard deviation
MAD = Median absolute deviation
MAD/0.675 = Robust estimate of variability, population standard deviation, σ
Skewness = Skewness statistic
Kurtosis = Kurtosis statistic
CV = Coefficient of Variation

The details of these summary statistics are described in an EPA (2006) guidance document and also in the ProUCL Technical Guide (A. Singh and A.K. Singh (EPA, 2007).

4.2 Summary Statistics with NDs

1. Click **Summary Statistics ► With NDs**



2. Select either Log-Transformed or Raw Statistics option.

3. The **Select Variables Screen** (Chapter 3) will appear.

- Select variable(s) from the list of variables.
- Only those variables that have nondetect values will be shown.
- If summary statistics are to be computed by a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables; then select the proper group variable.
- Click on the OK button to continue or on the Cancel button to cancel the summary statistics operations.

Note: *It should be noted that in ProUCL 4.0, "Summary Statistics" for a data set with nondetect observations represent* <u>*simple summary statistics*</u> *based upon the data set without using nondetect observations. All other parametric and nonparametric statistics and estimates of population mean, variance, percentiles (e.g., MLEs, KM, and ROS estimates) for variables with nondetect observations are given in other estimation menu options such as background statistics and UCL. The simple "Summary*

Statistics/With NDs" option only provides simple statistics (e.g., % NDs, max ND, Min ND, Mean of detected values) based upon detected values. These statistics (e.g., sd of log-transformed detected values) may help a user to determine the degree of skewness (e.g., mild, moderate, high) of the data set consisting of detected values. These statistics may also help the user to choose the most appropriate method (e.g., KM (BCA) UCL or KM (t) UCL) to compute UCLs, UPLs, and other limits.

Raw Statistics – Data Set with NDs

WND_Raw_Stats.ost

From File: D:\example.wst

Summary Statistics for Raw Data Sets with NDs

				Raw Statistics using Detected Observations							
Variable	NumObs	NumNDs	% NDs	Maximum	Minimum	Mean	Median	SD	MAD/0.675	Skewness	CV
Arsenic	17	3	15.00%	5	9.2	6.126	5.8	1.15	0.593	1.783	0.188
Mercury	25	5	16.67%	0.04	0.99	0.312	0.18	0.315	0.163	1.202	1.01

Log-Transformed Statistics – Data Set with NDs

WND_Log_Stats.ost

From File: D:\example.wst

Summary Statistics for Log-transformed Data Sets with NDs

				Statistics using Detected Log-transformed Observations							
Variable	NumObs	NumNDs	% NDs	Maximum	Minimum	Mean	Median	SD	MAD/0.675	Skewness	CV
Arsenic	17	3	15.00%	1.609	2.219	1.798	1.758	0.168	1.458	0.0935	0.106
Mercury	25	5	16.67%	-3.219	-0.0101	-1.66	-1.715	1.032	0.238	-0.622	1.4

- The Summary Statistics screen shown above can be saved as an Excel file. Click the save from the file menu.

- On the results screen, the following summary statistics are displayed for each selected variable from the data file.

  Num Obs = Number of Observations
  NumNDs = Number of Nondetects

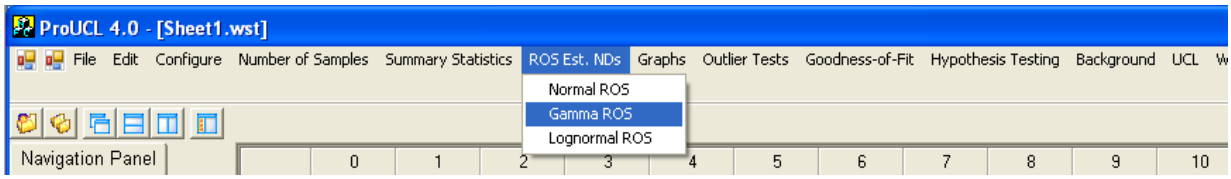
% NDs = Percentage of Nondetect observations
Minimum = Minimum value
Maximum = Maximum value
Mean = Sample average value
Median = Median value
SD = Classical sample standard deviation
MAD = Median Absolute Deviation
MAD/0.675 = Robust estimate of variability (standard deviation)
CV = Coefficient of Variation

Chapter 5

Estimating Nondetects Using ROS Methods

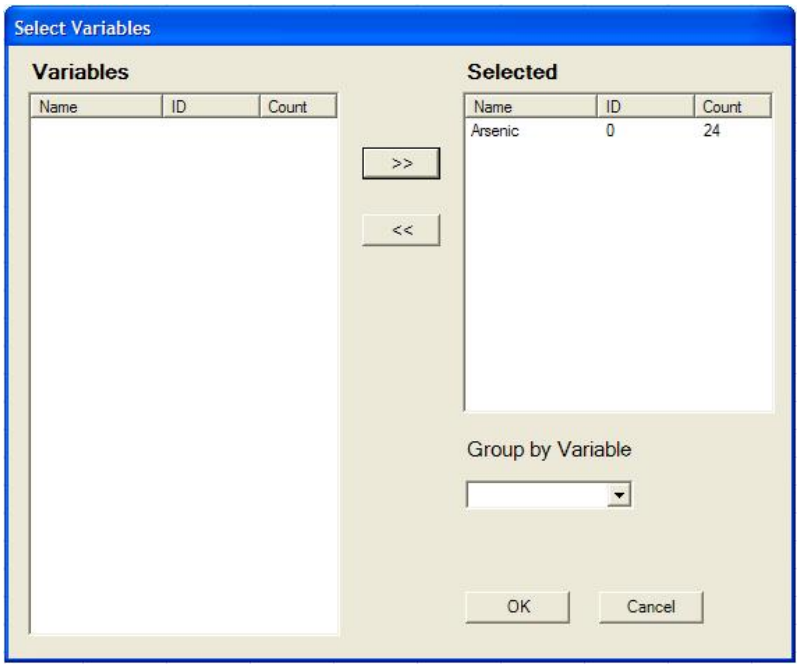
Regression on order statistics (ROS) can be used to extrapolate nondetect observations using a normal, lognormal, or gamma model. ProUCL 4.0 has three ROS estimation methods that can be used to estimate or extrapolate nondetect observations. The use of this option generates additional columns consisting of all extrapolated nondetects and detected observations. These columns are appended to the existing open spreadsheet. The user should save the updated file if they want to use the generated data for their other application(s).

1. Click **ROS Est. NDs ► Gamma ROS**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.



- Click on the OK button to continue or on the Cancel button to cancel the option.

Output Screen for ROS Est. NDs (Gamma) Option

	0	1	2
	Arsenic	D_Arsenic	GROS_Arsenic
1	1	0	1.0315131074651300
2	1	0	1.199323080983740
3	1.7	1	1.7
4	1	0	1.33761048919618
5	1	0	1.46028213348334
6	2	0	0.957865781528449
7	3.2	1	3.2
8	2	0	1.08600369175045
9	2	0	1.19098252772635
10	2.8	1	2.8
11	2	0	1.28363461914326
12	2	0	1.368290930400230
13	2	0	1.44721892249559
14	2	0	1.52178645574469
15	2	0	1.59289577776005
16	0.7	1	0.7
17	0.9	1	0.9
18	0.5	1	0.5
19	0.5	1	0.5
20	0.9	0	1.226870494768360
21	0.5	1	0.5
22	0.7	1	0.7
23	0.6	1	0.6
24	1.5	1	1.5

Note: *Columns with similar naming convention are generated for each selected variable and distribution using this ROS option.*

Chapter 6

Graphical Methods (Graph)

Three commonly used graphical displays are available under the Graphs option:

- Box Plot
- Histogram
- Multi-QQ

- The box plots and multiple Q-Q plots can be used for Full data sets without nondetects and also for data sets with nondetect values.

- Three options are available to draw Q-Q plots with nondetect (ND) observations. Specifically, Q-Q plots are displayed only for detected values, or with NDs replaced by ½ DL values, or with NDs replaced by the respective detection limits. The statistics displayed on a Q-Q plot (mean, *sd*, slope, intercept) are computed according to the method used. The NDs are displayed using the smaller font and in red color.

- ProUCL 4.0 can display box plots for data sets with NDs. This kind of graph may not be very useful when many NDs may be present in a data set.

  - A few choices are available to construct box plots for data sets with NDs. For an example, some texts (e.g., Helsel) display box plots only for the detected observations. Specifically, all nondetects below the largest detection limit (DL), and portion of the box plot (if any) below the largest DL are not shown on the box plot. A horizontal line is displayed at the largest detection limit level.
  - ProUCL 4.0 constructs a box plot using all detected and nondetect (using DL values) values. ProUCL 4.0 shows the full box plot. However, a horizontal line is displayed at the largest detection limit.

- When multiple variables are selected, one can choose to: 1) produce a multiple graphs on the same display by choosing the Graph by group variable option, or 2) produce separate graphs for each selected variable.

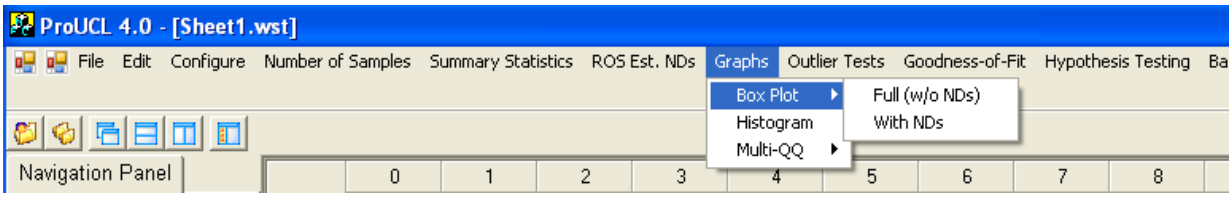
- The Graph by group variable option produces side-by-side box plots, or multiple Q-Q plots, or histograms for the groups of the selected variable representing samples obtained from multiple populations (groups). These multiple graphs are particularly useful to perform two (background versus site) or more sample visual comparisons.

  - Additionally, Box Plot has an optional feature, which can be used to draw lines at statistical (e.g., upper limits of background data set) limits computed from one population on the box plot obtained using the data from another population (a site area of concern). This type of box plot represents a useful visual comparison of site data with background threshold values (background upper limits).
  - Up to four (4) statistics can be added (drawn) on a box plot. If the user inputs a value in the value column, the check box in that row will get activated. For example, the

user may want to draw horizontal lines at 80th percentile, 90th percentile, 95th percentile, or a 95% UPL) on a box plot.

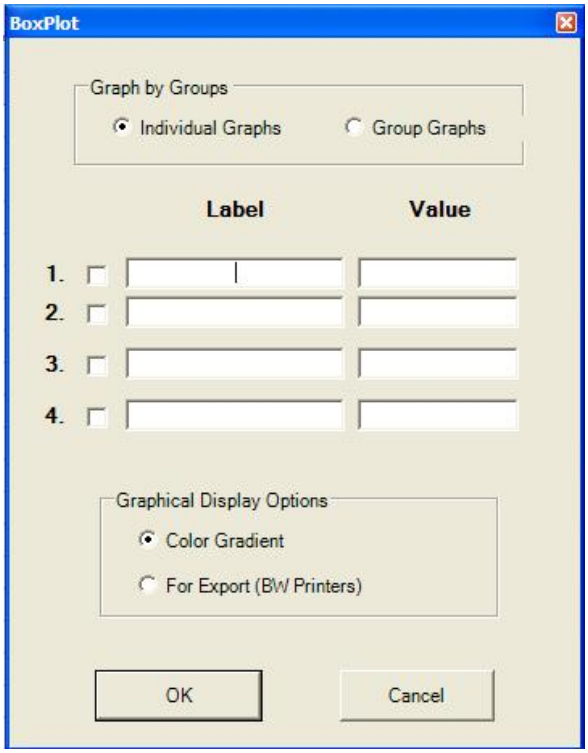
6.1 Box Plot

1. Click **Graphs ► Box Plot**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select an appropriate variable representing a group variable.
- When the **Group by variable** button is clicked, the following window is shown.

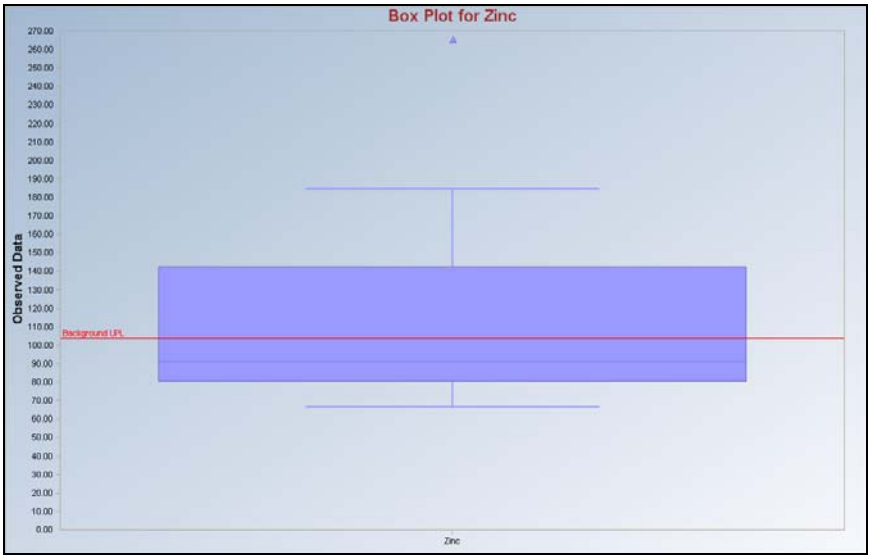


- The default option for **Graph by Groups** is **Individual Graphs**. This option will produce one graph for each selected variable. If you want to put all the selected variables into a single graph, then select the **Group Graphs** option. This **Group Graphs** option is used when multiple graphs categorized by a Group variable have to be produced on the same graph.

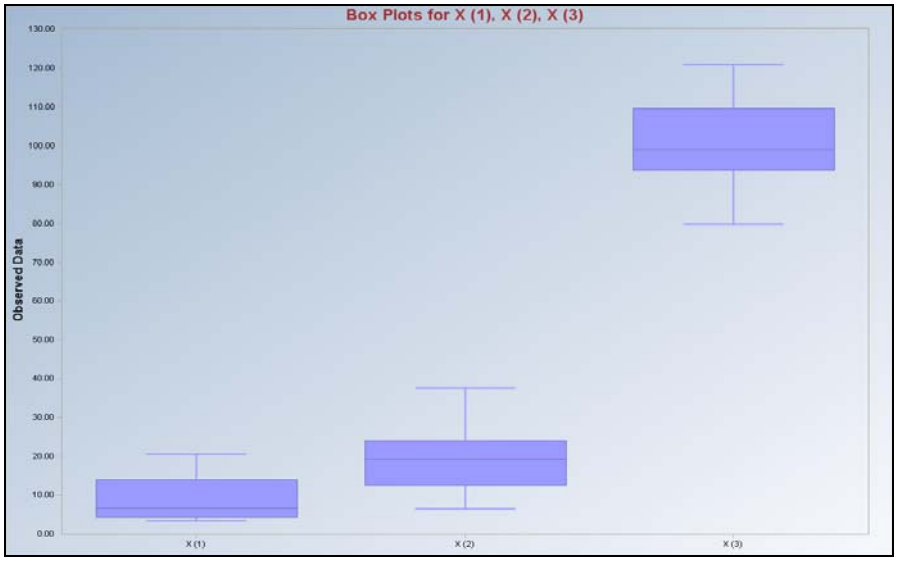
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to use and import graphs in black and white into a document or report, then check the radio button next to **For Export (BW Printers)**.

- Click on the OK button to continue or on the Cancel button to cancel the Box Plot (or other selected graphical) option.

Box Plot Output Screen (Single Graph)
Selected options: Label (Background UPL), Value (103.85), Individual Graphs, and Color Gradient.

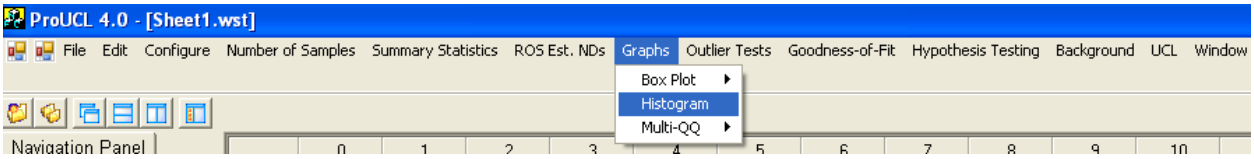


Box Plot Output Screen (Group Graphs)
Selected options: Group Graphs and Color Gradient.



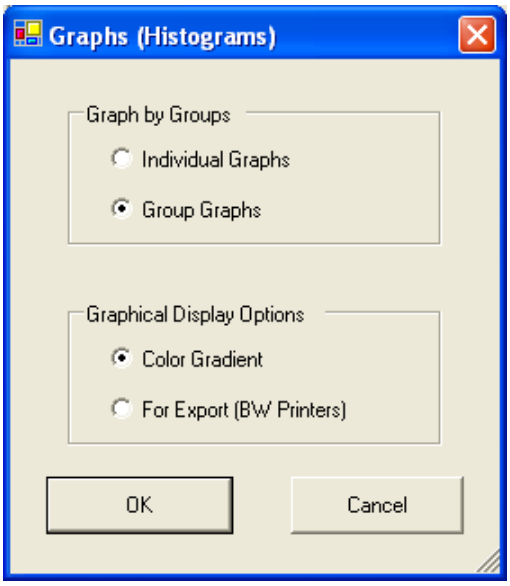
6.2 Histogram

1. Click **Graphs ► Histogram**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.
- When that option button is clicked, the following window will be shown.

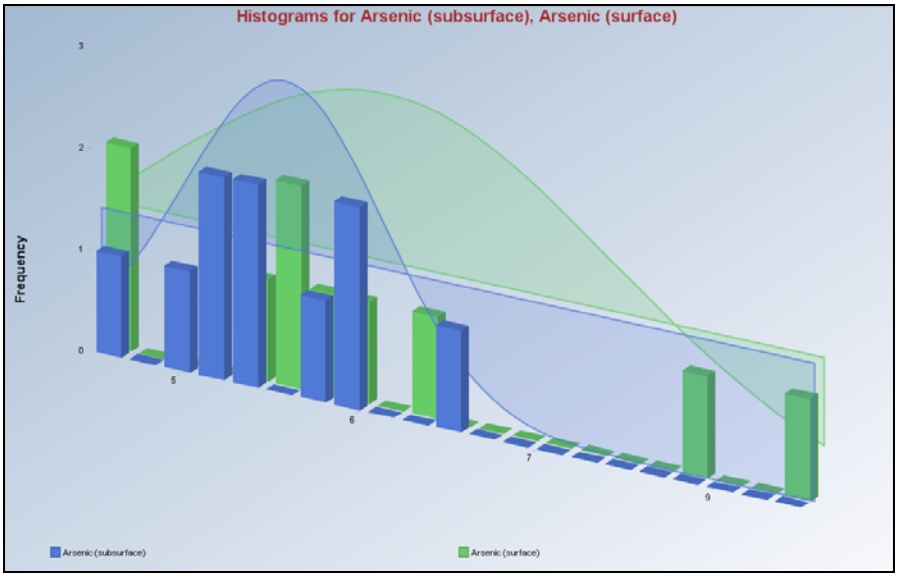


- The default selection for **Graph by Groups** is **Individual Graphs**. This option produces a histogram (or other graphs) separately for each selected variable. If multiple graphs or graphs by groups are desired, then check the radio button next to **Group Graphs**.

- The default option for **Graphical Display Options** is **Color Gradient**. If you want to use and import graphs in black and white into a document or report, then check the radio button next to **For Export (BW Printers)**.

- Click on the OK button to continue or on the Cancel button to cancel the Histogram (or other selected graphical) option.

Histogram Output Screen
Selected options: Group Graphs and Color Gradient.



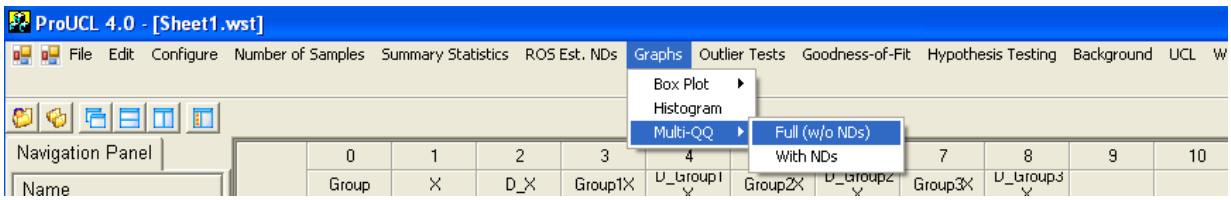
6.3 Multi-QQ

6.3.1 Multi-QQ (Full)

1. Click **Graphs** ► Multi-QQ

2. Multi-QQ can be obtained for data sets with (With NDs) and without NDs (Full).

   - When that option button is clicked, the following window will be shown.



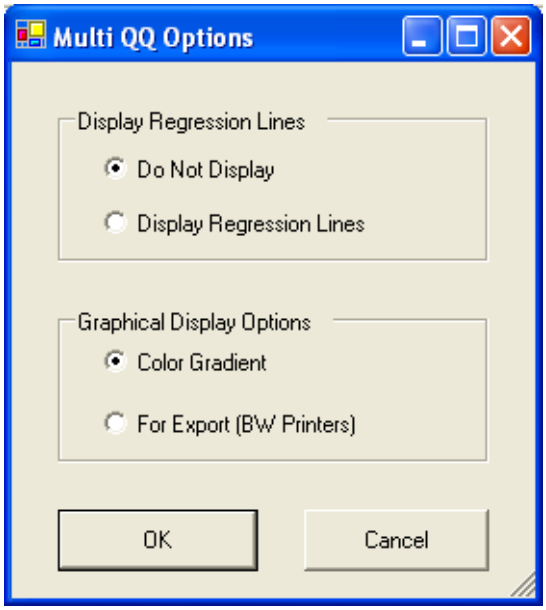
3. Select either Full or With NDs.

4. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.

   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

   - When the **Group by variable** option button is clicked, the following window will appear.



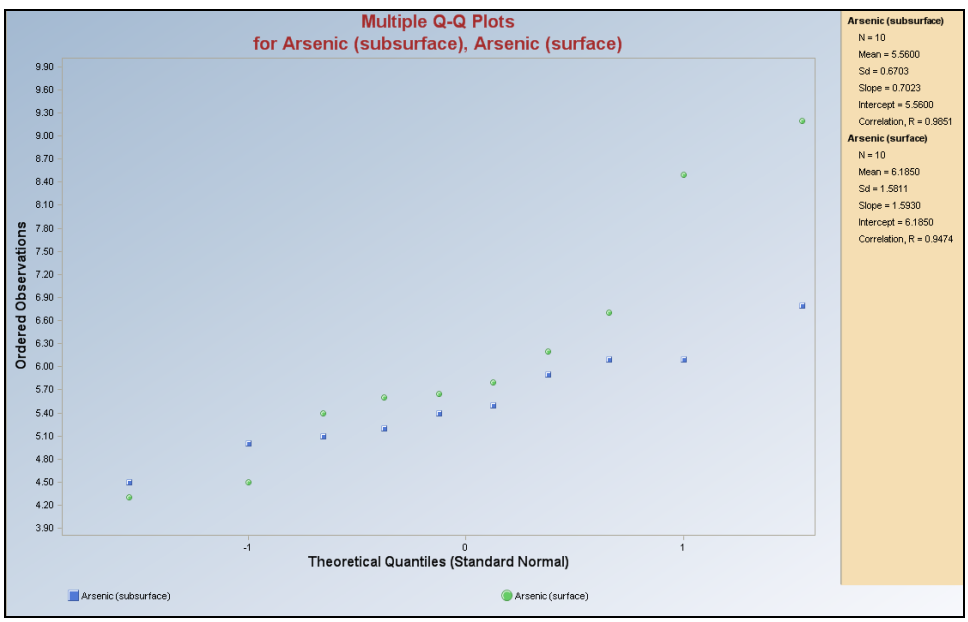
- The default option for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on graphs, then check the radio button next to **Display Regression Lines**.

- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, then check the radio button next to **For Export (BW Printers).**

- Click on the OK button to continue or on the Cancel button to cancel the selected Multi-QQ option.

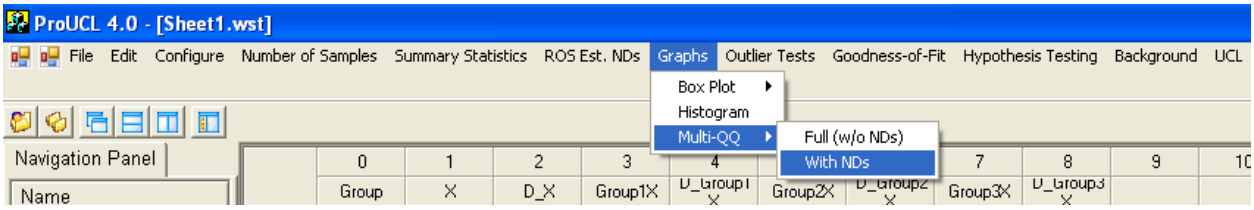
Note: *For Multi-QQ plot option, for both "Full" as well as for data sets "With NDs," the values along the horizontal axis represent quantiles of a standardized normal distribution (Normal distribution with mean 0 and standard deviation 1). Quantiles for other distributions (e.g., Gamma distribution) are used when using Goodness-of-Fit (GOF) test option.*

Output Screen for Multi-QQ (Full)
Selected options: Group Graph, Do Not Display Regression Lines, and Color Gradient.



6.3.2 Multi-QQ (with NDs)

1. Click **Graphs** ► Multi-QQ



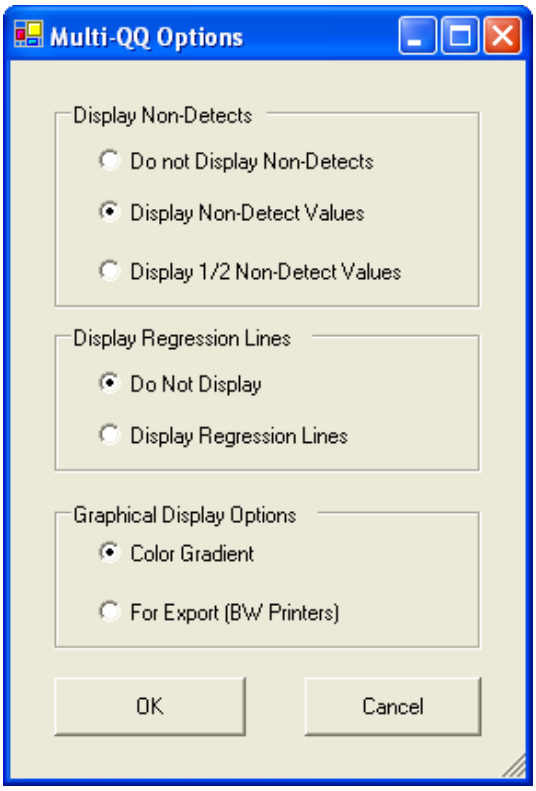
2. Select With NDs option by clicking on it.

3. The **Select Variables Screen** (Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.

- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

- When the **Group by variable** option button is clicked, the following screen appears.



- The default option for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines, then check the radio button next to **Display Regression Lines**.

- The default option for **Display Nondetects** is **Display Nondetect Values**.

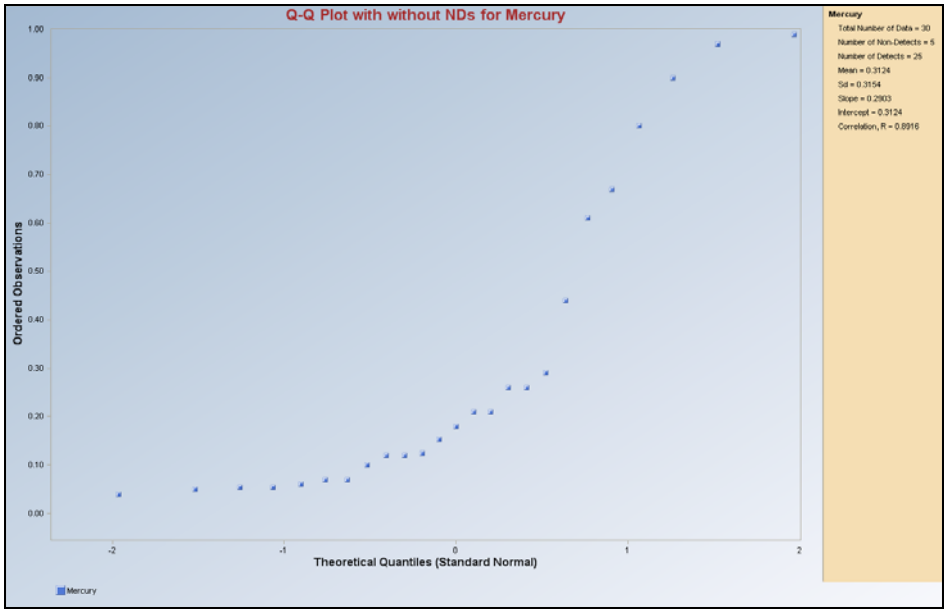
  - **Do not Display Nondetects**: Selection of this option excludes the NDs detects and plots only detected values on the associated Q-Q plot. The statistics are computed using only detected data.
  - **Display Nondetect Values**: Selection of this option treats detection limits as detected values and plots those detection limits and detected values on the Q-Q plot. The statistics are computed accordingly.

- **Display ½ Nondetect Values**: Selection of this option replaces the detection limits with their half values, and plots half detection limits and detected values on the Q-Q plot. The statistics are computed accordingly.

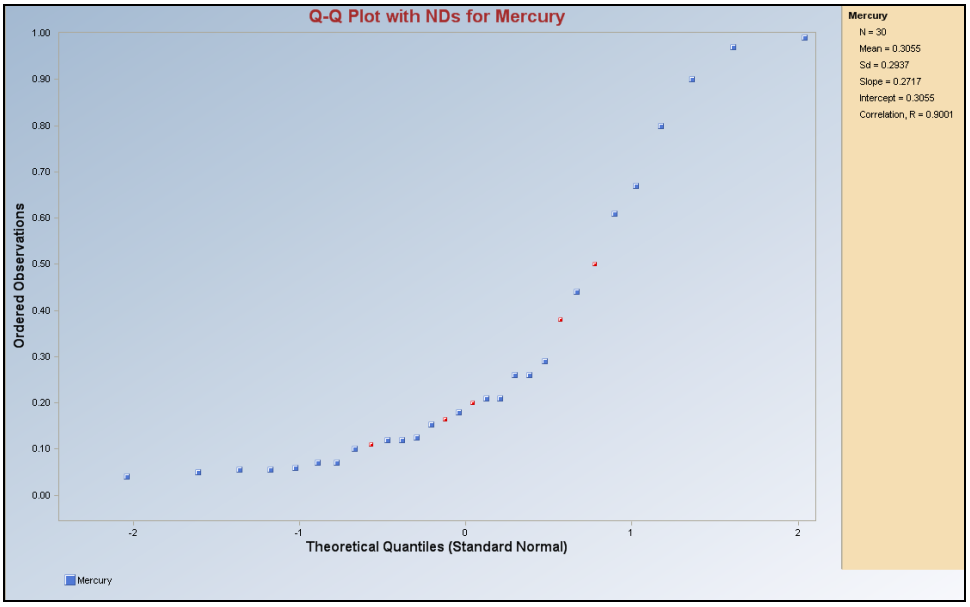
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, then check the radio button next to **For Export (BW Printers).**

- Click on the OK button to continue or on the Cancel button to cancel the Multi-QQ option.

Output Screen for Multi-QQ (without NDs)
Options: Do Not Display Regression Lines, Do not display Nondetects, and Color Gradient.

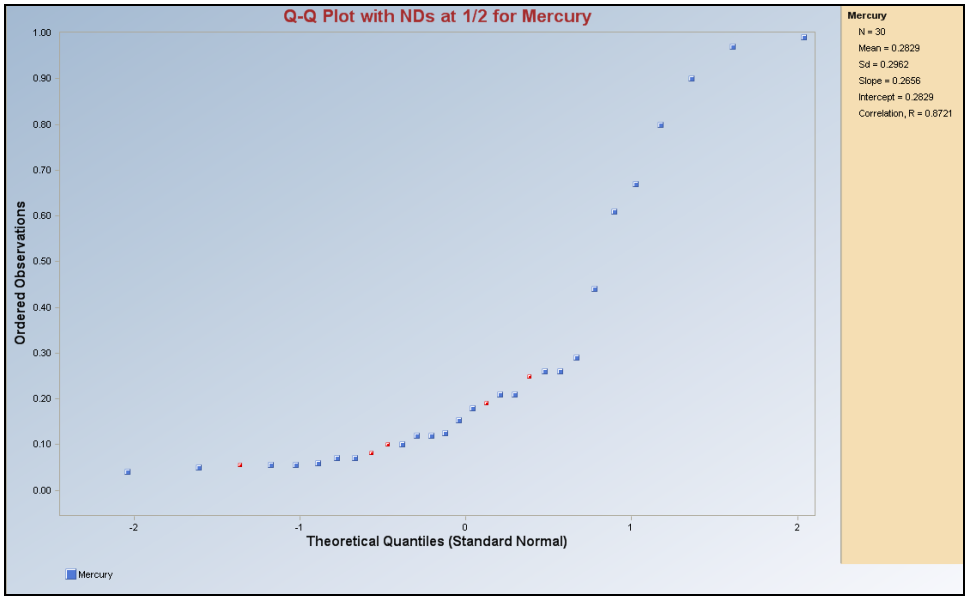


Output Screen for Multi-QQ (with NDs)
Options: Do not Display Regression Line, Display Nondetect Values, and Color Gradient.



Note: *The legend size of nondetect values is smaller than that of the detected values and is shown in red. The legend size is made smaller for BW printers.*

Output Screen for Multi-QQ (with NDs)
Selected options: Do not Display Regression Lines, Display ½ Nondetect Values, and Color Gradient.



Note: *The legend size of nondetect values is smaller than that of the detected values and is shown in red.*

Chapter 7

Simple Classical Outlier Tests

Outliers are inevitable in data sets originating from environmental applications. There are many graphical (Q-Q plots, Box plots), classical (Dixon, Rosner, Welch), and robust methods (biweight, Huber, PROP) available to identify outliers. It is well known that the classical outlier tests (e.g., Dixon test, Rosen test, EPA, 2006) suffer from masking (e.g., extreme outliers may mask intermediate outliers) effects. The use of robust outlier identification procedures is recommended to identify multiple outliers, especially when dealing with multivariate (having multiple contaminants) data sets. However, those preferred and more effective robust outlier identification methods are beyond the scope of ProUCL 4.0. Several robust outlier identification methods (e.g., based upon biweight, Huber, and PROP influence functions) are available in the Scout software package (EPA, 1999).

The two simple classical outlier tests (often cited in environmental literature): Dixon and Rosner tests are available in ProUCL 4.0. These tests can be used on data sets with and without nondetect observations. These tests also require the assumption of normality of the data set without the outliers. It should be noted that in environmental applications, one of the objectives is to identify high outlying observations that might be present in the right tail of a data distribution as those observations often represent contaminated locations of a polluted site. Therefore, for data sets with nondetects, two options are available in ProUCL 4.0 to deal with data sets with outliers. These options are: 1) exclude nondetects and 2) replace NDs by DL/2 values. These options are used only to identify outliers and not to compute any estimates and limits used in decision-making process.

It is suggested that these two classical outlier identification procedures be supplemented with graphical displays such as Q-Q plots, Box and Whisker plot (called box plot), and IQR (= upper quartile, Q3 - lower quartile, Q1). These graphical displays are available in ProUCL 4.0. Box plots with whiskers are often used to identify outliers (e.g., EPA, 2006). Typically, a box plot gives a good indication of extreme (outliers) observations that may present in a data set. The statistics (lower quartile, median, upper quartile, and IQR) used in the construction of a box plot do not get distorted by outliers. On a box plot, observations beyond the two whiskers may be considered as candidates for potential outliers.

Q-Q plots are also quite useful to identify outliers in a data set. For an example, on a normal Q-Q plot, observations that are well separated from the bulk (central part) of the data typically represent potential outliers needing further investigation. Also, significant and obvious jumps and breaks in a Q-Q plot (for any distribution) are indications of the presence of more than one population. Data sets exhibiting such behavior of Q-Q plots should be partitioned out in component sub-populations before estimating an EPC Term or a background threshold value (BTV). It is strongly recommended that both graphical and formal outlier identification tests should be used on the same data set to identify potential outliers that may be present in a data set under study. More details about the construction of graphical displays and outliers test can be found in the Technical Guide for ProUCL 4.0.

Dixon's Test (Extreme Value Test)

- This test is used to identify statistical outliers when the sample size is less than or equal to 25.

- This test can be used to identify outliers or extreme values in both the left tail (Case 1) and the right tail (Case 2) of a data distribution. In environmental data sets, extremes found in the right tail may represent potentially contaminated site areas needing further investigation or remediation. The extremes in the left tail may represent ND values.

- This test assumes that the data without the suspected outlier are normally distributed; therefore, it is necessary to perform a test for normality on the data without the suspected outlier before applying this test.

- This test may suffer from masking in the presence of multiple outliers. This means that if more than one outlier is suspected, this test may fail to identify all of the outliers. Therefore, if you decide to use the Dixon's test for multiple outliers, apply the test to the least extreme value first. Alternatively, use more effective robust outlier identification procedures. Those outlier identification procedures will be available in Scout (EPA, 1999) software.

Rosner's Test

- This test can be used to identify and detect up to 10 outliers in data sets of sizes 25 and higher.

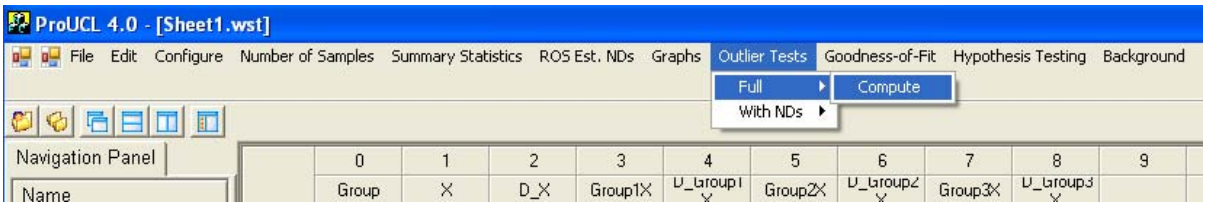
- This test also assumes that the data are normally distributed; therefore, it is necessary to perform a test for normality before applying this test.

Depending upon the selected variable(s) and the number of observations associated with them, either the Dixon's test or the Rosner's test will be performed.

NOTE: *Throughout this User Guide, and in ProUCL 4.0, it is assumed that the user is dealing with a single population. If multiple populations are present in a data set, it is recommended to separate them out using appropriate population partitioning methods and techniques. Appropriate tests and statistics (e.g., goodness-of-fit tests, 95% UCLs, 95% UPLs) should be computed separately for each of the identified populations. Also, outliers if any should be identified and thoroughly investigated. The presence of outliers distorts all statistics including the all of the upper limits (UCLs, UPLs, upper percentiles). The use of distorted statistics and limits may lead to incorrect conclusions having potential adverse effects on the human health and the environment. Decisions about the disposition of outliers: inclusion or exclusion in the data set to be used to compute the UCLs, UPLs, and other statistics should be made by all parties involved. Statistical methods supplemented with graphical displays (e.g., Q-Q plot and box plots) can only help identify statistical outliers that may be present in a data set. The project team and experts familiar with the site should interpret and assign physical meaning and significance to those identified outliers. The entire project team should be involved in taking decisions about the appropriate disposition (include or not include) of outliers*.

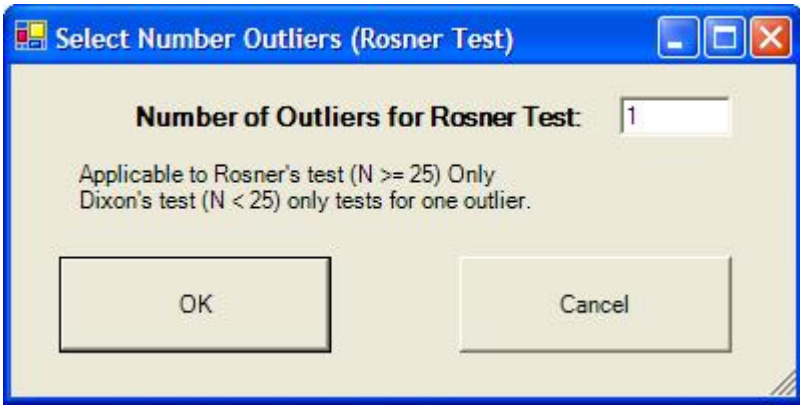
7.1 Outlier Test for Full Data Set

1. Click **Outlier Tests ► Full ►Compute**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.
- If at least one of the selected variables has 25 or more observations, then click the option button for the Rosner Test.



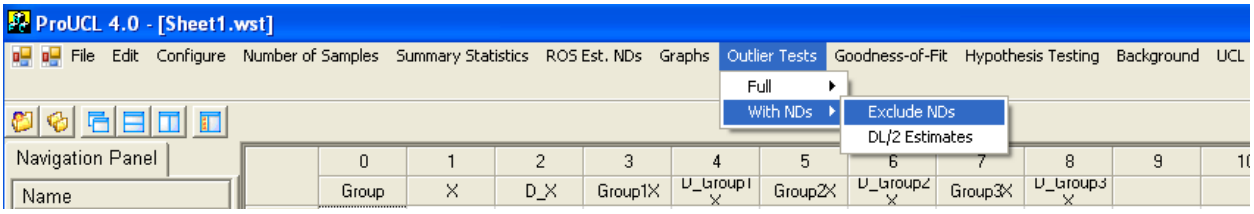
- The default option for the number of suspected outliers is 1. In order to use this test, the user has to obtain an initial guess about the number of outliers that may be present in the data set. This can be done by using graphical displays such as a Q-Q plot. On this graphical Q-Q plot, higher observations that are well separated from the rest of the data may be considered as potential or suspected outliers.
- Click on the OK button to continue or on the Cancel button to cancel the Outlier Tests.

7.2 Outlier Test for Data Set with NDs

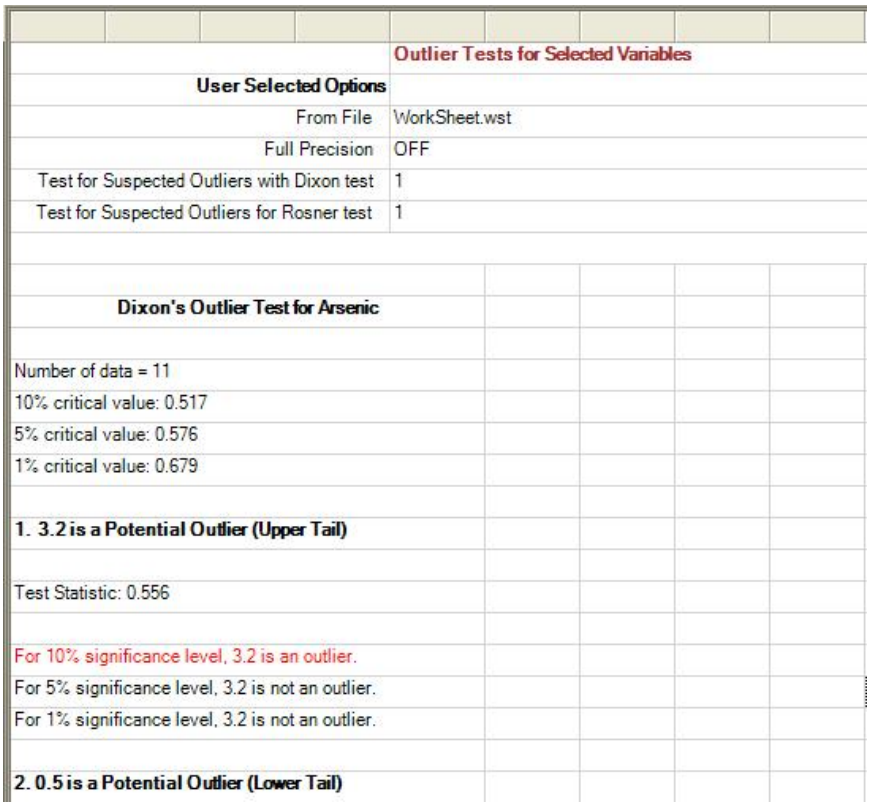
Typically, in environmental applications, one is interested in identifying high outliers (perhaps representing contaminated parts of a site area, hot spots) that might be present in the right tail of the data distribution. Therefore, one may want to use the same outlier identification procedures (e.g., Dixon test,

Rosner test) that are used on full-uncensored data sets (without any NDs). The processes to perform such tests using ProUCL 4.0 are described as follows.

1. Click **Outlier Tests ► With NDs ► Exclude NDs**



Output Screen for Dixon's Outlier Test



	Outlier Tests for Selected Variables
User Selected Options	
From File	WorkSheet.wst
Full Precision	OFF
Test for Suspected Outliers with Dixon test	1
Test for Suspected Outliers for Rosner test	1

Dixon's Outlier Test for Arsenic

Number of data = 11
10% critical value: 0.517
5% critical value: 0.576
1% critical value: 0.679

1. 3.2 is a Potential Outlier (Upper Tail)

Test Statistic: 0.556

For 10% significance level, 3.2 is an outlier.
For 5% significance level, 3.2 is not an outlier.
For 1% significance level, 3.2 is not an outlier.

2. 0.5 is a Potential Outlier (Lower Tail)

Output Screen for Rosner's Outlier Test

There are many observations in this data set that may represent potential outliers. The potential outliers can also be seen in the graphical displays associated with this data set.

- Selected Options: Number of Suspected Outliers for the Rosner Test = 4

	Outlier Tests for Selected Variables
User Selected Options	
From File	D:\Narain\ProUCL 4.0\Data\Aroclor 1254.wst
Full Precision	OFF
Test for Suspected Outliers with Dixon test	1
Test for Suspected Outliers for Rosner test	2

Rosner's Outlier Test for Aroclor1254

Number of data: 44

Number of suspected outliers: 2

#	Mean	sd	Potential outlier	Test value	Critical value (5%)	Critical value (1%)
1	1531.88	3316.59	19000.00	5.27	3.08	3.43
2	1125.65	1998.60	8300.00	3.59	3.07	3.41

For 5% significance level, there are 2 Potential Outliers

Therefore, Potential Statistical Outliers are

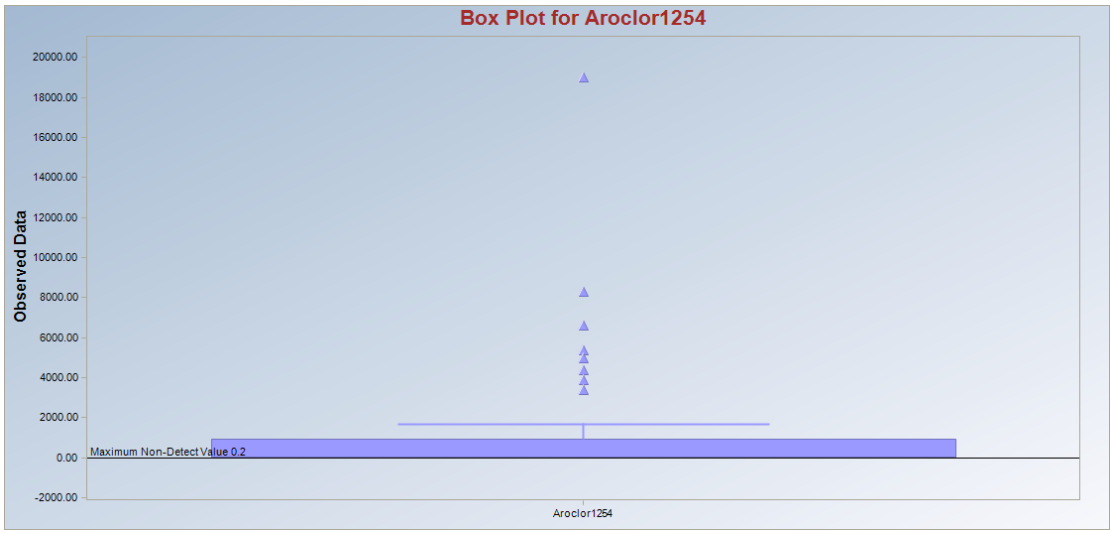
19000.00, 8300.00

For 1% Significance Level, there are 2 Potential Outliers

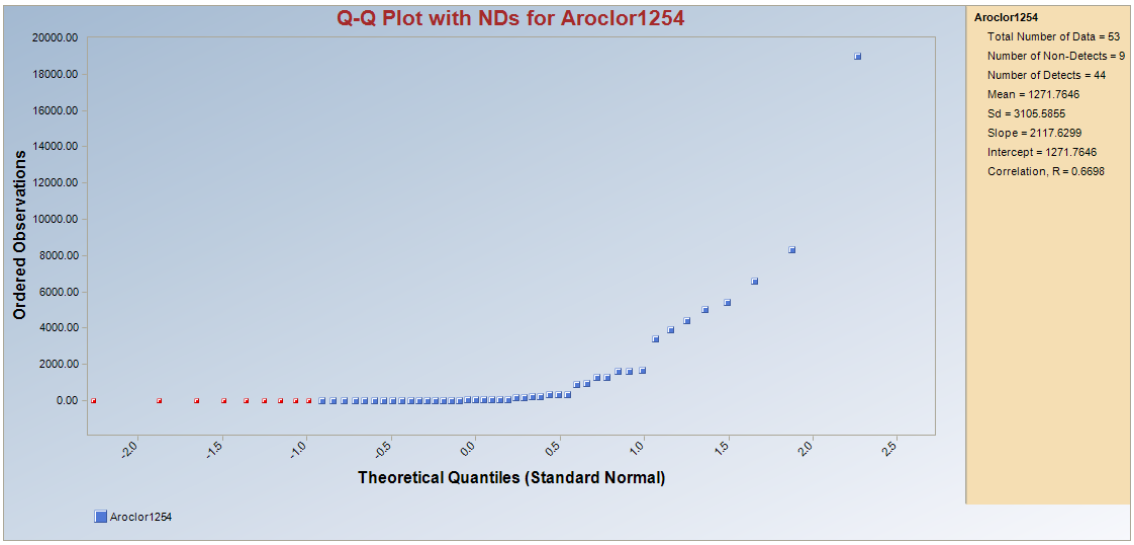
Therefore, Potential Statistical Outliers are

19000.00, 8300.00

Box plot of the Aroclor1254 Data Set



Q-Q Plot of the Aroclor1254 Data Set



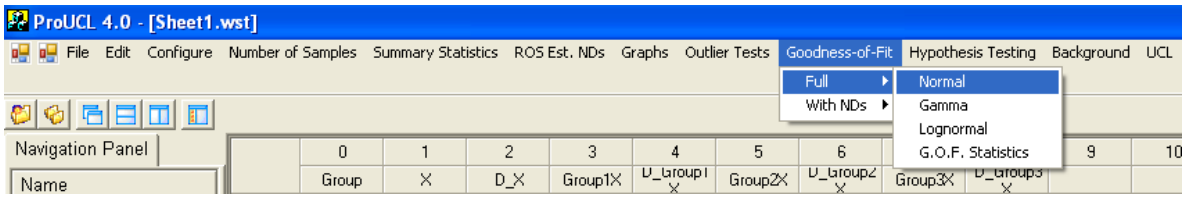
Chapter 8

Goodness-of-Fit (G.O.F.) Tests

Several goodness-of-fit (G.O.F.) tests for full data sets (without nondetects) and for data sets with NDs are available in ProUCL 4.0. Details of those tests are described in the ProUCL 4.0 Technical Guide. In this User Guide, those tests and available options have been illustrated using screen shots generated by ProUCL 4.0.

Two choices are available for Goodness-of-Fit menu: Full and With NDs.

- **Full**
    - This option is used to analyze full data sets without any nondetect observations. Throughout this User Guide and in ProUCL 4.0, "Full" represents data sets without nondetect observations.
    - This option tests for normal, gamma, or lognormal distribution of the variable(s) selected using the Select Variables option.
    - G.O.F. Statistics: This option is available for both full data sets and for data sets with NDs. This option simply generates output log of GOF test statistics and derived conclusions about the data distributions of all selected variables. This option is also available for variables categorized by a group variable.



- **With NDs**
    - Analyzes data sets that have both nondetected and detected values.
    - Six sub-menu items listed and shown below are available for this option.

1. **Exclude NDs**: tests for normal, gamma, or lognormal distribution of the selected variable(s) using only the detected values.

2. **ROS Estimates**: tests for normal, gamma, or lognormal distribution of the selected variable(s) using the detected values and the extrapolated values for the nondetects.

    - Three ROS methods for normal, lognormal, and gamma distributions are available. This option is used to estimate or extrapolate the NDs based upon the specified distribution.

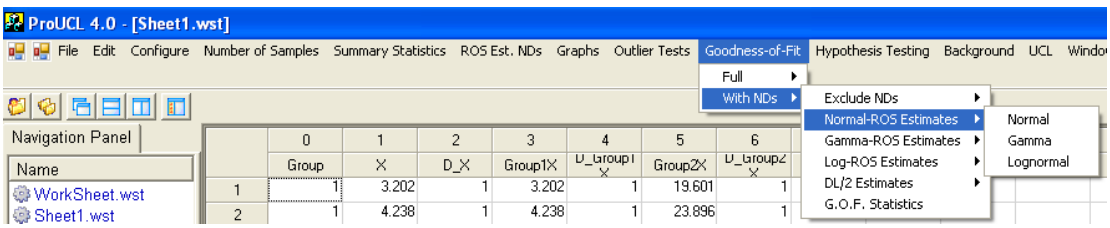
    - By using the menu item ROS Est. NDs, ProUCL 4.0 actually generates additional column(s) of ROS estimated NDs based upon the selected distribution. This

option should be used for variables with NDs. This is further illustrated by a screen shot given in the following

3. **DL/2 Estimates**: tests for normal, gamma, or lognormal distribution of the selected variable(s) using the detected values and the ND values replaced by their respective DL/2 values. This option is included for historical reasons and also for curious users.

   Note: *The use of fabricated data obtained using DL/2 (DL or 0 values) values is not a recommended method. At best, the user may use the fabricated data (e.g., DL/2 option) for exploratory reasons. It is suggested that these substitution methods should not be used for estimation and for hypotheses testing approaches.*

4. **G.O.F. Statistics:** As for the full data sets, this option simply generates output log of GOF test statistics and other relevant statistics for data sets with nondetects. The conclusions about the data distributions for all selected variables are also listed on the generated output file (Excel-type spreadsheet). This option is also available for variables categorized by a group variable.



- When multiple variables are selected from the Select Variables screen, you can choose either:

  - Group Graph option to produce multiple Q-Q plots for all selected variables in a single graph. The relevant statistics (e.g., slope, intercept, correlation, test statistic and critical value) associated with the selected variables are shown on the right panel (in tan color; see page 77). In order to capture all the graphs and results shown on the window screen, it is preferable to print the graph using the Landscape option. *The use of this option is recommended when a selected variable has data coming from two or more groups or populations.*
  - Group Graph option is particularly useful to generate multiple Q-Q plots for the groups associated with a selected variable. In order to capture all the graphs and results shown on the window screen, it is preferable to print the graph using the Landscape option. The user may also want to turn off the Navigation Panel and Log Panel.

- Individual Graph option is used to generate individual Q-Q plots and the associated statistics separately for each of the selected variable, one variable at a time.

  - The linear pattern displayed by a Q-Q plot suggests an approximate goodness-of-fit for the selected distribution.
  - The program computes the intercept, slope, and the correlation coefficient for the linear pattern displayed by the Q-Q plot. A high value of the correlation coefficient

(e.g., > 0.95) is an indication of a good fit for that distribution. This high correlation should exhibit a definite linear pattern in the Q-Q plot. Specifically, when data are sparse and correlation is high, the use of correlation statistic to determine data distribution is not desirable. Note that these statistics are displayed on the Q-Q plot.

- On a Q-Q plot, observations that are well separated from the bulk (central part) of the data typically represent potential outliers needing further investigation.
- Significant and obvious jumps and breaks in a Q-Q plot (for any distribution) are indications of the presence of more than one population. Data sets exhibiting such behavior of Q-Q plots should be partitioned out in component sub-populations before estimating an EPC term or a background threshold value (BTV). It is strongly recommended that both graphical and formal goodness-of-fit tests should be used on the same data set to determine the distribution of the data set under study.

- **Normality or Lognormality Tests**: In addition to informal graphical normal and lognormal Q-Q plots, a formal Goodness-of-Fit (GOF) test is also available to test the normality or lognormality of the data set.

  - Lilliefors Test: a test typically used for samples of size larger than 50 (> 50). However, the Lilliefors test (generalized Kolmogorov Smirnov test) is available for samples of all sizes. There is no applicable upper limit for sample size for the Lilliefors test.
  - Shapiro and Wilk (SW) Test: a test used for samples of size smaller than or equal to 2000 (<= 2000). In ProUCL 4.0, the SW test uses the SW Critical values for samples of size 50 or less. For samples of size, greater than 50, the SW Test statistic is displayed along with the P-value of the test (Royston, 1982). These values are not as yet available in ProUCL 4.0. This extension of SW test will be available in Scout (EPA, 1999) software package.
  - It should be noted that sometimes these two tests might lead to different conclusions. Therefore, the user should exercise caution interpreting the results. Specifically, the user should the pattern exhibited by the associated Q-Q plot.

- **GOF test for Gamma Distribution:** In addition to the graphical gamma Q-Q plot, two formal empirical distribution function (EDF) procedures are also available to test the gamma distribution of a data set. These tests are the Anderson-Darling test and the Kolmogorov-Smirnov test.

  - It is noted that these two tests might lead to different conclusions. Therefore, the user should exercise caution interpreting the results.
  - These two tests may be used for samples of sizes in the range of 4-2500. Also, for these two tests, the value of the shape parameter, *k* (*k* hat) should lie in the interval [0.01, 100.0]. Consult the ProUCL 4.0 Technical Guide (A. Singh and A.K. Singh (EPA, 2007)) for a detailed description of gamma distribution and its parameters, including *k*. Extrapolation beyond these sample sizes and values of *k* is not recommended.

- ProUCL computes the relevant test statistic and the associated critical value, and prints them on the associated Q-Q plot. On this Q-Q plot, the program informs the user if the data are gamma, normally, or lognormally distributed.

- Even though, the G.O.F. Statistics option prints out all GOF test statistics for all selected variables, it is suggested that the user should look at the graphical Q-Q plot displays to gain extra insight (e.g., outliers, multiple population) into the data set.

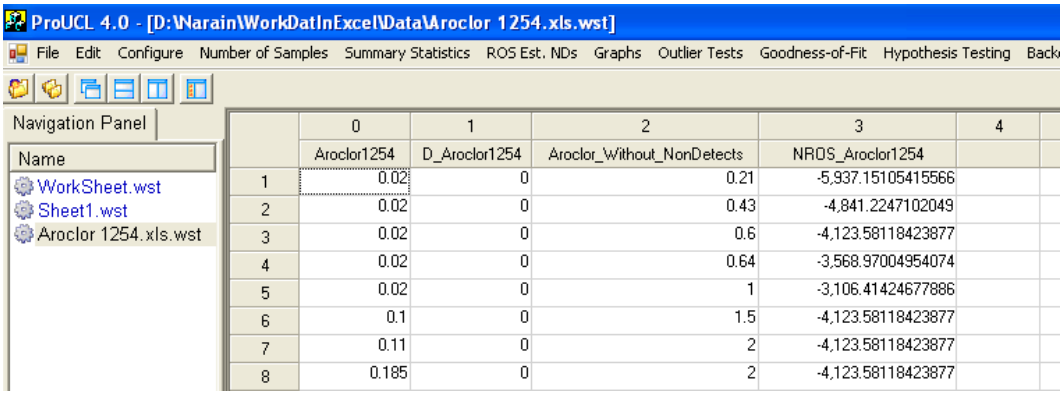
Note: *It is highly recommended not to skip the use of a graphical Q-Q plot to determine the data distribution as a Q-Q plot also provides a useful information about the presence of multiple populations or outliers.*

8.1 ROS Estimated (Est.) NDs – Saving Extrapolated NDs

- As mentioned before, for a variable with NDs, ProUCL 4.0 can generate additional column(s) consisting of detected data and the estimated (extrapolated) values of NDs using the ROS method assuming a normal, lognormal, or a gamma distribution.

- The user may want to use the resulting full data set (detected and estimated NDs) thus obtained to compute the statistics of interest such as a bootstrap BCA UCL95 or a gamma 95% upper percentile.

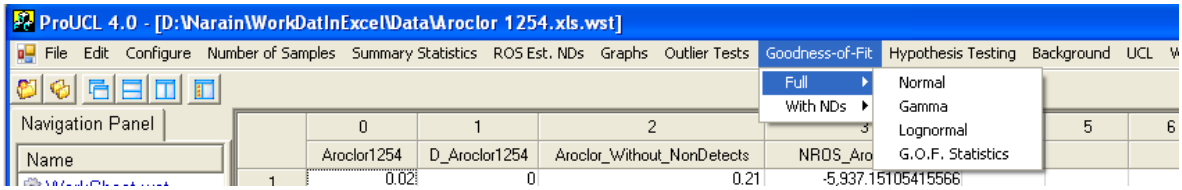
- This option of saving estimated NDs is provided only for experienced users and researchers. It is expected that the user knows and understands the theory behind these methods. Therefore, it is suggested that this option be used with care. For an example, often, the use of a ROS method yields infeasible (e.g., negative, exceeding the DLs) estimates of NDs, and therefore, the associated estimates of EPC terms and of BTVs may be biased and not reliable. This is especially true when the data set contains potential outlier(s).



	0	1	2	3	4
	Aroclor1254	D_Aroclor1254	Aroclor_Without_NonDetects	NROS_Aroclor1254	
1	0.02	0	0.21	-5,937.15105415566	
2	0.02	0	0.43	-4,841.2247102049	
3	0.02	0	0.6	-4,123.58118423877	
4	0.02	0	0.64	-3,568.97004954074	
5	0.02	0	1	-3,106.41424677886	
6	0.1	0	1.5	-4,123.58118423877	
7	0.11	0	2	-4,123.58118423877	
8	0.185	0	2	-4,123.58118423877	

8.2 Goodness-of-Fit Tests with Full Data Sets

1. Click **Goodness-of-Fit ► Full**

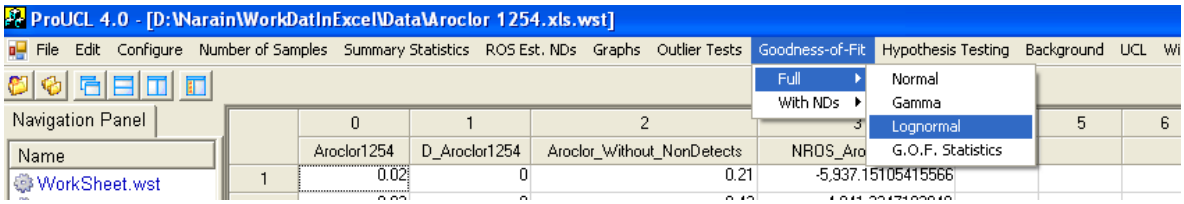


2. Select the distribution to be tested: Normal, Lognormal, or Gamma

- To test your variable for normality, click on Normal from the drop-down menu list.
- To test variable for lognormality, click on Lognormal from the drop-down menu list.
- To test your variable for gamma distribution, click on Gamma from the drop-down menu list.

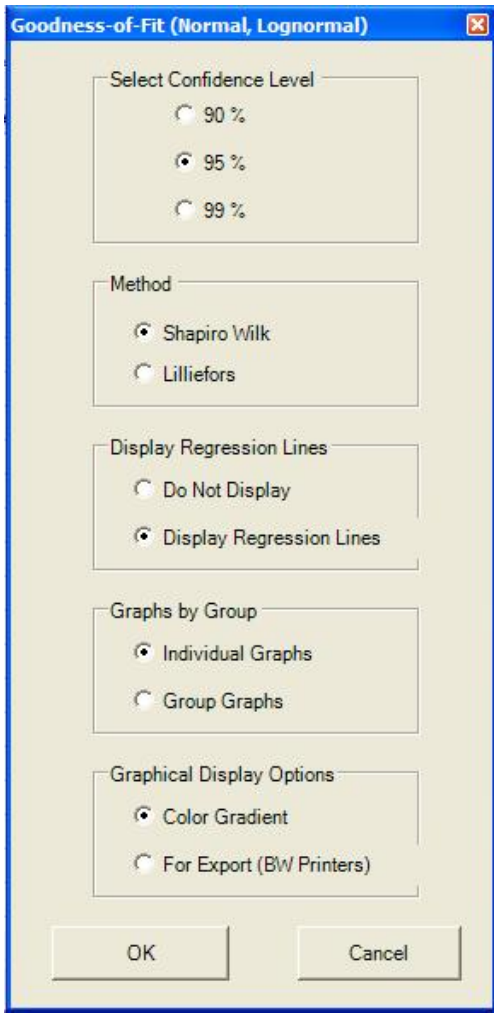
8.2.1 GOF Tests for Normal and Lognormal Distribution

1. Click **Goodness-of-Fit ► Full ► Normal or Lognormal**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.
- When the option button is clicked, the following window will be shown.



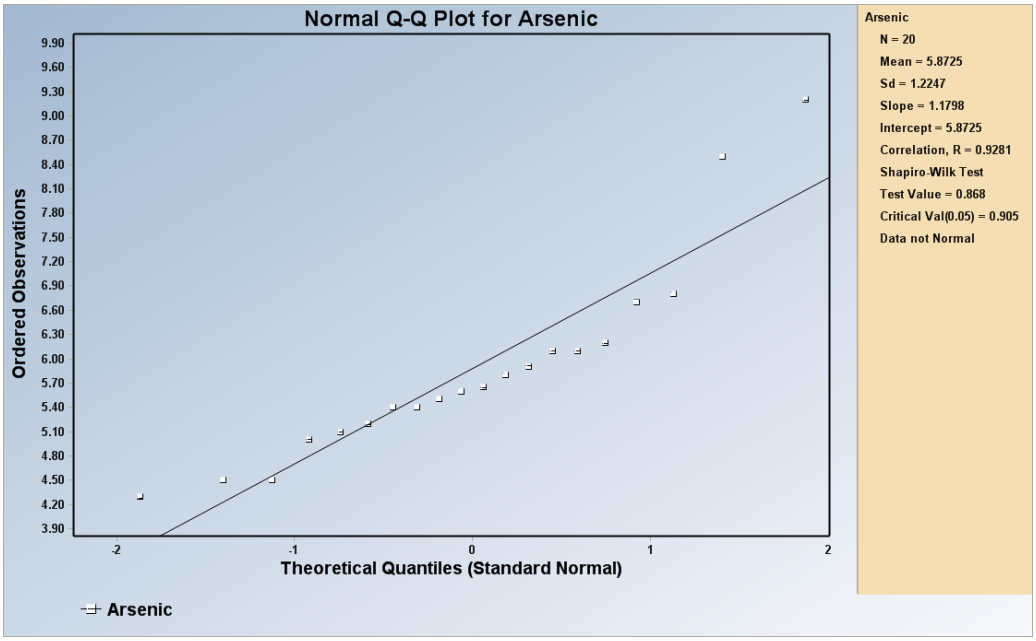
- The default option for the **Confidence Level** is **95%**.
- The default GOF **Method** is **Shapiro Wilk**. If the sample size is greater than 50, the program automatically uses the **Lilliefors** test.
- The default method for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on a Q-Q plot, then check the radio button next to **Display Regression Lines**.
- The default option for **Graphs by Group** is **Individual Graphs**. If you want to see the plots for all selected variables on a single graph, then check the radio button next to **Group Graphs**.

Note: *This option for Graphs by Group is specifically provided when the user wants to display multiple graphs for a variable by a group variable (e.g., site AOC1, site AOC2, background). This kind of display represents a useful visual comparison of the values of a variable (e.g., concentrations of COPC-Arsenic) collected from two or more groups (e.g., upgradient wells, monitoring wells, residential wells).*

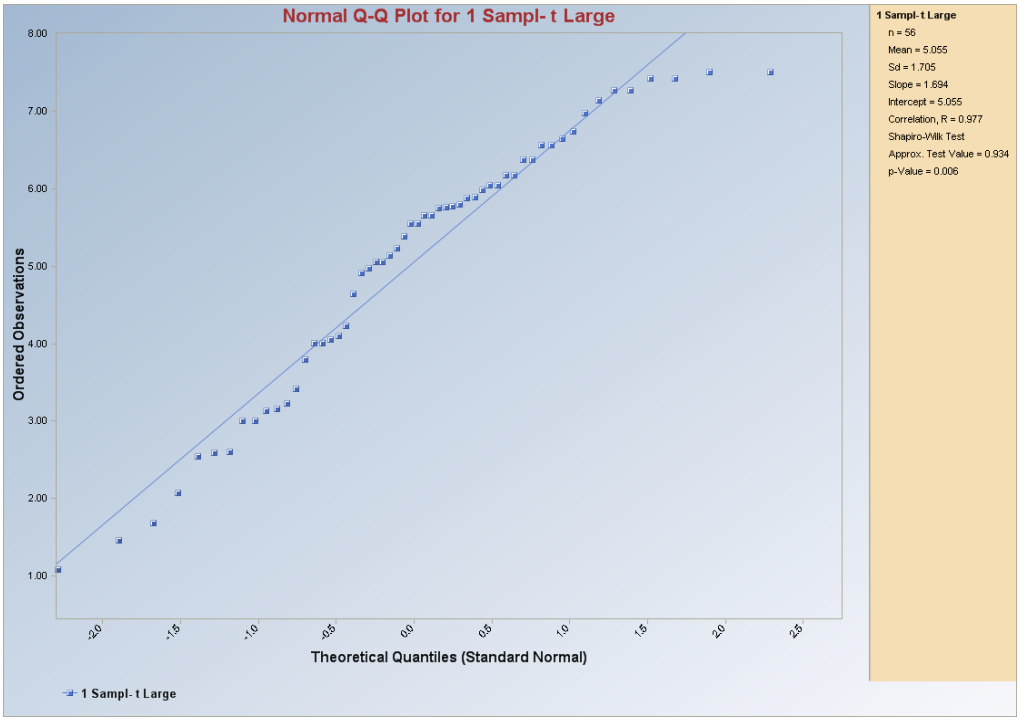
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white to be included in reports for later use, then check the radio button next to **For Export (BW Printers)**.

- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

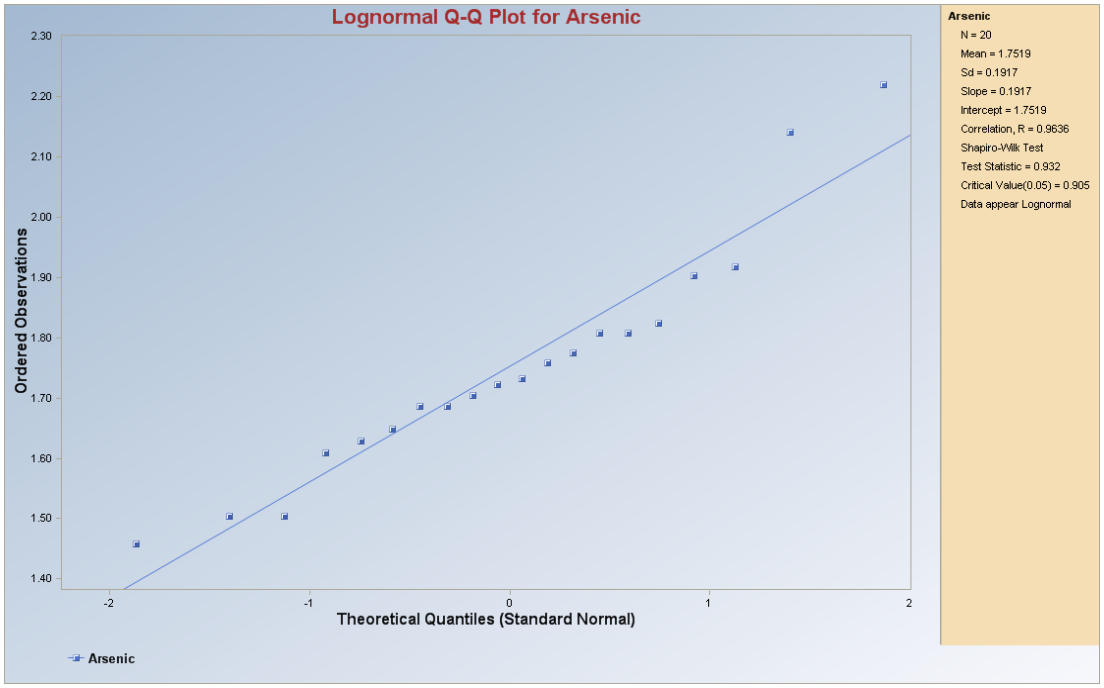
Output Screen for Normal Distribution (Full)
Selected Options: Shapiro Wilk, Display Regression Line, and For Export (BW Printers).



Output Screen for Normal Distribution (Full)
Selected Options: Shapiro Wilk (n > 50), Display Regression Line, and For Export (BW Printers).

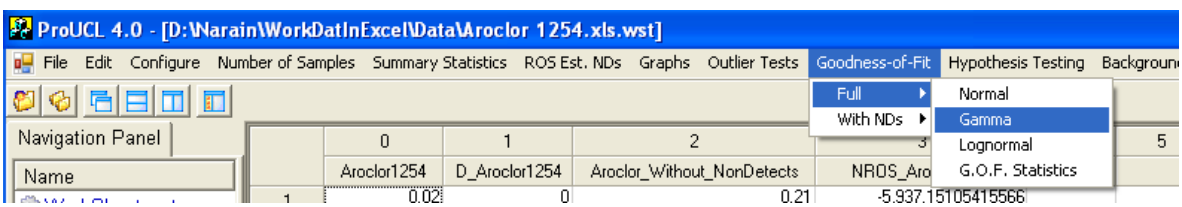


Output Screen for Lognormal Distribution (Full)
Selected options: Shapiro Wilk, Display Regression Lines, and Color Gradient.



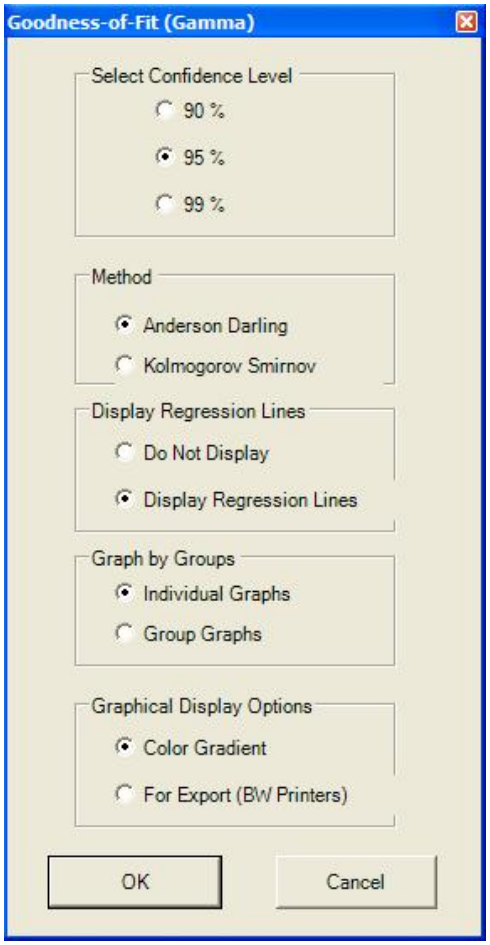
8.2.2 GOF Tests for Gamma Distribution

1. Click **Goodness-of-Fit ► Full ► Gamma**



2. The **Select Variables Screen** (described in Chapter 3) will appear.

- Select one or more variable(s) from the Select Variables screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.
- When the option button is clicked, the following window will be shown.

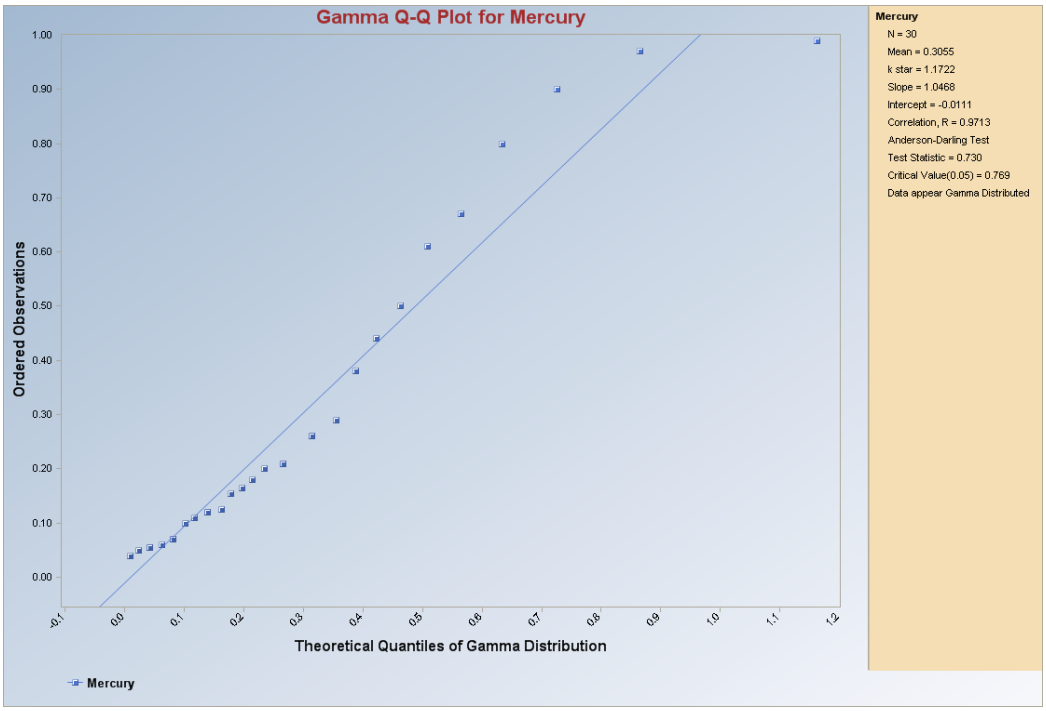


- The default option for the **Confidence Level** is **95%**.
- The default GOF method is **Anderson Darling**.
- The default option for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the Gamma Q-Q plot, then check the radio button next to **Display Regression Lines**.
- The default option for **Graph by Groups** is **Individual Graphs**. If you want to see the graphs for all the selected variables into a single graph, then check the radio button next to **Group Graphs**.
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**.

- Click the OK button to continue or the Cancel button to cancel the option.

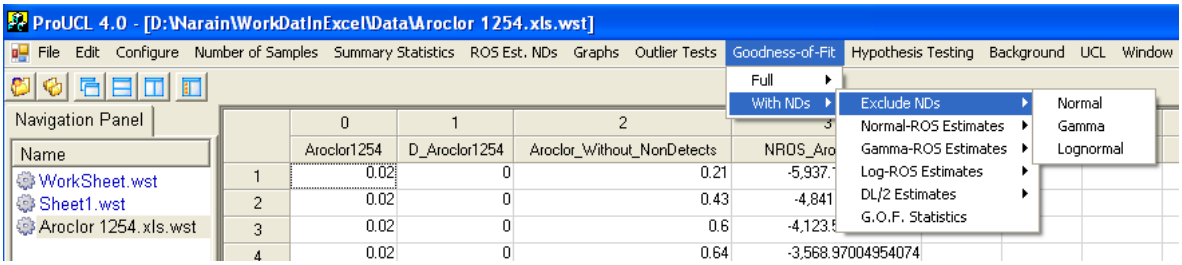
- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

Output Screen for Gamma Distribution (Full)
Selected options: Anderson Darling, Display Regression Lines, Individual Graphs, and Color Gradient.



8.3 Goodness-of-Fit Tests Excluding NDs

1. Click **Goodness-of-Fit ► With NDs ► Exclude NDs**

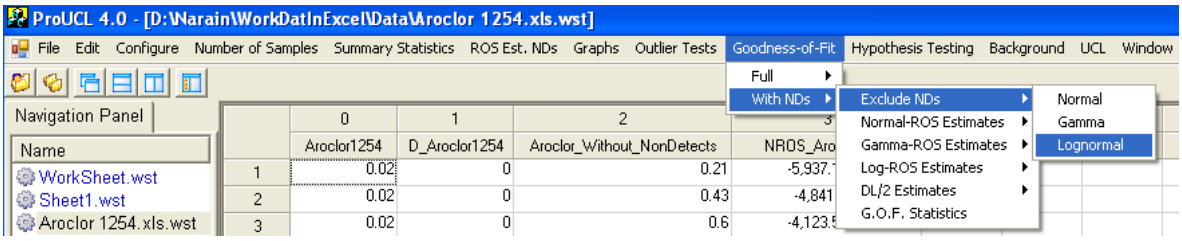


2. Select distribution to be tested: Normal, Gamma, or Lognormal.

- To test for normality, click on Normal from the drop-down menu list.
- To test for lognormality, click on Lognormal from the drop-down menu list.
- To test for gamma distribution, click on Gamma from the drop-down menu list.

8.3.1 Normal and Lognormal Options

1. Click **Goodness-of-Fit ► With NDs ►Excluded NDs ► Normal or Lognormal**

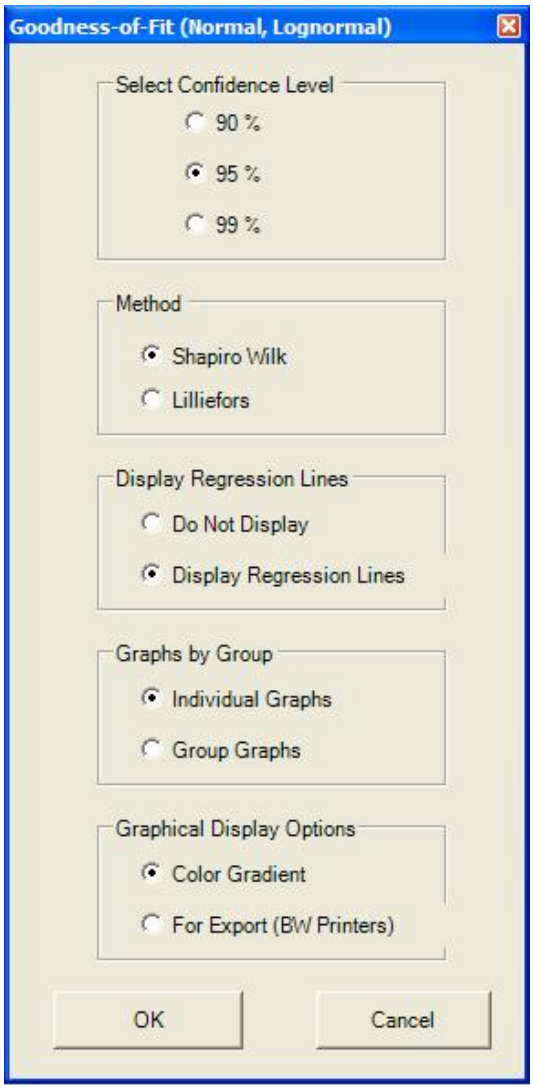


2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.

   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

   - When the option button (Normal or Lognormal) is clicked, the following window will be shown.



- The default option for the **Confidence Level** is **95%**.
- The default GOF **Method** is **Shapiro Wilk**. If the sample size is greater than 50, the program defaults to **Lilliefors** test.
- The default for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the associated Q-Q plot, check the radio button next to **Display Regression Lines**.
- The default option for **Graphs by Group** is **Individual Graphs**. If you want to see the plots for all selected variables on a single graph, check the radio button next to **Group Graphs**.

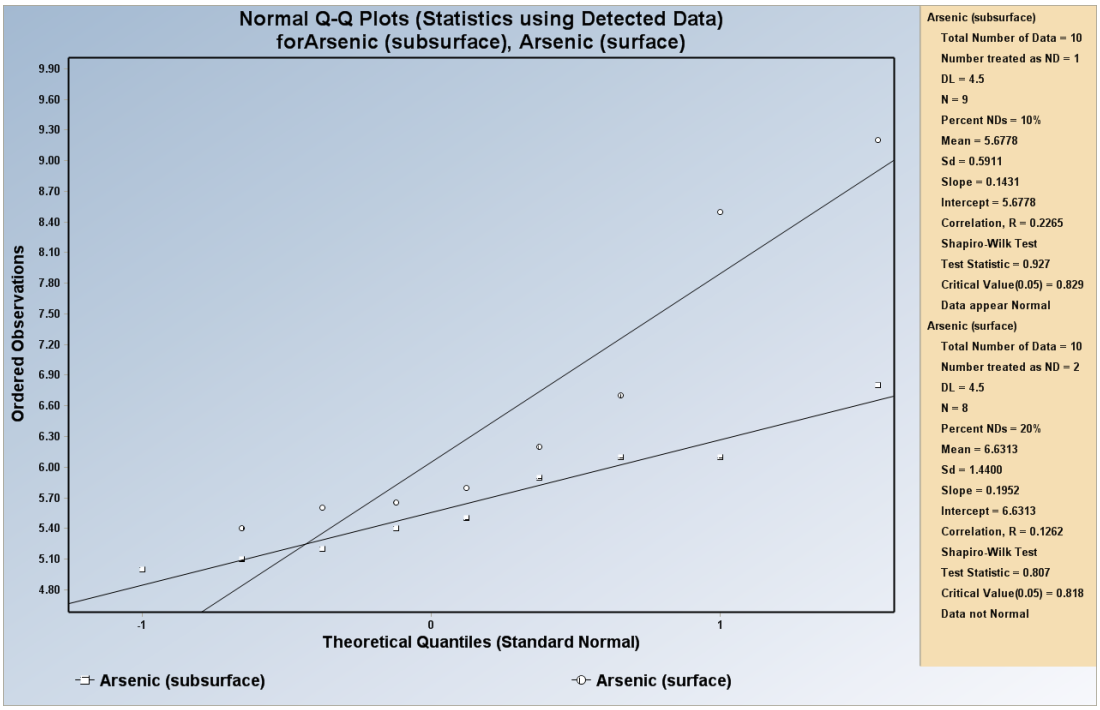
Note: *This option for Graphs by Group is specifically useful when the user wants to display multiple graphs for a variable by a group variable (e.g., site AOC1, Site AOC2, background). This kind of display represents a useful visual comparison of the values of a variable (e.g., concentrations of COPC-Arsenic) collected from two or more groups (e.g., upgradient wells, monitoring wells, and residential wells).*

- The default option for Graphical Display Option is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**.

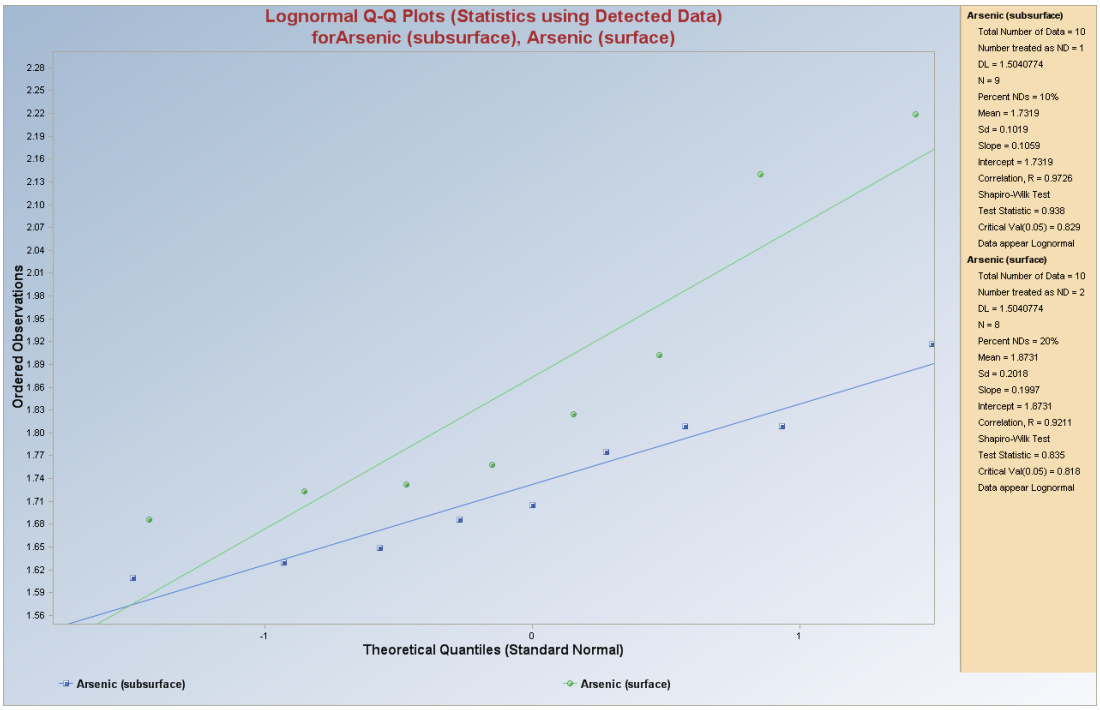
- Click the OK button to continue or the Cancel button to cancel the option.

- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

Output Screen for Normal Distribution (Exclude NDs)
Selected options: Shapiro Wilk, Display Regression Lines, Group Graphs, and For Export (BW Printers).

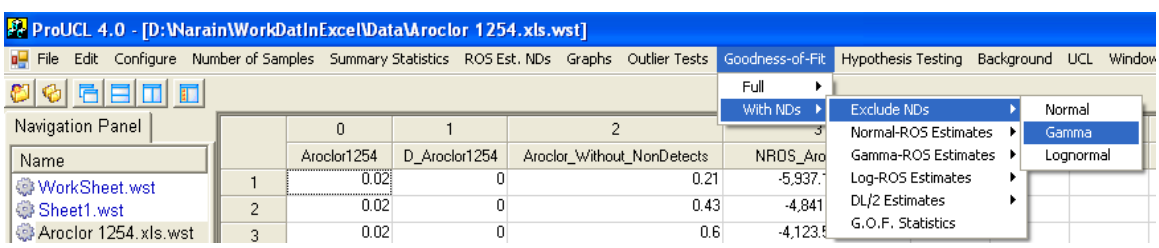


Output Result for Lognormal Distribution (Exclude NDs)
Selected options: Shapiro Wilk, Display Regression Lines, Group Graphs, and Color Gradient.



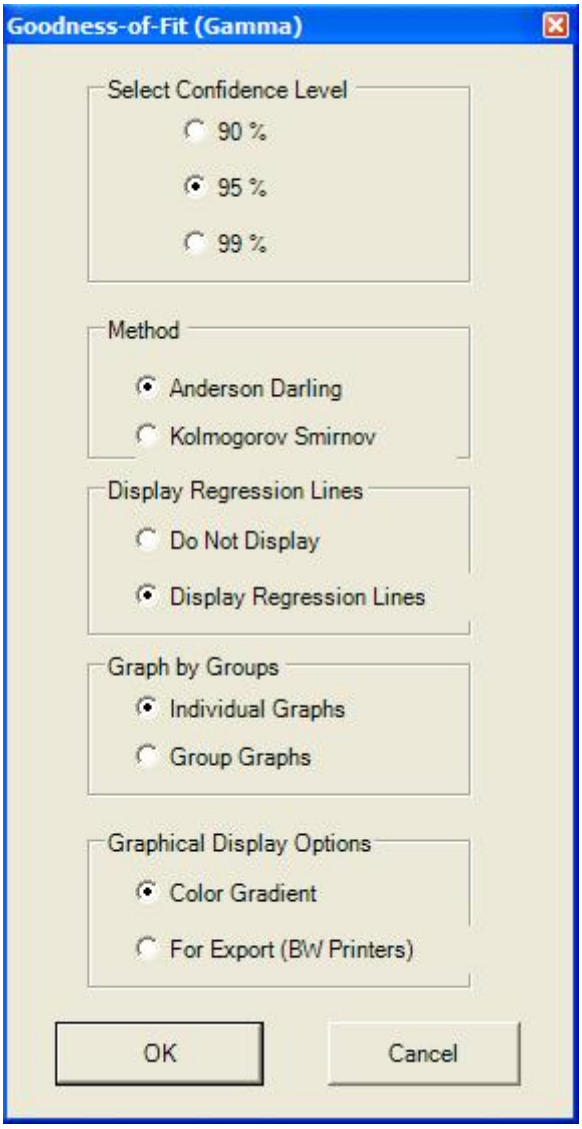
8.3.2 Gamma Distribution Option

1. Click **Goodness-of-Fit ► With NDs ►Excluded NDs ► Gamma**



2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.
   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.
   - When the option button (Gamma) is clicked, the following window is shown.

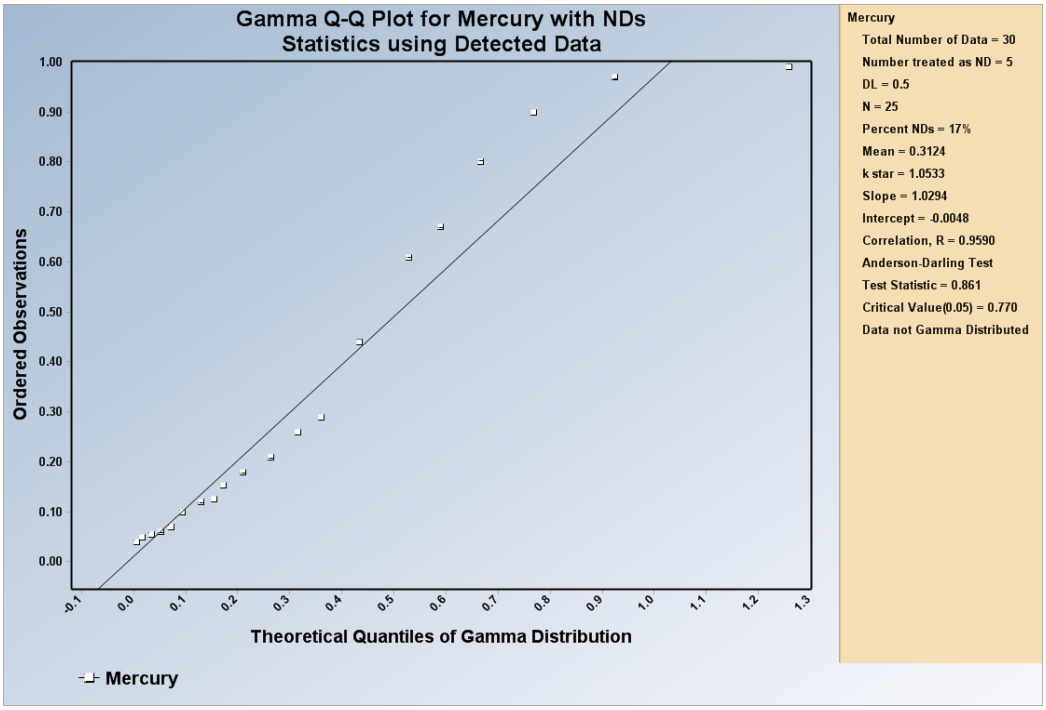


- The default option for the **Confidence Level** is **95%**.
- The default GOF test **Method** is **Anderson Darling**.
- The default method for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the normal Q-Q plot, check the radio button next to **Display Regression Lines**.
- The default option for **Graph by Groups** is **Individual Graphs**. If you want to display all selected variables on a single graph, check the radio button next to **Group Graphs**.
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**.

- Click the OK button to continue or the Cancel button to cancel the option.

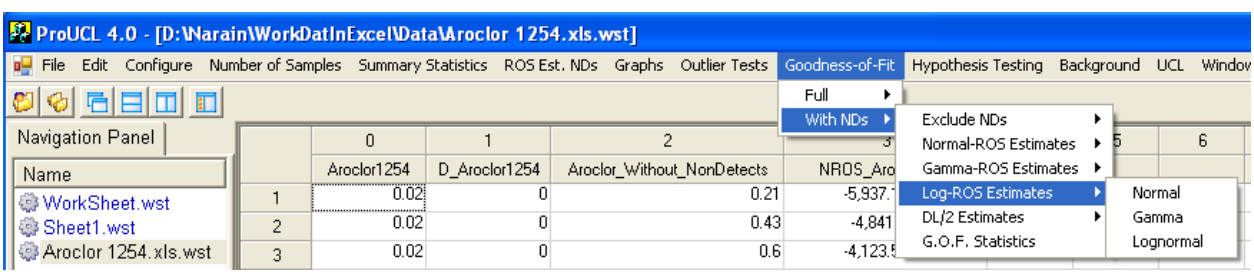
- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

Output Screen for Gamma Distribution (Exclude NDs)
Selected options: Anderson Darling, Do Not Display, Individual Graphs, and For Export (BW Printers).



8.4 Goodness-of-Fit Tests with Log-ROS Estimates

1. Click **Goodness-of-Fit ► With NDs ► Log-ROS Estimates**

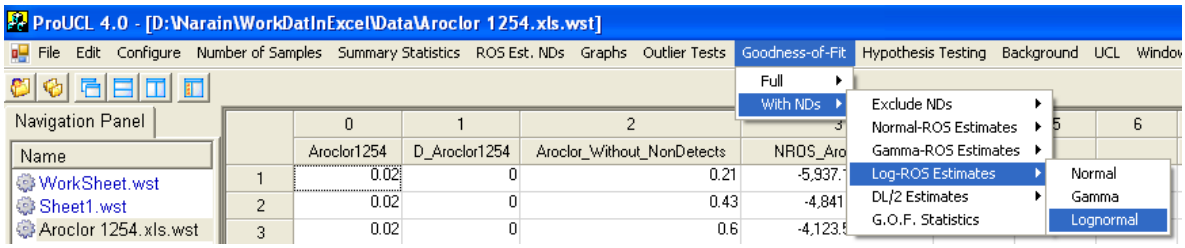


2. Select the distribution to be tested: Normal, Lognormal, or Gamma

- To test your variable for normality, click on Normal from the drop-down menu list.
- To test a variable for gamma distribution, click Gamma from drop-down menu list.
- To test your variable for lognormality, click on Lognormal from drop-down menu.

8.4.1 Normal or Lognormal Distribution (Log-ROS Estimates)

1. Click **Goodness-of-Fit ► With NDs ► Log-ROS Estimates ► Normal, Lognormal**

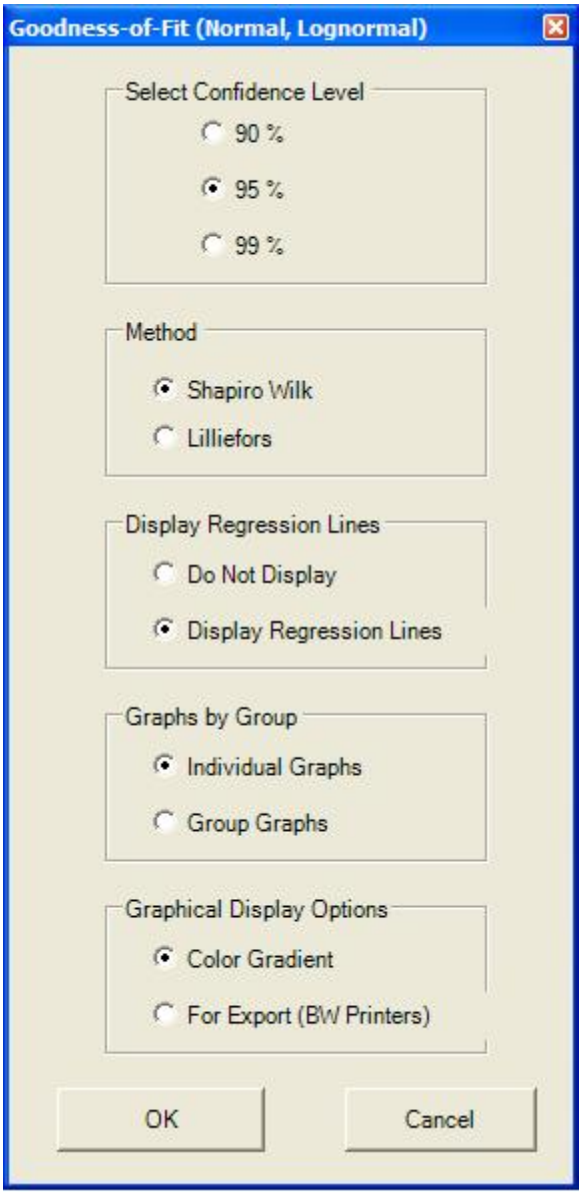


2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.

   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

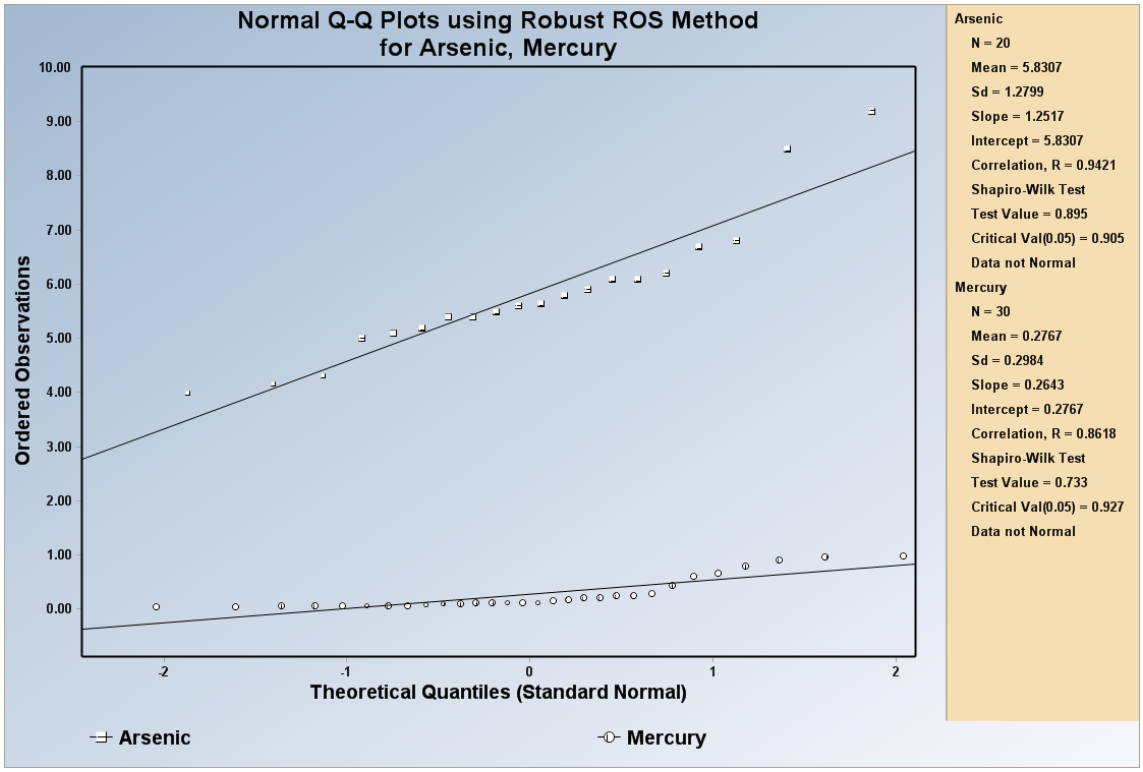
   - When the option button (Normal or Lognormal) is clicked, the following window will be shown.



- The default option for the **Confidence Level** is **95%**.
- The default GOF test **Method** is **Shapiro Wilk**. If the sample size is greater than 50, the program defaults to use the **Lilliefors** test.
- The default method for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the normal Q-Q plot, check the radio button next to **Display Regression Lines**.
- The default option for **Graphs by Group** is **Individual Graphs**. If you want to display all selected variables into a single graph, check the radio button next to **Group Graphs**.
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**.

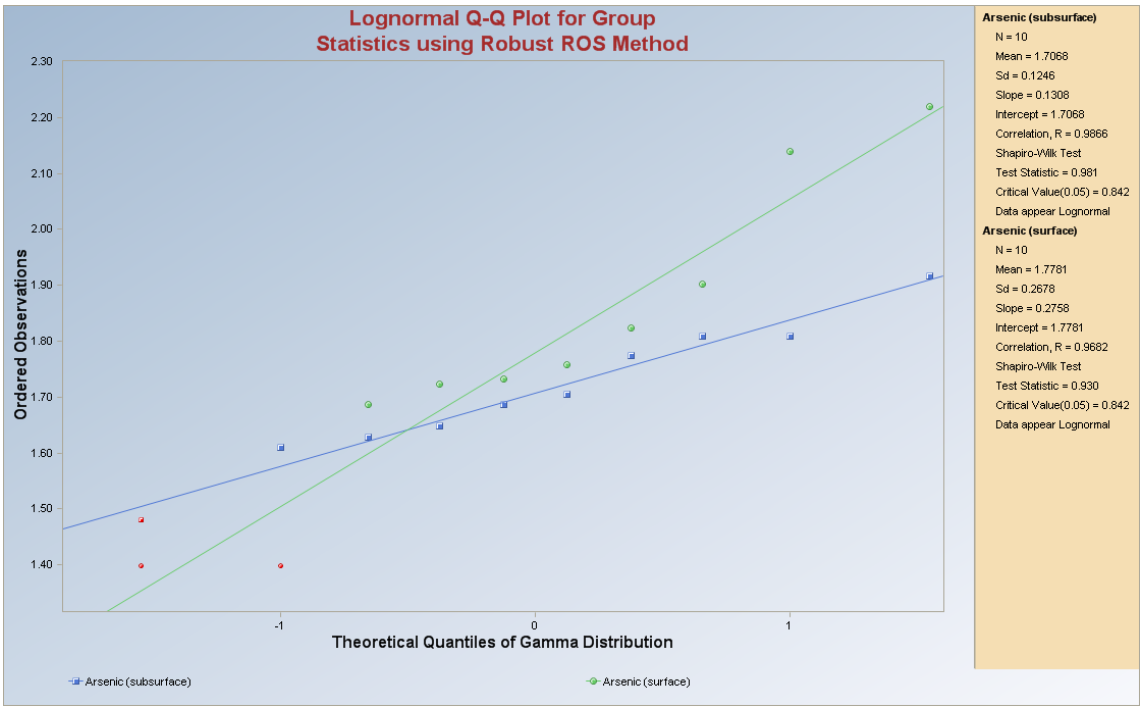
- Click the OK button to continue or the Cancel button to cancel the option.
- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

Output Screen for Normal Distribution (Log-ROS Estimates)
Selected options: Shapiro Wilk, Display Regression Lines, Group Graphs, and For Export (BW Printers).



Note: *The legend size of nondetect values is smaller than that of the detected values.*

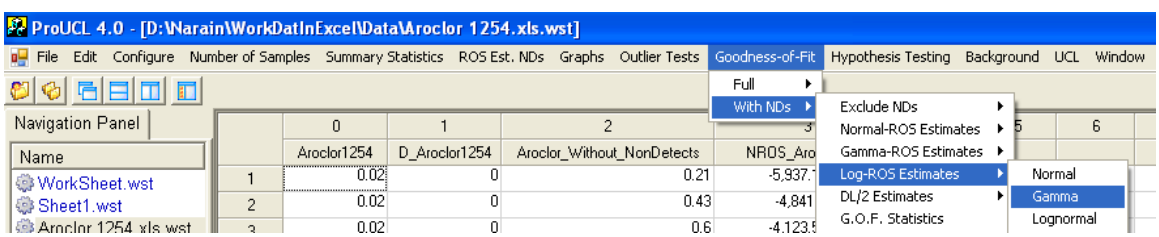
Output Screen for Lognormal Distribution (Log-ROS Estimates)
Selected options: Shapiro Wilk, Display Regression Lines, Group Graphs, and Color Gradient.



Note: *The legend size of nondetect values is smaller than that of the detected values and is shown in red.*

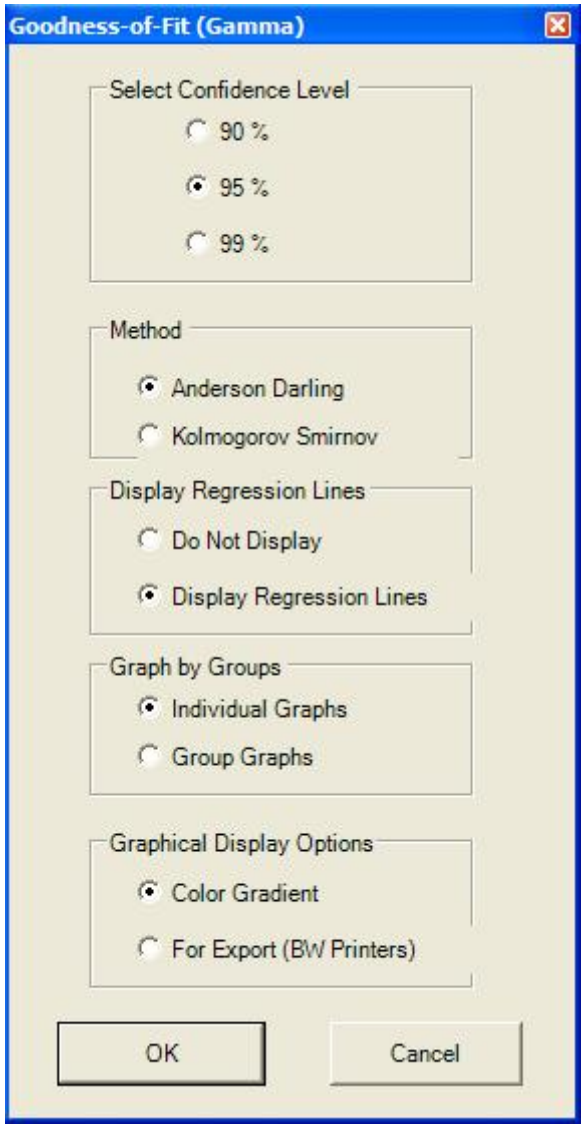
8.4.2 Gamma Distribution (Log-ROS Estimates)

1. Click **Goodness-of-Fit ► With NDs ► Log-ROS Estimates ► Gamma**



2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.
   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.
   - When the option button (Gamma) is clicked, the following window will be shown.

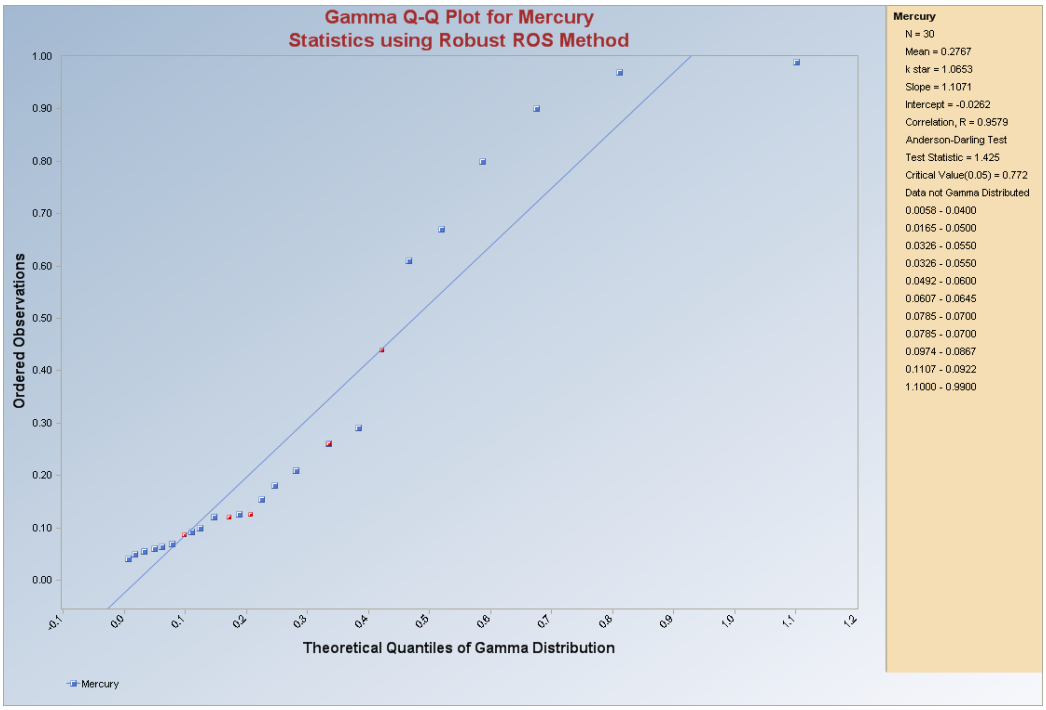


- The default option for the **Confidence Level** is **95%**.
- The default GOF test **Method** is **Anderson Darling**.
- The default method for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the normal Q-Q plot, check the radio button next to **Display Regression Lines**.
- The default option for **Graph by Groups** is **Individual Graphs**. If you want to put all of the selected variables into a single graph, check the radio button next to **Group Graphs**.
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**.

- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

Output Screen for Gamma Distribution (Log-ROS Estimates)

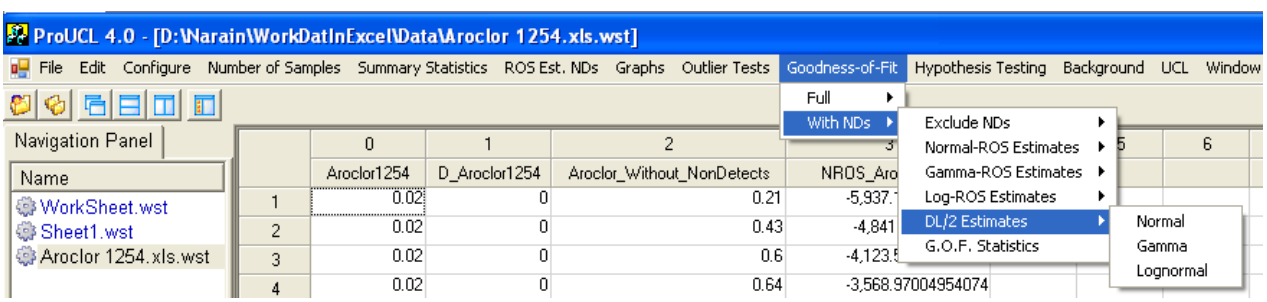
Selected options: Anderson Darling, Display Regression Lines, Individual Graphs, and Color Gradient.



Note: The *legend size of nondetect values is smaller than that of detected values and is shown in red.*

8.5 Goodness-of-Fit Tests with DL/2 Estimates

1. Click **Goodness-of-Fit ► With NDs ► DL/2 Estimates**

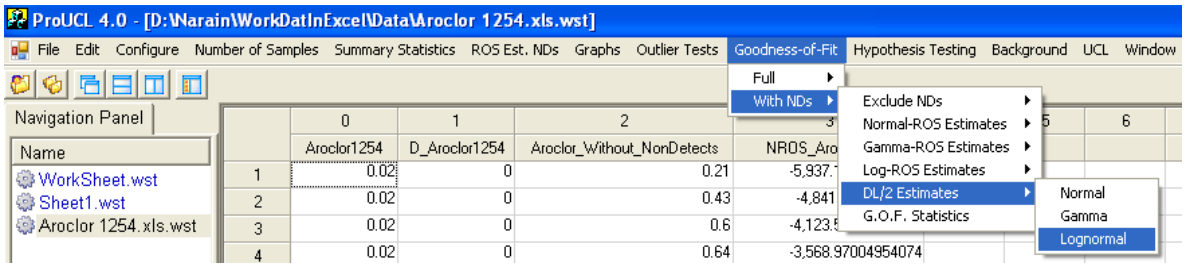


2. Select the distribution to be tested: Normal, Gamma, or Lognormal

- To test the variable for normality, click on Normal from the drop-down menu list.
- To test the variable for lognormality, click on Lognormal from the drop-down menu list.
- To test your variable for gamma distribution, click on Gamma from the drop-down menu list.

8.5.1 Normal or Lognormal Distribution (DL/2 Estimates)

1. Click **Goodness-of-Fit ► With NDs ► DL/2 Estimates ► Normal or Lognormal**

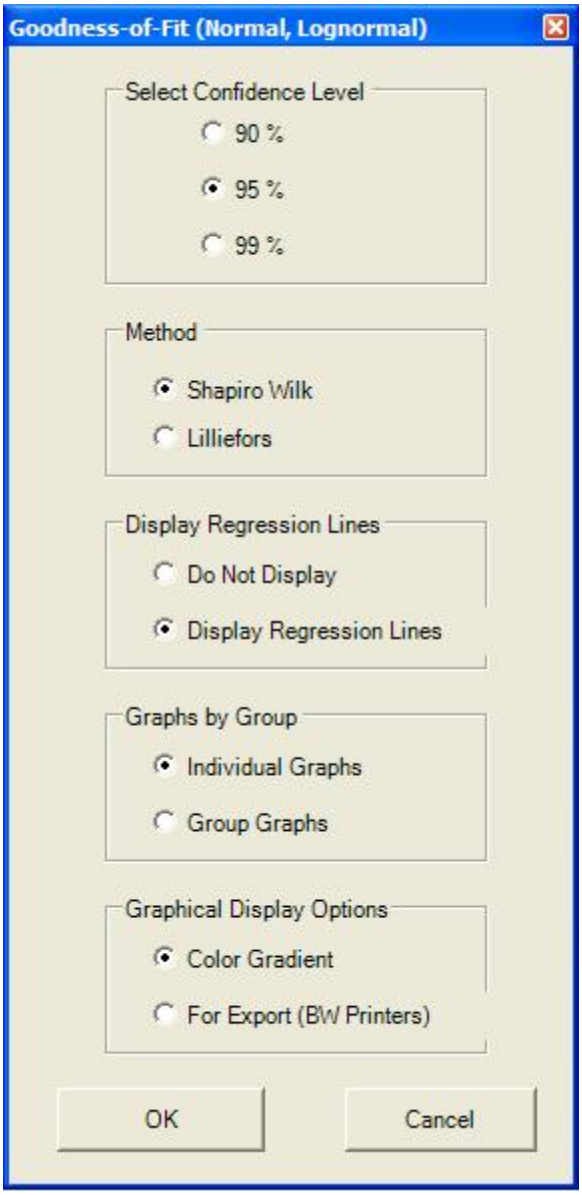


2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.

   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

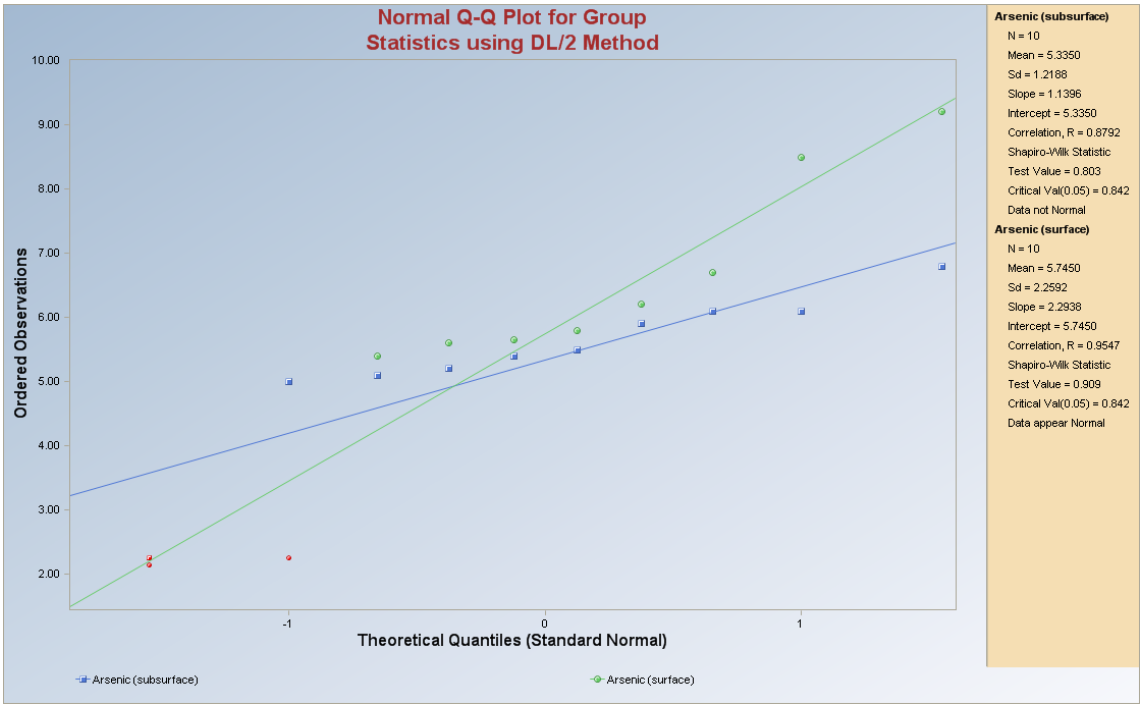
   - When Normal or Lognormal button is clicked, following window is displayed.



- The default option for the **Confidence Level** is **95%**.
- The default **Method** is **Shapiro Wilk**. If the sample size is greater than 50, the program defaults to the **Lilliefors** test.
- The default method for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the normal Q-Q plot, check the radio button next to **Display Regression Lines**.
- The default option for **Graphs by Group** is **Individual Graphs**. If you want to put all of the selected variables into a single graph, check the radio button next to **Group Graphs**.
- The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**

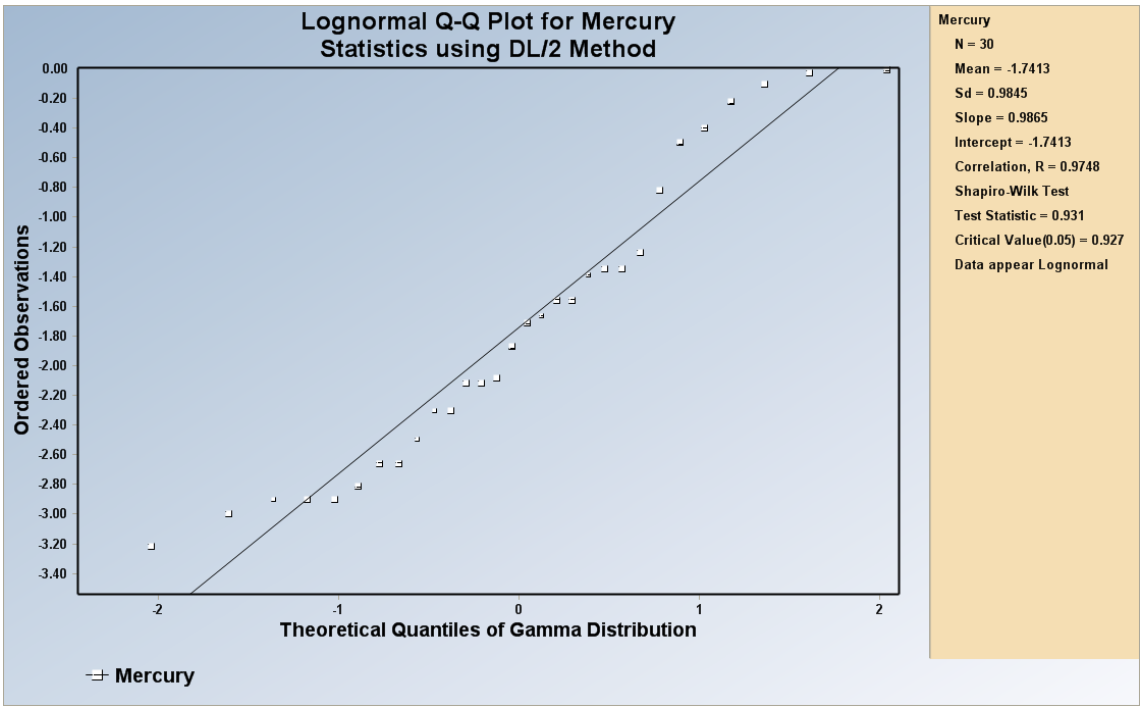
- Click the OK button to continue or the Cancel button to cancel the option.
- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

Output Screen for Normal Distribution (DL/2 Estimates)
Selected options: Shapiro Wilk, Display Regression Lines, Group Graphs, and Color Gradient.



Note: *The legend size of nondetect values is smaller than that of the detected values and is shown in red.*

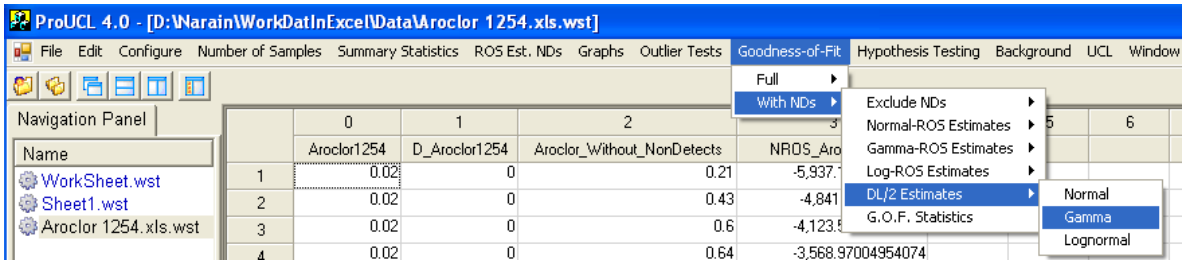
Output Screen for Lognormal Distribution (DL/2 Estimates)
Selected options: Shapiro Wilk, Display Regression Lines, Individual Graphs, and For Export (BW Printers).



Note: *The legend size of nondetect values is smaller than that of the detected values. The color is not shown on this graph as this graph is for BW printers.*

8.5.2 Gamma Distribution (DL/2 Estimates)

1. Click **Goodness-of-Fit ► With NDs ►DL/2 Estimates ► Gamma**

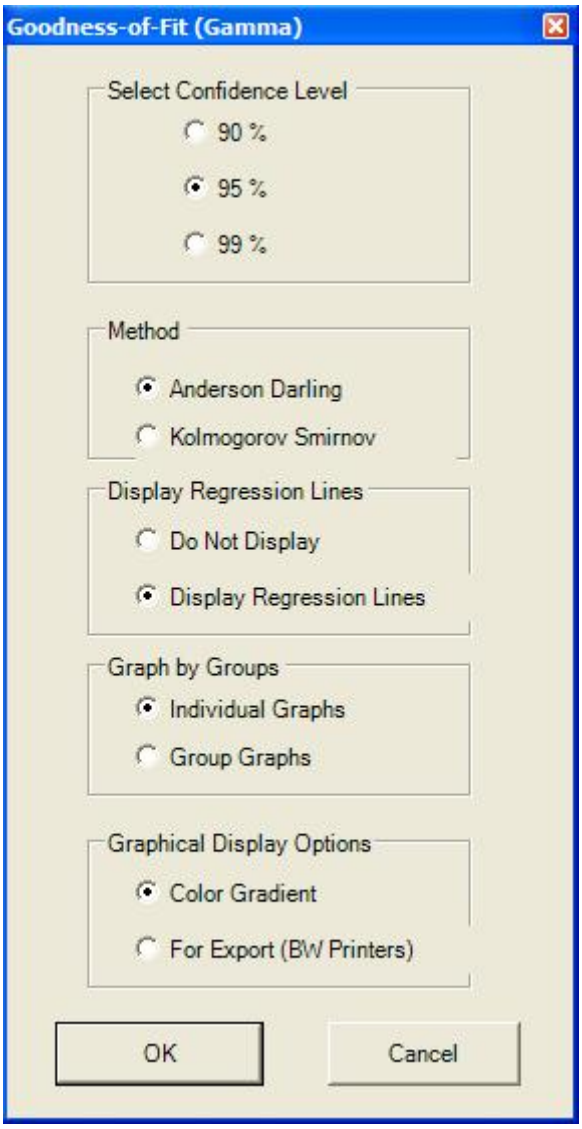


2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.

   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Group by variable** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group variable.

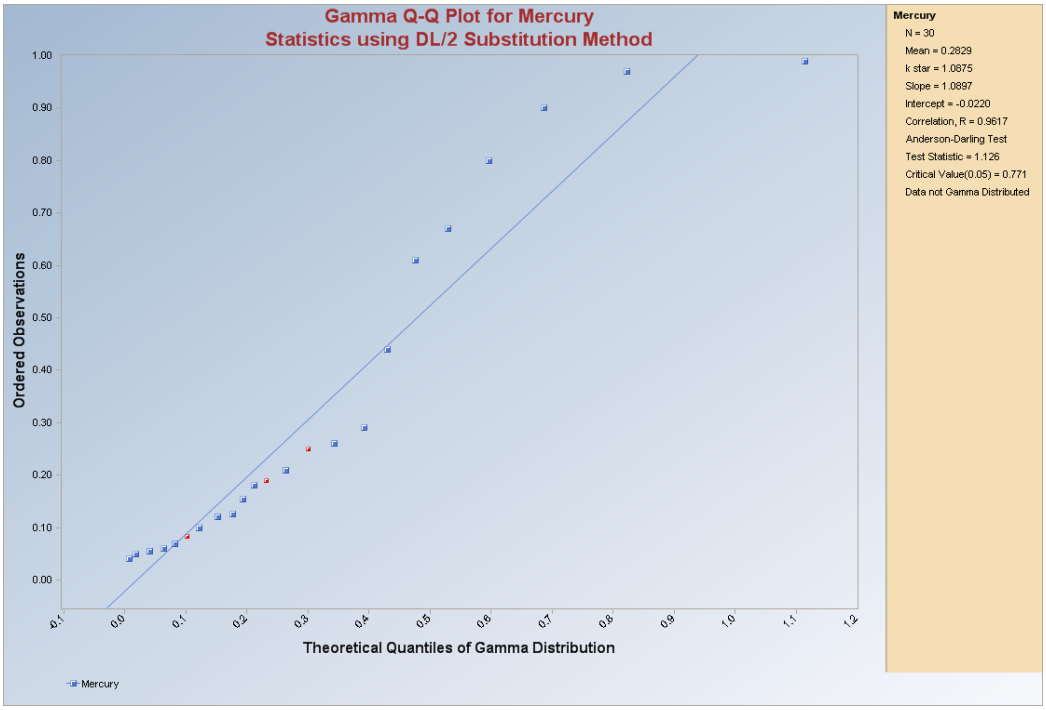
- When the Gamma option button is clicked, the following window will be shown.



  - The default option for the **Confidence Level** is **95%**.
  - The default **Method** is **Anderson Darling**.
  - The default method for **Display Regression Lines** is **Do Not Display**. If you want to see regression lines on the normal Q-Q plot, check the radio button next to **Display Regression Lines**.
  - The default option for **Graph by Groups** is **Individual Graphs**. If you want to put all of the selected variables into a single graph, check the radio button next to **Group Graphs**.
  - The default option for **Graphical Display Options** is **Color Gradient**. If you want to see the graphs in black and white, check the radio button next to **For Export (BW Printers)**.

- Click the OK button to continue or the Cancel button to cancel the Goodness-of-Fit tests.

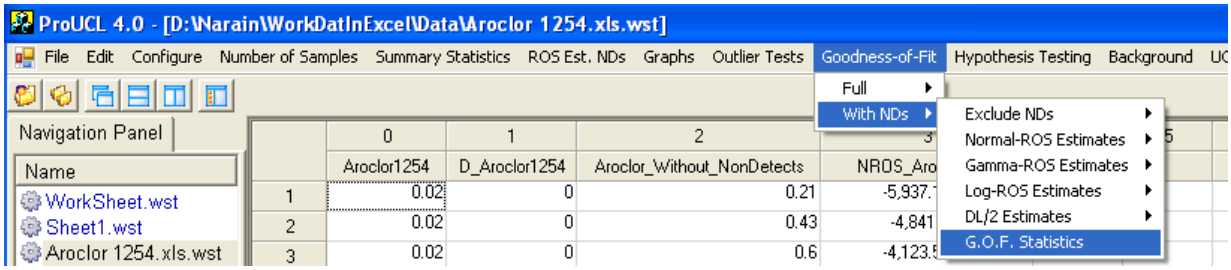
Output Screen for Gamma Distribution (DL/2 Estimates)
Selected options: Anderson Darling, Display Regression Lines, Individual Graphs, and Color Gradient.



Note: *The legend size of nondetect values is smaller than that of the detected values and is shown in red.*

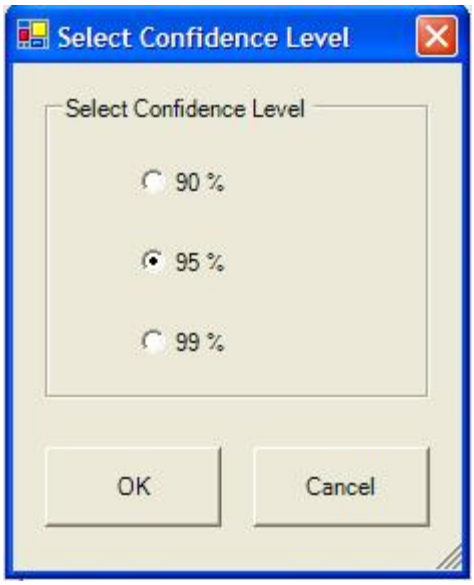
8.6 Goodness-of-Fit Tests Statistics

1. Click **Goodness-of-Fit ► With NDs ► G.O.F. Statistics**



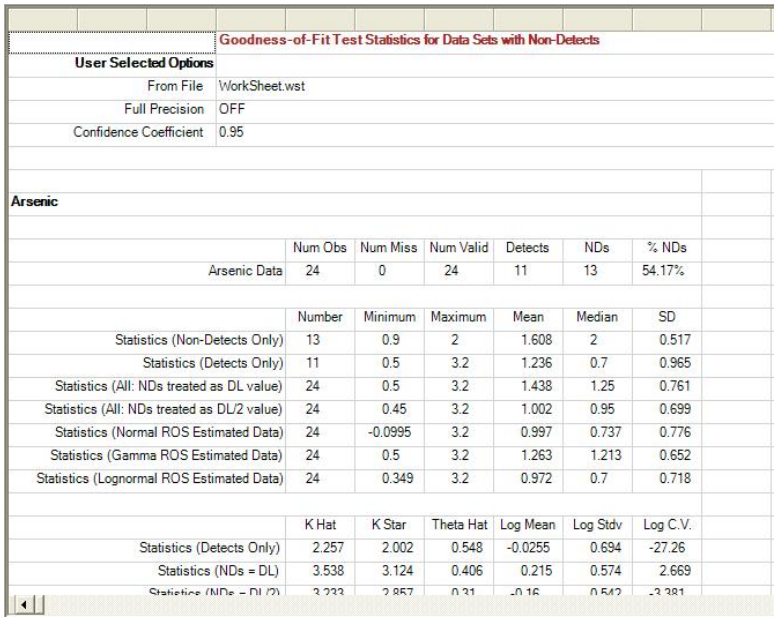
2. The **Select Variables Screen** (Chapter 3) will appear.

   - Select one or more variable(s) from the Select Variables screen.
   - When the option button is clicked, the following window will be shown.



- The default confidence level is 95%.
- Click the OK button to continue or the Cancel button to cancel the option.

Sample Output Screen for G.O.F. Test Statistics



Goodness-of-Fit Test Statistics for Data Sets with Non-Detects

User Selected Options	
From File	WorkSheet.wst
Full Precision	OFF
Confidence Coefficient	0.95

Arsenic

	Num Obs	Num Miss	Num Valid	Detects	NDs	% NDs
Arsenic Data	24	0	24	11	13	54.17%

	Number	Minimum	Maximum	Mean	Median	SD
Statistics (Non-Detects Only)	13	0.9	2	1.608	2	0.517
Statistics (Detects Only)	11	0.5	3.2	1.236	0.7	0.965
Statistics (All: NDs treated as DL value)	24	0.5	3.2	1.438	1.25	0.761
Statistics (All: NDs treated as DL/2 value)	24	0.45	3.2	1.002	0.95	0.699
Statistics (Normal ROS Estimated Data)	24	-0.0995	3.2	0.997	0.737	0.776
Statistics (Gamma ROS Estimated Data)	24	0.5	3.2	1.263	1.213	0.652
Statistics (Lognormal ROS Estimated Data)	24	0.349	3.2	0.972	0.7	0.718

	K Hat	K Star	Theta Hat	Log Mean	Log Stdv	Log C.V.
Statistics (Detects Only)	2.257	2.002	0.548	-0.0255	0.694	-27.26
Statistics (NDs = DL)	3.538	3.124	0.406	0.215	0.574	2.669
Statistics (NDs = DL/2)	3.233	2.857	0.31	-0.16	0.542	-3.381

Output Screen for the G.O.F. Test Statistics – (continued)

Normal Distribution Test Results

	Test value	Crit. (0.95)	Conclusion with Alpha(0.05)
Shapiro-Wilks (Detects Only)	0.777	0.85	Data Not Normal
Lilliefors (Detects Only)	0.273	0.267	Data Not Normal
Shapiro-Wilks (NDs = DL)	0.89	0.916	Data Not Normal
Lilliefors (NDs = DL)	0.217	0.181	Data Not Normal
Shapiro-Wilks (NDs = DL/2)	0.701	0.916	Data Not Normal
Lilliefors (NDs = DL/2)	0.335	0.181	Data Not Normal
Shapiro-Wilks (Normal ROS Estimates)	0.868	0.916	Data Not Normal
Lilliefors (Normal ROS Estimates)	0.17	0.181	Data Appear Normal

Gamma Distribution Test Results

	Test value	Crit. (0.95)	Conclusion with Alpha(0.05)
Anderson-Darling (Detects Only)	0.787	0.738	
Kolmogorov-Smirnov (Detects Only)	0.254	0.258	Data appear Approximate Gamma Distribution
Anderson-Darling (NDs = DL)	0.98	0.75	
Kolmogorov-Smirnov (NDs = DL)	0.214	0.179	Data Not Gamma Distributed
Anderson-Darling (NDs = DL/2)	1.492	0.751	
Kolmogorov-Smirnov (NDs = DL/2)	0.261	0.179	Data Not Gamma Distributed
Anderson-Darling (Gamma ROS Estimates)	0.524	0.747	
Kolmogorov-Smirnov (Gamma ROS Est.)	0.127	0.179	Data Appear Gamma Distributed

Chapter 9

Single Sample and Two-Sample Hypotheses Testing Approaches

This chapter illustrates single sample and two-sample parametric and nonparametric hypotheses testing approaches as incorporated in ProUCL 4.0. ProUCL 4.0 can perform these hypotheses tests on data sets with and without nondetect observations. It should be pointed out that, when one wants to use two-sample hypotheses tests on data sets with NDs, ProUCL 4.0 assumes that samples from both of the groups have nondetect observations. All this means that, a ND column (with 0 or 1 entries only) needs to be provided for the variable in each of the two groups. This has to be done even if one of the groups has all detected entries; in this case the associated ND column will have all entries equal to "1." This will allow the user to compare two groups (e.g., arsenic in background vs. site samples) with one of the groups having some NDs and the other group having all detected data.

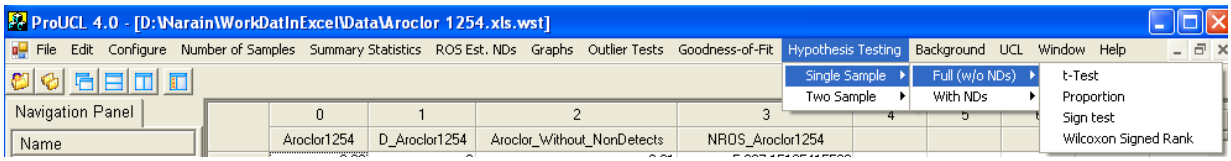
9.1 Single Sample Hypotheses Tests

In many environmental applications, single sample hypotheses tests are used to compare site data (provided enough site data are available) with pre-specified cleanup standards or compliance limits. ProUCL 4.0 contains single sample parametric and nonparametric tests including Student's t-test, sign test, Wilcoxon Signed Rank (WSR) test, and test for proportion. The single sample hypotheses tests are useful when the environmental parameters such as the clean standard, action level, or compliance limits (CLs) are known, and the objective is to compare site concentrations with those known threshold values. Specifically, a t-test (or a sign test) may be used to verify the attainment of cleanup levels at an AOC after remediation activity; and a test for proportion may be used to verify if the proportion of exceedances of an action level (or a compliance limit) by sample concentrations collected from an AOC (or a MW) exceeds a certain specified proportion (e.g., 1%, 5%, 10%).

ProUCL 4.0 can perform these hypotheses on data sets with and without nondetect observations. However, it should be noted that for single sample hypotheses tests (e.g., sign test, proportion test) used to compare site mean/median concentration level with a cleanup standard, $C_{s,}$ or a compliance limit (e.g., proportion test), all NDs (if any) should lie below the cleanup standard, $C_{s.}$ For proper use of these hypotheses testing approaches, the differences between these tests should be noted and understood. Specifically, a t-test or a WSR test are used to compare the measures of location and central tendencies (e.g., mean, median) of a site area (e.g., AOC) to a cleanup standard, C_s, or action level also representing a measure of central tendency (e.g., mean, median); whereas, a proportion test compares if the proportion of site observations from an AOC exceeding a compliance limit (*CL*) exceeds a specified proportion, P_0 (e.g., 5%, 10%). ProUCL 4.0 has useful graphical methods that may be used to visually compare the concentrations of a site area of concern (AOC) with an action level. This can be done using a box plot of site data with action level superimposed on that graph. The details of the various single sample hypotheses testing approaches can be found in EPA guidance documents (1989, 2006). A brief discussion of these methods is also given in the ProUCL 4.0 Technical Guide.

9.1.1 Single Sample Hypothesis Testing for Full Data without Nondetects

1. Click **Hypothesis Testing► Single Sample**

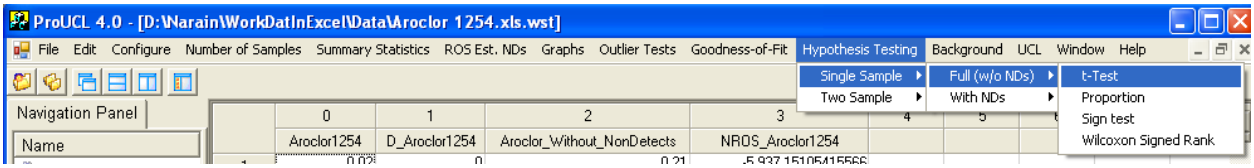


2. Select Full (w/o NDs) – This option is used for full data sets without nondetects.

- To perform a t-test, click on t-Test from the drop-down menu as shown above.
- To perform a proportion test, click on Proportion from the drop-down menu.
- To run a sign test, click on Sign test from the drop-down menu.
- To run a Wilcoxon Signed Rank test, click on Wilcoxon Signed Rank from the drop-down menu.

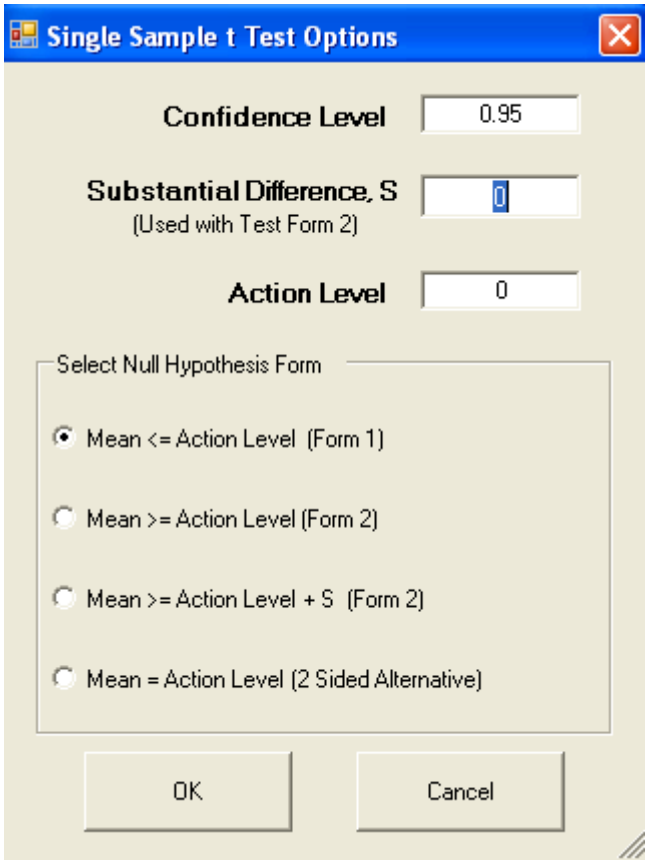
9.1.1.1 Single Sample t-Test

1. Click **Hypothesis Testing ► Single Sample ► Full (w/o NDs) ► t-Test**



2. The Select Variables Screen (see page 130) will appear.

- Select variable (variables) from the Select Variables screen.
- When the Options button is clicked, the following window will be shown.



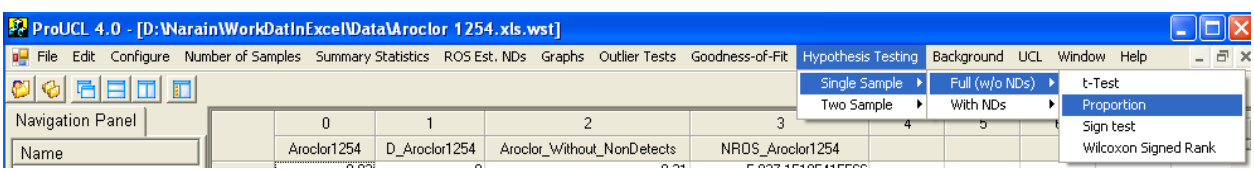
- Specify the **Confidence Level**; default is 0.95.
- Specify meaningful values for **Substantial Difference, S** and the **Action Level**. The default choice for S is "0."
- Select the form of Null Hypothesis; default is Mean <= Action Level (Form 1).
- Click on OK button to continue or on Cancel button to cancel the test.

Output for Single Sample t-Test (Full Data without NDs)

1 Sample- t	
Single Sample t-Test	
Raw Statistics	
Number of Valid Samples	9
Number of Distinct Samples	9
Minimum	82.39
Maximum	113.2
Mean	99.38
Median	103.5
SD	10.41
SE of Mean	3.468
H0: Site Mean = 100	
Test Value	-0.178
Two Sided Critical Value (0.05)	2.306
P-Value	0.863
Conclusion with Alpha = 0.05	
Do Not Reject H0, Conclude Mean = 100	
P-Value > Alpha (0.05)	

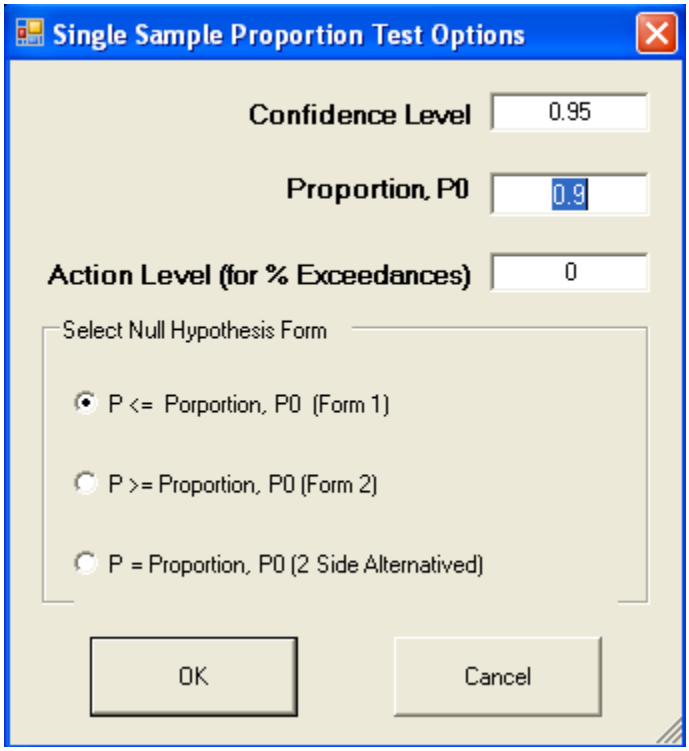
9.1.1.2 Single Sample Proportion Test

1. Click **Hypothesis Testing ► Single Sample ► Full (w/o NDs) ► Proportion**



2. The Select Variables Screen will appear.

- Select variable (variables) from the Select Variables screen.
- When the Options button is clicked, the following window will be shown.



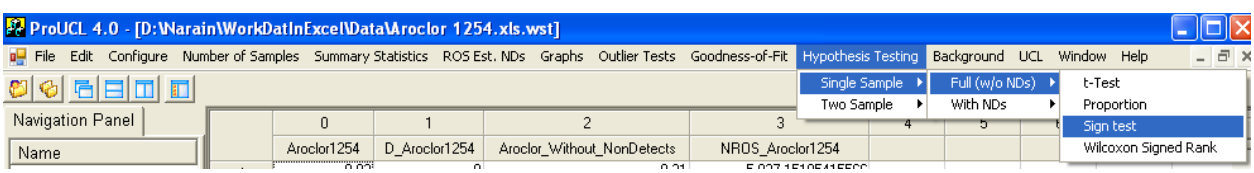
- Specify the **Confidence level**; default is 0.95.
- Specify the **Proportion** level and a meaningful **Action Level**.
- Select the form of Null Hypothesis; default is P <= Proportion (Form 1).
- Click on OK button to continue or on Cancel button to cancel the test.

Output for Single Sample Proportion Test (Full Data without NDs)

One-Sample Proportion Test	
Raw Statistics	
Number of Valid Samples	85
Number of Distinct Samples	83
Minimum	0.598
Maximum	7.676
Mean	5.183
Median	5.564
SD	1.588
SE of Mean	0.172
Number of Exceedances	27
Sample Proportion of Exceedances	0.318
H0: Site Proportion <= 0.3 (Form 1)	
Large Sample z-Test Value	0.237
Critical Value (0.05)	1.645
P-Value	0.406
Conclusion with Alpha = 0.05	
Do Not Reject H0, Conclude Site Proportion <= 0.3	
P-Value > Alpha (0.05)	

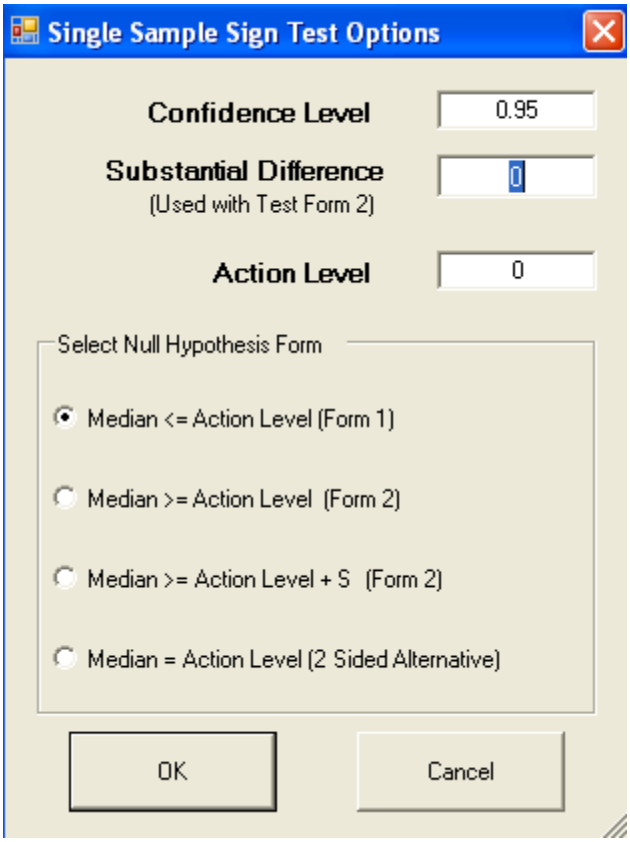
9.1.1.3 Single Sample Sign Test

1. Click **Hypothesis Testing ► Single Sample ► Full (w/o NDs) ► Sign test**



2. The Select Variables Screen will appear.

   - Select variable (variables) from the Select Variables screen.
   - When the Options button is clicked, the following window will be shown.



- Specify the **Confidence Level**; default choice is 0.95.
- Specify meaningful values for **Substantial Difference, S** and **Action Level**.
- Select the form of Null Hypothesis; default is Median <= Action Level (Form 1).
- Click on OK button to continue or on Cancel button to cancel the test.

Output for Single Sample Proportion Test (Full Data without NDs)

Single Sample Sign Test

Raw Statistics

Number of Valid Samples	10
Number of Distinct Samples	10
Minimum	750
Maximum	1161
Mean	925.7
Median	888
SD	136.7
SE of Mean	43.24
Number Above Limit	3
Number Equal Limit	0
Number Below Limit	7

H0: Site Median >= 1000 (Form 2)

Test Value	3
Lower Critical Value (0.05)	1
P-Value	0.172

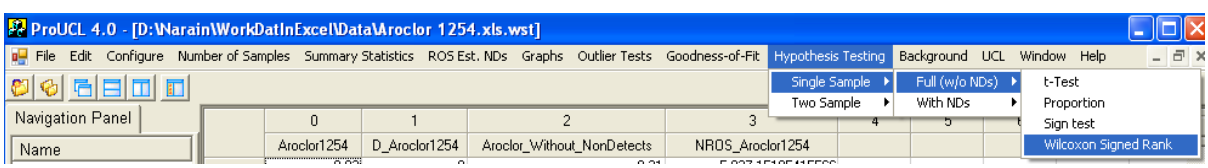
Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Median >= 1000

P-Value > Alpha (0.05)

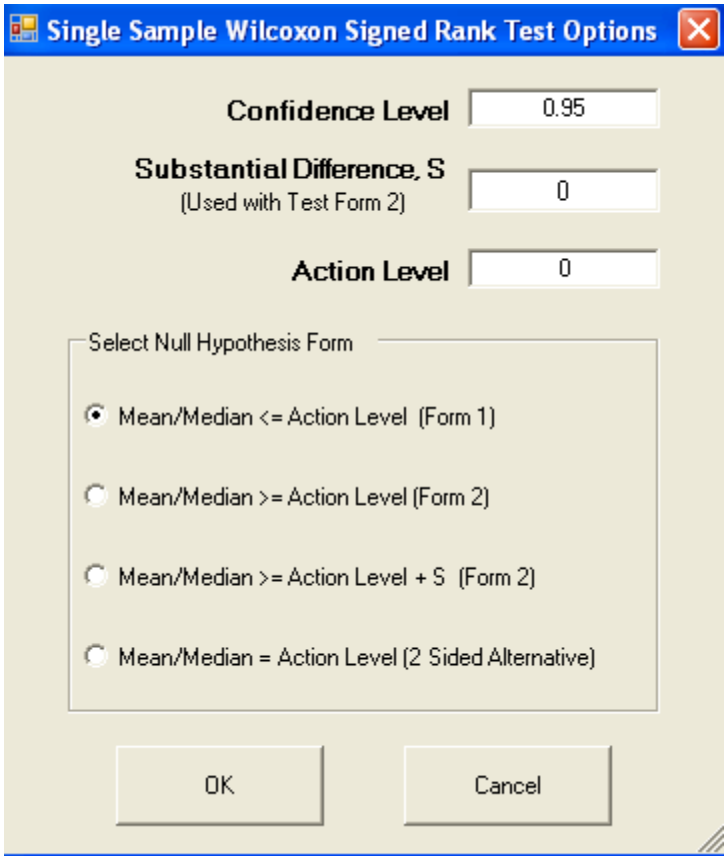
9.1.1.4 Single Sample Wilcoxon Signed Rank (WSR) Test

1. Click **Hypothesis Testing ► Single Sample ► Full (w/o NDs) ► Wilcoxon Signed Rank**



2. The Select Variables Screen will appear.

- Select variable (variables) from the Select Variables screen.
- When the Options button is clicked, the following window will be shown.



- Specify the **Confidence Level**; default is 0.95.
- Specify meaningful values for **Substantial Difference, S,** and **Action Level**.
- Select the form of Null Hypothesis; default is Mean/Median <= Action Level (Form 1).
- Click on OK button to continue or on Cancel button to cancel the test.

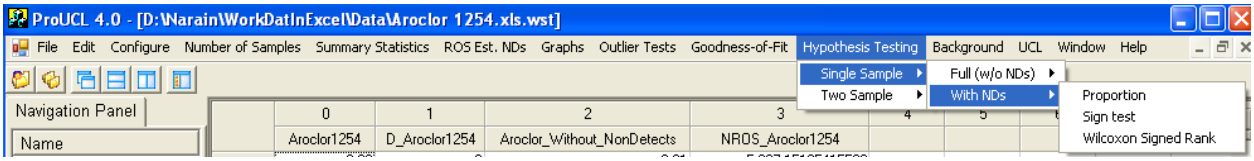
Output for Single Sample Proportion Test (Full Data without NDs)

Single Sample Wilcoxon Signed Rank Test		
Raw Statistics		
Number of Valid Samples	10	
Number of Distinct Samples	10	
Minimum	750	
Maximum	1161	
Mean	925.7	
Median	888	
SD	136.7	
SE of Mean	43.24	
Number Above Limit	3	
Number Equal Limit	0	
Number Below Limit	7	
T-plus	11.5	
T-minus	43.5	
H0: Site Median <= 1000 (Form 1)		
Test Value	11.5	
Critical Value (0.05)	45	
P-Value	0.947	
Conclusion with Alpha = 0.05		
Do Not Reject H0, Conclude Mean/Median <= 1000		
P-Value > Alpha (0.05)		

9.1.2 Single Sample Hypothesis Testing for Data Sets with Nondetects

Most of the one-sample tests such as the Proportion test and the Sign test on data sets with nondetect values assume that all nondetect observations lie below the compliance limit (CL) or an action level, A_0. The single sample tests cannot be performed if ND observations exceed the CL or action levels.

1. Click on **Hypothesis Testing► Single Sample**

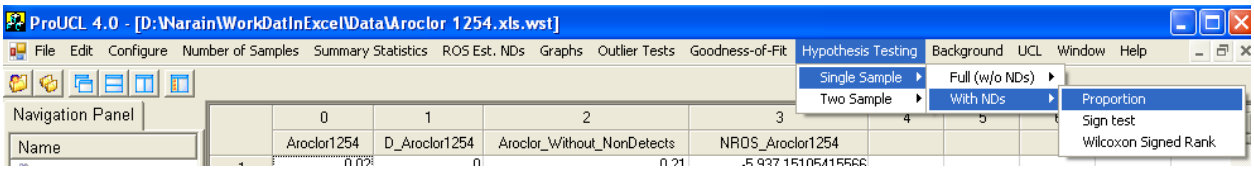


2. Select the With NDs option

- To perform a proportion test, click on Proportion from the drop-down menu.
- To perform a sign test, click on Sign test from the drop-down menu.
- To perform a Wilcoxon Signed Rank test, click on Wilcoxon Signed Rank from the drop-down menu list.

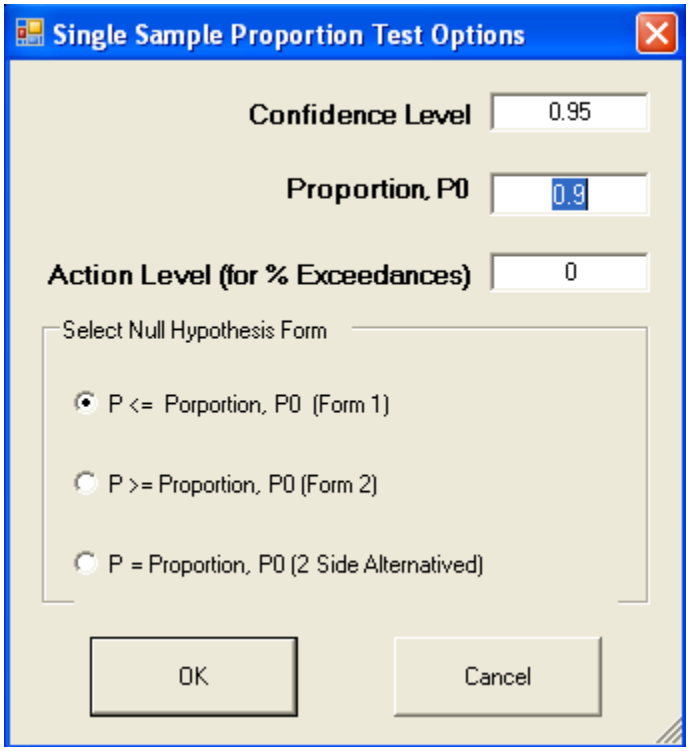
9.1.2.1 Single Proportion Test on Data Sets with NDs

1. Click **Hypothesis Testing ► Single Sample ► With NDs ► Proportion**



2. The Select Variables screen will appear.

- Select variable (variables) from the Select Variables screen.
- When the Options button is clicked, the following window will be shown.



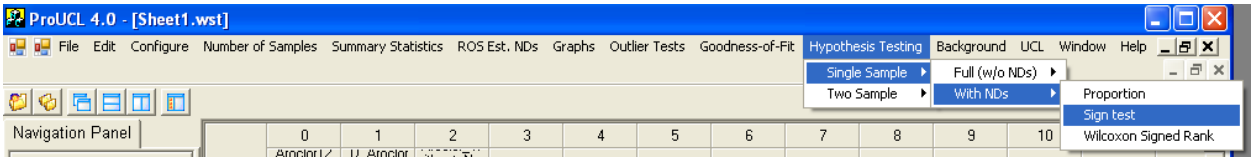
- Specify the **Confidence Level**; default is 0.95.
- Specify meaningful values for **Proportion** and the **Action Level**.
- Select the form of Null Hypothesis; default is P <= Proportion (Form 1).
- Click on OK button to continue or on Cancel button to cancel the test.

Output for Single Sample Proportion Test (with NDs)

Arsenic	
Single Sample Proportion Test	
Raw Statistics	
Number of Valid Samples	24
Number of Distinct Samples	10
Number of Non-Detect Data	13
Number of Detected Data	11
Percent Non-Detects	54.17%
Minimum Non-detect	0.9
Maximum Non-detect	2
Minimum Detected	0.5
Maximum Detected	3.2
Mean of Detected Data	1.236
Median of Detected Data	0.7
SD of Detected Data	0.965
Number of Exceedances	2
Sample Proportion of Exceedances	0.0833
Some Non-Detect Values Exceed	
The User Selected Action/Compliance Limit	
Unable to do Proportion Test with such parameters	

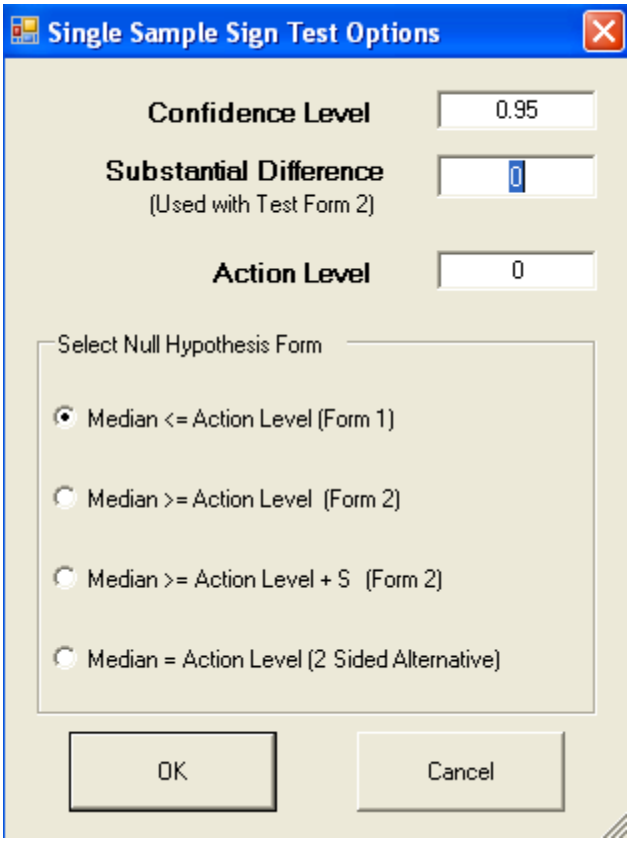
9.1.2.2 Single Sample Sign Test with NDs

1. Click **Hypothesis Testing ► Single Sample ► With NDs ► Sign test**



2. The Select Variables screen will appear.

- Select variable (variables) from the Select Variables screen.
- When the Options button is clicked, the following window will be shown.



- Specify the **Confidence Level**; default is 0.95.
- Specify meaningful values for **Substantial Difference, S** and **Action Level**.
- Select the form of Null Hypothesis; default is Median <= Action Level (Form 1).
- Click on OK button to continue or on Cancel button to cancel the test.

Output for Single Sample Sign Test (Data with Nondetects)

Arsenic

Single Sample Sign Test

Raw Statistics

Number of Valid Samples	24
Number of Distinct Samples	10
Number of Non-Detect Data	13
Number of Detected Data	11
Percent Non-Detects	54.17%
Minimum Non-detect	0.9
Maximum Non-detect	2
Minimum Detected	0.5
Maximum Detected	3.2
Mean of Detected Data	1.236
Median of Detected Data	0.7
SD of Detected Data	0.965
Number Above Limit	0
Number Equal Limit	0
Number Below Limit	24

H0: Site Median <= 5 (Form 1)

Test Value	0
Upper Critical Value (0.05)	17
P-Value	1

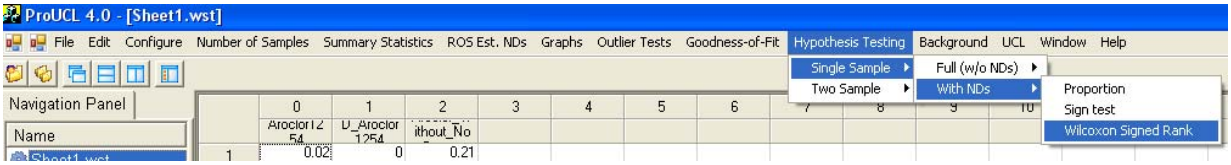
Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Median <= 5

P-Value > Alpha (0.05)

9.1.2.3 Single Sample Wilcoxon Signed Rank Test with NDs

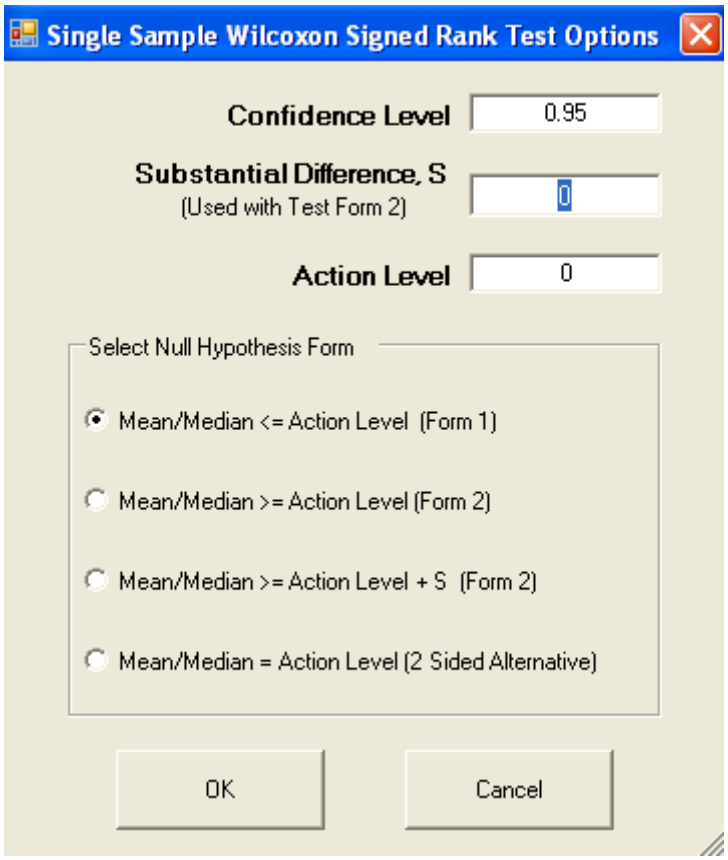
1. Click **Hypothesis Testing ► Single Sample ► With NDs ► Wilcoxon Signed Rank**



2. The Select Variables Screen will appear.

- Select variable (variables) from the Select Variables screen.

- When the Options button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; default is 0.95.
  - Specify meaningful values for **Substantial Difference, S** and **Action Level**.
  - Select the form of Null Hypothesis; default is Mean/Median <= Action Level (Form 1).
  - Click on OK button to continue or on Cancel button to cancel the test.

Output for Single Sample Wilcoxon Signed Rank Test (Data with Nondetects)

Arsenic

Single Sample Wilcoxon Signed Rank Test

Raw Statistics

Number of Valid Samples	24
Number of Distinct Samples	10
Number of Non-Detect Data	13
Number of Detected Data	11
Percent Non-Detects	54.17%
Minimum Non-detect	0.9
Maximum Non-detect	2
Minimum Detected	0.5
Maximum Detected	3.2
Mean of Detected Data	1.236
Median of Detected Data	0.7
SD of Detected Data	0.965
Number Above Limit	0
Number Equal Limit	0
Number Below Limit	24
T-plus	0
T-minus	300

H0: Site Median <= 6 (Form 1)

Large Sample z-Test Value	-4.293
Critical Value (0.05)	1.645
P-Value	1

Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Mean/Median <= 6

P-Value > Alpha (0.05)

Dataset contains multiple Non-Detect values!

All Observations < 2 are treated as Non-Detects

9.2 Two-Sample Hypotheses Testing Approaches

In this section, the two-sample hypotheses testing approaches as incorporated in ProUCL 4.0 have been illustrated. These approaches are used to compare the parameters and distributions of the two populations (e.g., Background vs. AOC) based upon data sets collected from those populations. Both forms (Form 1 and Form 2, Form 2 with Substantial Difference, S) of two-sample hypothesis testing approaches have been included in ProUCL 4.0. The methods are available for full data sets as well as for data sets with below detection limit (BDL) values.

- Full – analyzes data sets consisting of all detected values. The following parametric and nonparametric tests are available:
    - Student's t and Satterthwaite tests to compare the means of two populations (e.g. Background versus AOC).
    - F-test to the check the equality of dispersions of two populations.
    - Two-sample nonparametric Wilcoxon-Mann-Whitney (WMW) test. This test is equivalent to Wilcoxon Rank Sum (WRS) test.
    - Quantile test is often used to compare upper tails of two data distributions. This test is normally performed in parallel with WMW test.

- With NDs – analyzes data sets consisting of both nondetected and detected values. The following tests are available:
    - Wilcoxon-Mann-Whitney test. All observations (including detected values) below the highest detection limit are treated as ND (less than the highest DL) values.
    - Quantile test is used to compare upper tails of two data distributions. This test is performed in parallel with WMW test.
    - Gehan's test, useful when multiple detection limits may be present.

The details of these methods can be found in the ProUCL 4.0 Technical Guide and are also available in EPA (1997, 2006). It is re-stated that the use of informal graphical displays (e.g., side-by-side box plots, multiple Q-Q plots) should always accompany the formal hypothesis testing approaches listed above. This is especially warranted when the data sets may consist of observations from multiple populations (e.g., mixture samples collected from various onsite locations) and outliers.

Note: *As mentioned before, it is pointed out that, when one wants to use two-sample hypotheses tests on data sets with NDs, ProUCL 4.0 assumes that samples from both of the groups have nondetect observations. This may not be the case, as data from a polluted site may not have any ND observations. ProUCL can handle such data sets. However, the user will have to provide a ND column (with 0 or 1 entries only) for the selected variable of each of the two groups. Thus when one of the groups (e.g., site arsenic) has no ND value, the user supplies an associated ND column with all entries equal to "1." This will allow the user to compare two groups (e.g., arsenic in background vs. site samples) with one of the group having some NDs and the other group having all detected data*.

9.2.1 Two-Sample Hypothesis Tests for Full Data

Full – This option is used to analyze data sets consisting of all detected values. The following two-sample tests are available in ProUCL 4.0.

- Student's t and Satterthwaite tests to compare the means of two populations (e.g., Background versus AOC).

- F-test is also available to test the equality of dispersions of two populations.

- Two-sample nonparametric Wilcoxon-Mann-Whitney (WMW) test.

- Two-sample quantile test.

- Student's t-Test

  - This test can be used to compare the site mean concentration of a COPC with that of the background mean concentration provided the populations are normally distributed. The data sets are given by independent random observations, $X_1, X_2, \ldots, X_n$ collected from a site, and independent random observations, $Y_1, Y_2, \ldots, Y_m$ collected from a background population. The same terminology is used for all other two-sample tests in ProUCL 4.0.
  - Student's t-test also assumes that the spread (variance) of the two populations are approximately equal.
  - The F-test can be used to the check the equality of dispersions of two populations.

- Satterthwaite t-Test

  - This test is used to compare the population means of two populations when the variances or Spreads of those populations may not be equal. As mentioned before, the F-distribution based test can be used to verify the equality of dispersions of two populations.

- Test for Equality of two Dispersions (F-test)

  - This test is used to determine whether the true underlying variances of two populations are equal. Usually the F-test is employed as a preliminary test, before conducting the two-sample t-test for testing the equality of means of two populations.
  - The assumptions underlying the F-test are that the two-samples represent independent random samples from two normal populations. The F-test for equality of variances is highly sensitive to departures from normality.

- Two-Sample Nonparametric WMW Test

  - This test is used to determine the approximate equality of the two continuous data distributions. This test also assumes that the shapes (e.g., as determined by spread, skewness, and graphical displays) of the two populations are roughly equal. The test is often used to determine if the measures of central locations of the two populations

are significantly different. Specifically, the test can be used to determine if the site concentrations exceed the background concentrations.

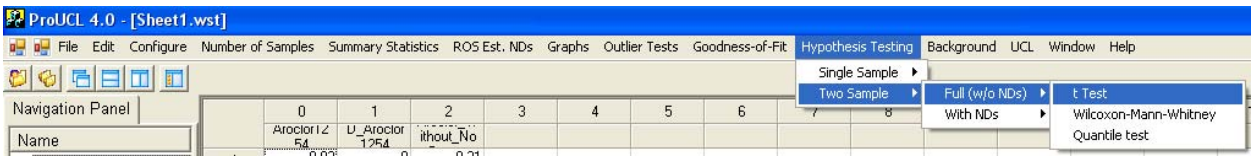
- The Wilcoxon-Mann-Whitney test does not assume that the data are normally or log-normally distributed. For large samples (e.g., ≥ 20), the distribution of the WMW test statistic can be approximated by a normal distribution.
- This test is used to determine if measurements from one population consistently tend to be larger (or smaller) than those from the other population.

- Two-Sample Quantile Test

  - The nonparametric quantile test does not assume that the data are normally or log-normally distributed. For large samples (e.g., ≥ 20), the distribution of the quantile test statistic can be approximated by a normal distribution.
  - This test is used in parallel with the WMW test. This test is often used in Background Test Form 1 to determine if the concentrations from the upper tail of site data distribution are comparable to (lower than or equal to) that of the background data distribution. The critical values for this Form 1 test are available in EPA, 1994. The details of the test are given in EPA (1994, 2006).

Note: *The use of the tests listed above is not recommended on log-transformed data sets, especially when the parameters of interests are the population means. In practice, the cleanup and remediation decisions have to be made in original scale based upon statistics and estimates computed in the original scale. The equality of means in log-scale does not necessarily imply the equality of means in the original scale. This topic is discussed in detail in Chapter 3 of the revised background document (EPA, 2002) for CERCLA sites (currently under revision).*

1. Click on **Hypothesis Testing ► Two Sample**



2. Select the Full (w/o NDs) option

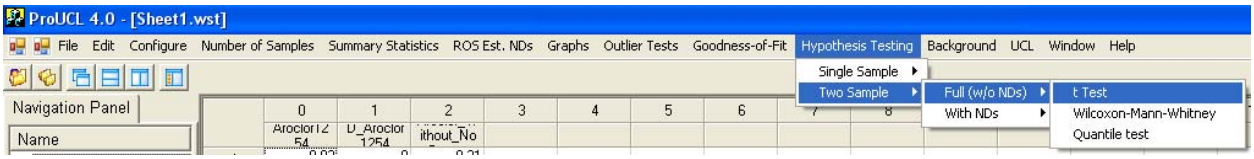
- To perform a t-test, click on t Test from the drop-down menu.

- To perform a Wilcoxon-Mann-Whitney, click on Wilcoxon-Mann-Whitney from the drop-down menu list.

- To perform a quantile test, click on Quantile test from the drop-down menu.

9.2.1.1 Two-Sample t-Test without NDs

1. Click **Hypothesis Testing ► Two Sample ► Full (w/o NDs) ► t Test**



2. The Select Variables screen will appear.

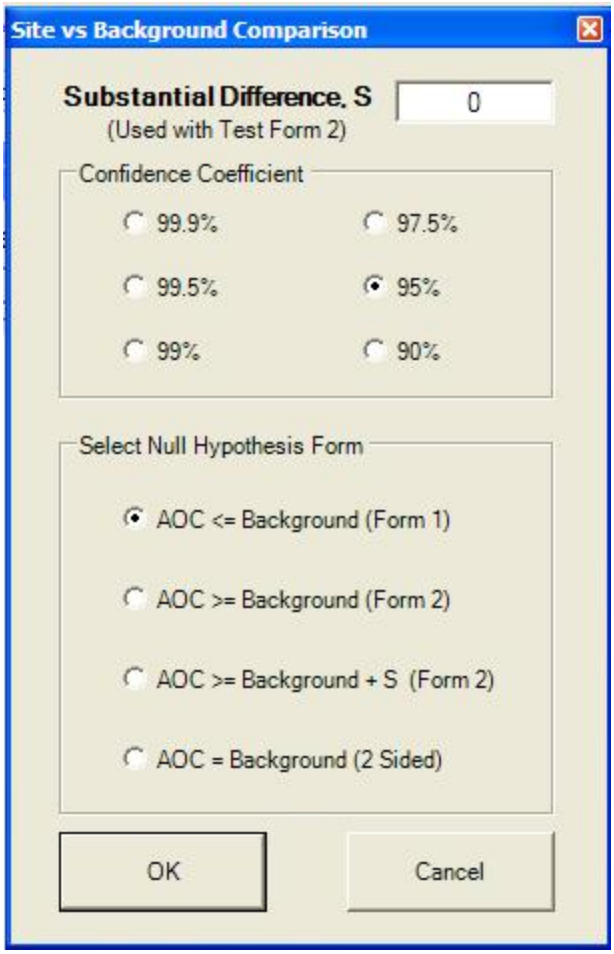
- Select variable (variables) from the Select Variables screen.

- **Without Group Variable**: This option is used when the data values of the variable (COPC) for the site and the background are given in separate columns.

- **With Group Variable**: This option is used when data values of the variable (COPC) for the site and the background are given in the same column. The values are separated into different populations (groups) by the values of an associated Group Variable. The group variable may represent several populations (e.g., several AOCs, MWs). The user can compare two groups at a time by using this option.

- When using this option, the user should select a group variable by clicking the arrow next to the **Group Var** option for a drop-down list of available variable. The user selects an appropriate (meaningful) variable representing groups such as Background and AOC. The user is allowed to use letters, numbers, or alphanumeric labels for the group names.

    - When the options button is clicked, the following window will be shown.



- Specify a useful **Substantial Difference, S** value. The default choice is **0**.
- Choose the **Confidence** level. The default choice is **95%**.
- Select the form of Null Hypothesis. The default is AOC <= Background (Form 1).
- Click on OK button to continue or on Cancel button to cancel the option.

- Click on the OK button to continue or on the Cancel button to cancel the Site versus Background Comparison.

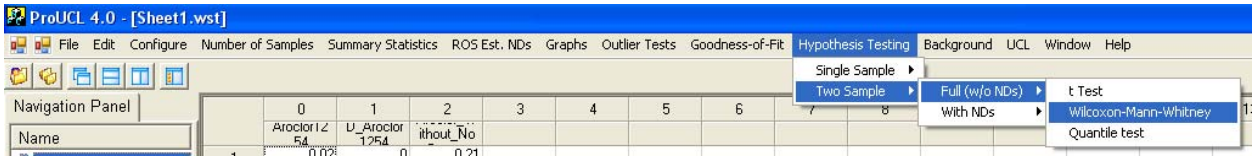
Output for Two-Sample t-Test (Full Data without NDs)

Raw Statistics			
	Site	Background	
Number of Valid Samples	77	47	
Number of Distinct Samples	57	40	
Minimum	0.09	0.22	
Maximum	168.6	1.33	
Mean	3.915	0.599	
Median	0.43	0.54	
SD	20.02	0.284	
SE of Mean	2.281	0.0414	

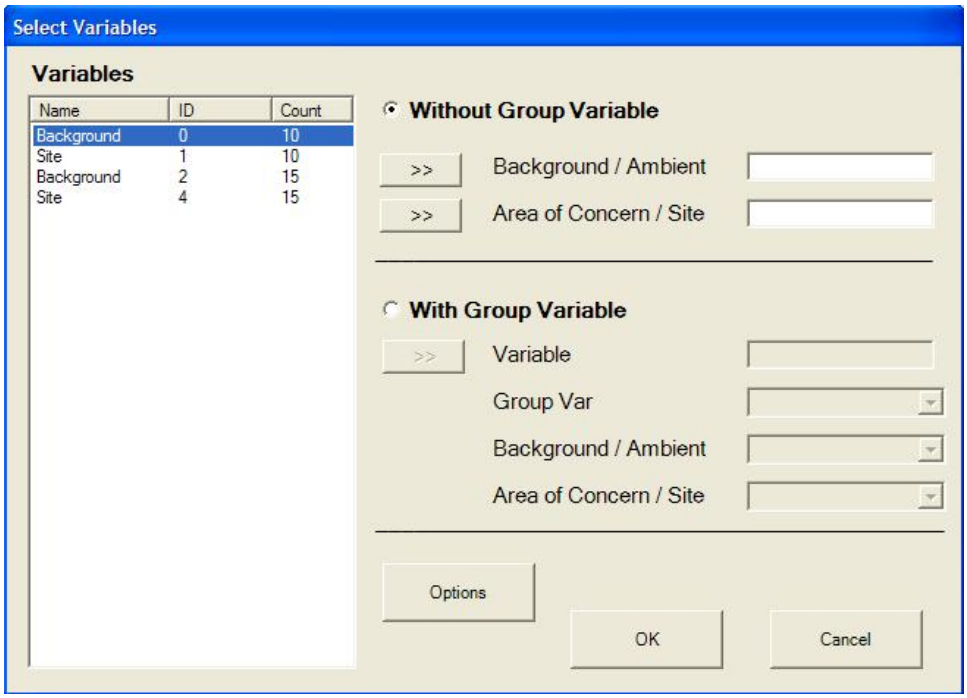
Site vs Background Two-Sample t-Test				
H0: Mu of Site - Mu of Background <= 0				
		t-Test	Critical	
Method	DF	Value	t (0.050)	P-Value
Pooled (Equal Variance)	122	1.134	1.657	0.129
Satterthwaite (Unequal Variance)	76.1	1.454	1.665	0.075
Pooled SD 15.799				
Conclusion with Alpha = 0.050				
* Student t (Pooled) Test: Do Not Reject H0, Conclude Site <= Background				
* Satterthwaite Test: Do Not Reject H0, Conclude Site <= Background				

9.2.1.2 Two-Sample Wilcoxon-Mann-Whitney (WMW) Test without NDs

1. Click **Hypothesis Testing ► Two Sample ► Full (w/o NDs) ► Wilcoxon-Mann-Whitney Test**



2. The Select Variables Screen will appear.



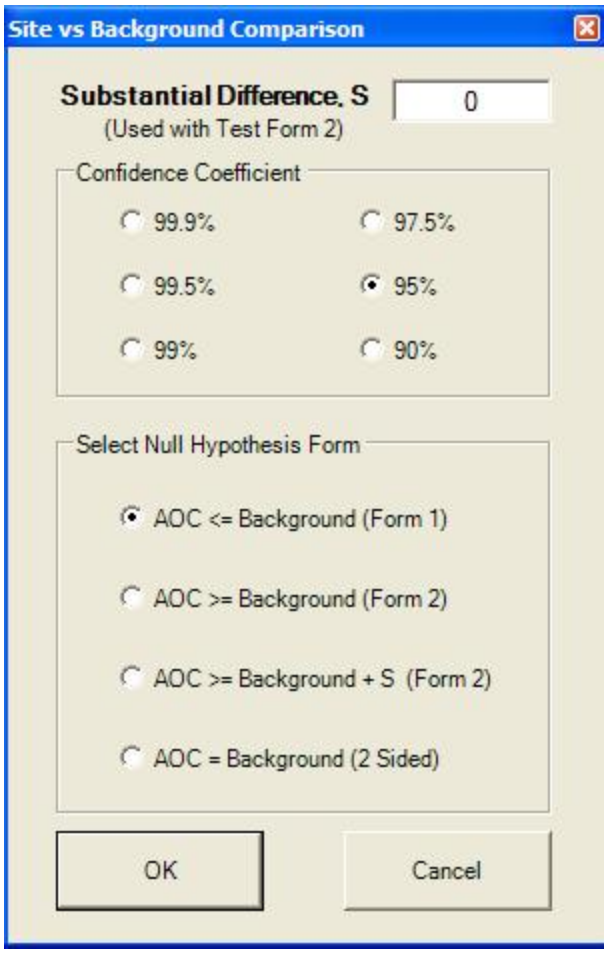
- Select variable (variables) from the Select Variables screen.

- **Without Group Variable**: This option is used when the data values of the variable (COPC) for the site and the background are given in separate columns.

- **With Group Variable**: This option is used when data values of the variable (COPC) for the site and the background are given in the same column. The values are separated into different populations (groups) by the values of an associated Group Variable. When using this option, the user should select a group variable by clicking the arrow next to the **Group Var** option for a drop-down list of available variables.

- ProUCL 4.0 has been written using environmental terminology such as performing background versus site comparisons. However, all the tests and procedures in ProUCL 4.0 can be used for any other application if used properly. The user selects an appropriate group variable representing groups such as Background and AOC. For other applications such as comparing a new treatment drug versus older treatment drug, the group variable may represent the two groups: old drug group and new drug group. The user is allowed to use letters, numbers, or alphanumeric labels for the group names.

- When the Options button is clicked, the following window is shown.



- Specify a **Substantial Difference, S** value. The default choice is **0**.
- Choose the **Confidence** level. The default choice is **95%**.
- Select the form of Null Hypothesis. The default is AOC <= Background (Form 1).
- Click on OK button to continue or on Cancel button to cancel the selected options.

- Click on the OK button to continue or on the Cancel button to cancel the Site versus Background Comparison.

Output for Two-Sample Wilcoxon-Mann-Whitney Test (Full Data)

Area of Concern Data: Site

Background Data: Background

Raw Statistics

	Site	Background
Number of Valid Samples	10	10
Number of Distinct Samples	9	8
Minimum	15	23
Maximum	100	79
Mean	48.7	49.5
Median	34.5	50.5
SD	33.36	16.76
SE of Mean	10.55	5.3

Wilcoxon-Mann-Whitney (WMW) Test

H0: Mean/Median of Site or AOC <= Mean/Median of Background

Site Rank Sum W-Stat	97.5
WMW Test U-Stat	42.5
WMW Critical Value (0.050)	72
Approximate P-Value	0.727

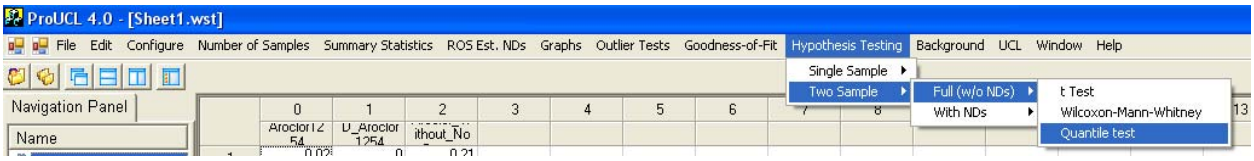
Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Site <= Background

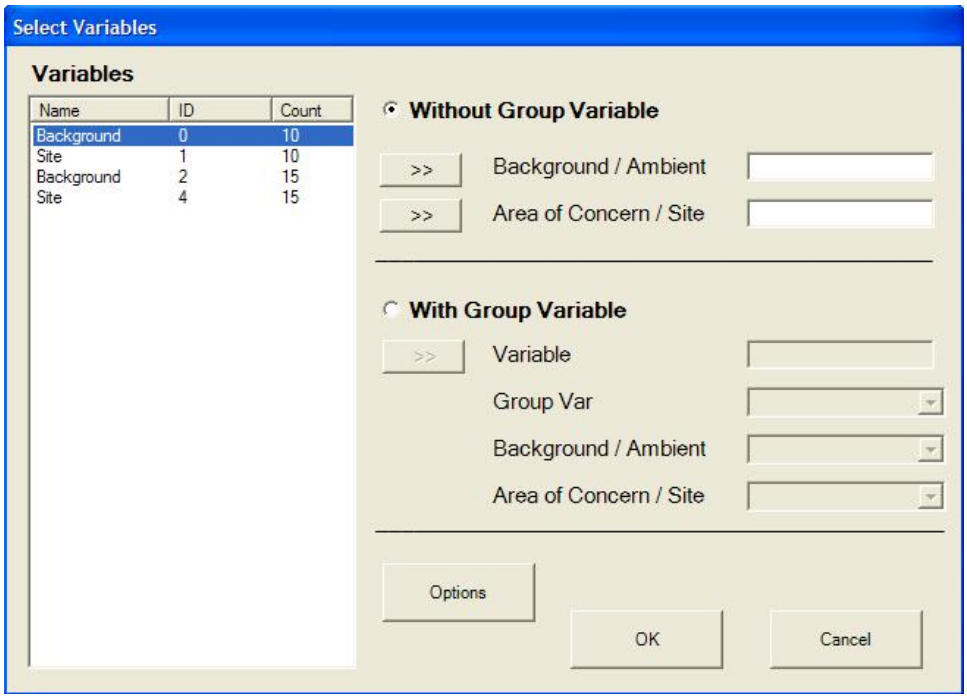
9.2.1.3 Two-Sample Quantile Test for Full Data without NDs

As mentioned before, the quantile test is often used in parallel with the WMW test. Typically, both tests are performed on the same data set before coming to the conclusion about comparability (or non-comparability) of the data distributions of the two populations.

1. Click **Hypothesis Testing ► Two Sample ► Full (w/o NDs) ► Quantile test**



2. The Select Variables Screen shown below will appear.

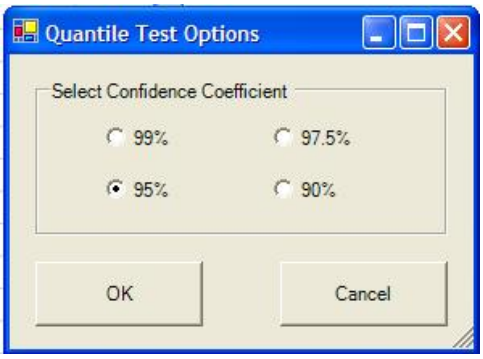


- Select variable (variables) from the Select Variables screen.

- **Without Group Variable**: This option is used when the data values of the variable (COPC) for the site and the background are given in different columns.

- **With Group Variable**: This option is used when data values of the variable (COPC) for the site and the background are given in the same column. The values are separated into different groups by using the values of the associated Group Variable. When using this option, the user should select a group variable by clicking the arrow next to the **Group Var** option for a drop-down list of available variables. The user selects an appropriate group variable representing groups such as Background and AOC. The user is allowed to use letters, numbers, or alphanumeric labels for the group names.

- When the Options button is clicked, the following window will be shown.



- Choose the **Confidence** level; the default choice is **95%**.
- Click on OK button to continue or on Cancel button to cancel the option.

- Click on the OK button to continue or on the Cancel button to cancel the Site versus Background Comparison.

Output for Two-Sample Quantile Test (Full Data)

Area of Concern Data: Site

Background Data: Background

Raw Statistics

	Site	Background
Number of Valid Samples	10	10
Number of Distinct Samples	9	8
Minimum	15	23
Maximum	100	79
Mean	48.7	49.5
Median	34.5	50.5
SD	33.36	16.76
SE of Mean	10.55	5.3

Quantile Test

H0: Site Concentration <= Background Concentration (Form 1)

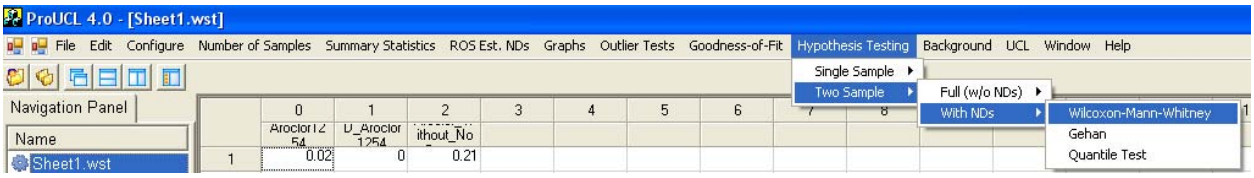
Approximate R Value (0.043)	4		
Approximate K Value (0.043)	4		
Number of Site Observations in 'R' Largest	3		
Calculated Alpha	0.0433		

Conclusion with Alpha = 0.043

Do Not Reject H0, Perform Wilcoxon-Mann-Whitney Ranked Sum Test

9.2.2 Two-Sample Hypothesis Testing for Data Sets with Nondetects

1. Click **Hypothesis Testing► Two Sample**

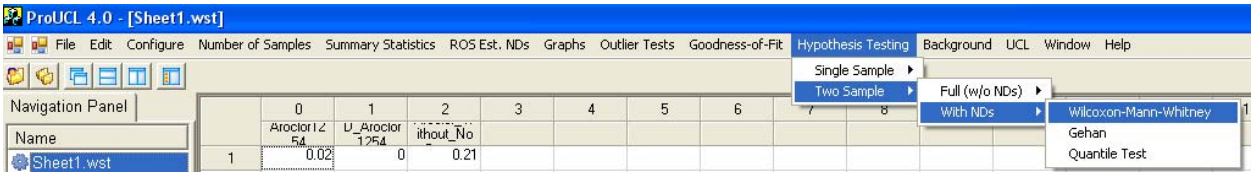


2. Select the With NDs option. A list of available tests will appear (shown above).

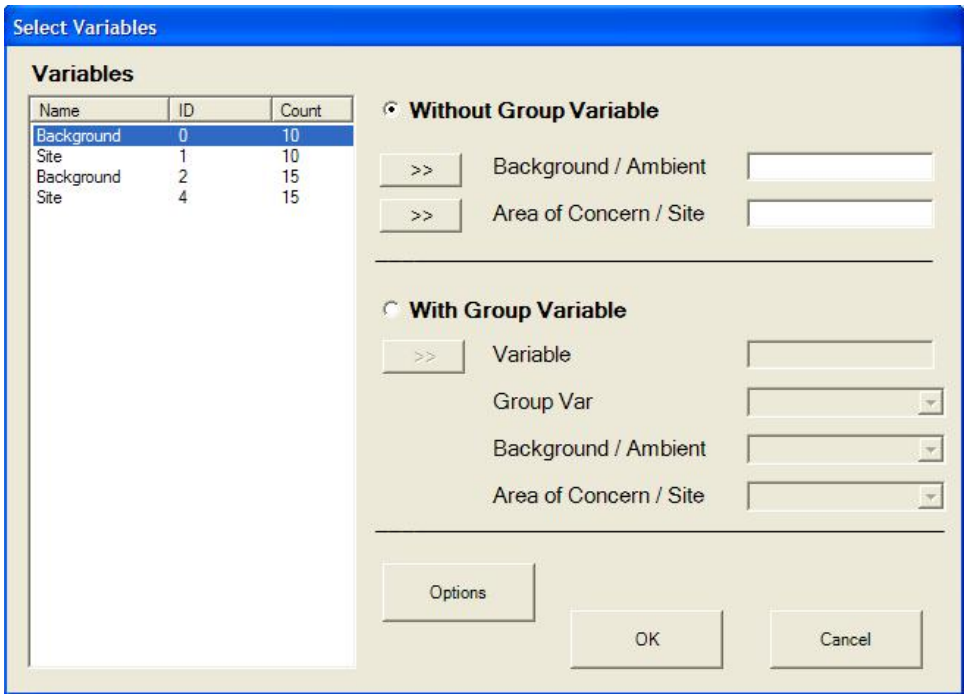
- To perform a Wilcoxon-Mann-Whitney test, click on Wilcoxon-Mann-Whitney from the drop-down menu list.
- To perform a Gehan test, click on Gehan from the drop-down menu.
- To perform a quantile test, click on Quantile Test from the drop-down menu.

9.2.2.1 Two-Sample Wilcoxon-Mann-Whitney Test with Nondetects

1. Click **Hypothesis Testing ► Two Sample ► With NDs ► Wilcoxon-Mann-Whitney**



2. The Select Variables Screen shown below will appear.

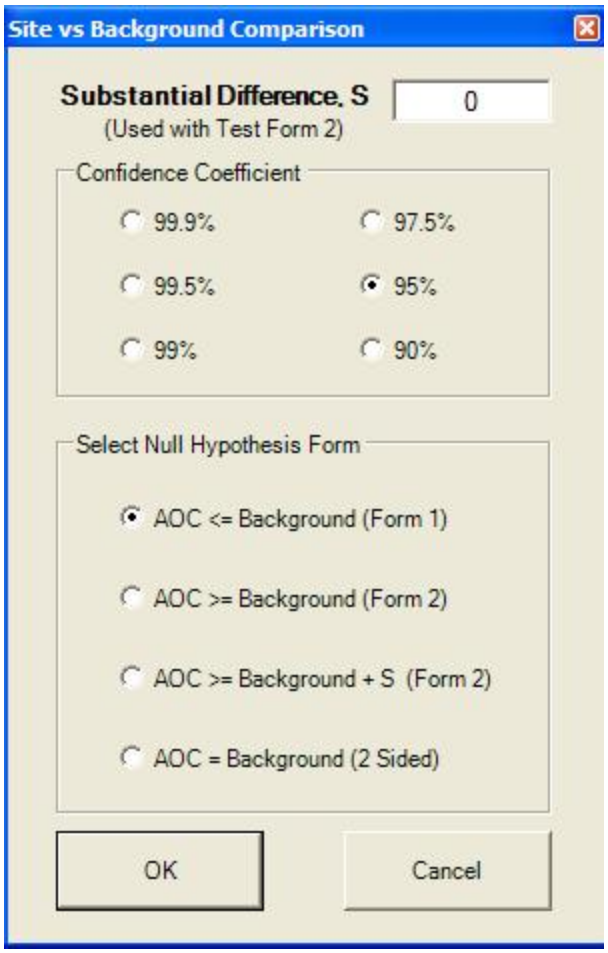


- Select variable (variables) from the Select Variables screen.

- **Without Group Variable**: This option is used when the data values of the variable (COPC) for the site and the background are given in separate columns.

- **With Group Variable**: This option is used when data values of the variable (COPC) for the site and the background are given in the same column. The values are separated into different populations (groups) by the values of an associated Group Variable. When using this option, the user should select a group variable by clicking the arrow next to the **Group Var** option for a drop-down list of available variables. The user selects an appropriate variable representing groups such as Background and AOC. The user is allowed to use letters, numbers, or alphanumeric labels for the group names.

- When the Options button is clicked, the following window will be shown.



- Specify a meaningful **Substantial Difference, S** value. The default choice is **0**.
- Choose the **Confidence** level. The default choice is **95%**.
- Select the form of Null Hypothesis. The default is AOC <= Background (Form 1).
- Click on the OK button to continue or on the Cancel button to cancel the selected options.

- Click on OK button to continue or on Cancel button to cancel the Site versus Background Comparison.

Output for Two-Sample Wilcoxon-Mann-Whitney Test (with Nondetects)

Area of Concern Data: Site

Background Data: Background

Raw Statistics

	Site	Background
Number of Valid Samples	15	15
Number of Non-Detect Data	8	12
Number of Detect Data	7	3
Minimum Non-Detect	5	3
Maximum Non-Detect	300	25
Percent Non detects	53.33%	80.00%
Minimum Detected	11	8
Maximum Detected	200	22
Mean of Detected Data	74.43	15
Median of Detected Data	70	15
SD of Detected Data	68.42	7

Wilcoxon-Mann-Whitney Site vs Background Test

All observations <= 300 (Max DL) are ranked the same

Wilcoxon-Mann-Whitney (WMW) Test

H0: Mean/Median of Site or AOC <= Mean/Median of Background

Site Rank Sum W-Stat	232.5
WMW Test U-Stat	112.5
WMW Critical Value (0.050)	152
Approximate P-Value	0.508

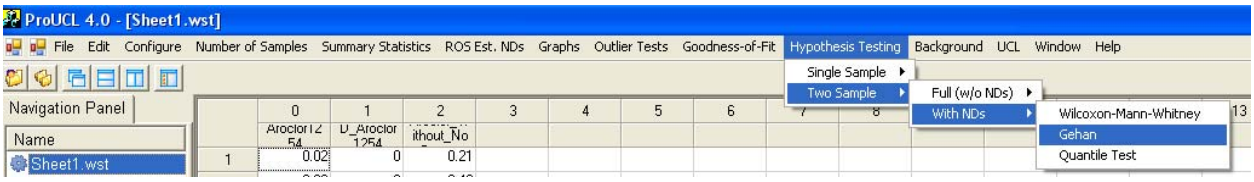
Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Site <= Background

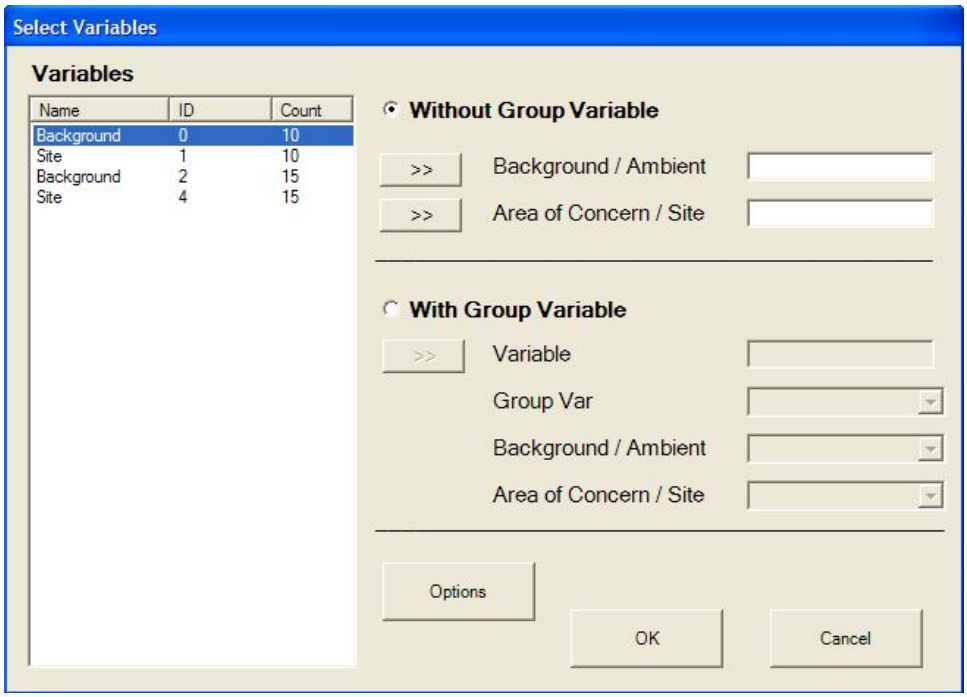
Note: *In the WMW test, all observations below the largest detection limit are considered as NDs (potentially including some detected values) and hence they all receive the same average rank. This action may reduce the associated power of the WMW test considerably. This in turn may lead to incorrect conclusion. As mentioned before, all hypotheses testing approaches should be supplemented with graphical displays such as Q-Q plots and box plots. When multiple detection limits are present, the use of the Gehan test is preferable.*

9.2.2.2 Two-Sample Gehan Test for Data Sets with Nondetects

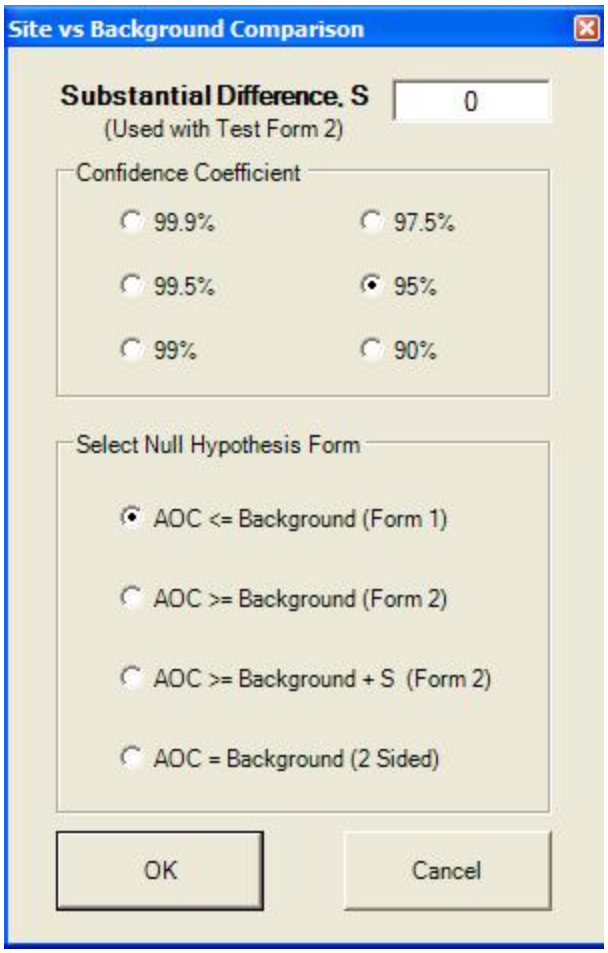
1. Click **Hypothesis Testing ► Two Sample ► With NDs ► Gehan**



2. The Select Variables Screen will appear.



- Select variable (variables) from the Select Variables screen.
- **Without Group Variable**: This option is used when the data values of the variable (COPC) for the site and the background are given in separate columns.
- **With Group Variable**: This option is used when data values of the variable (COPC) for the site and the background are given in the same column. The values are separated into different populations (groups) by the values of an associated Group Variable. When using this option, the user should select a group variable by clicking the arrow next to the **Group Var** option for a drop-down list of available variables. The user selects a group variable representing groups such as Background and AOC.
- When the Options button is clicked, the following window will be shown.



- Specify a **Substantial Difference, S** value. The default choice is **0**.
- Choose the **Confidence** level. The default choice is **95%**.
- Select the form of Null Hypothesis. The default is AOC <= Background (Form 1).
- Click on OK button to continue or on Cancel button to cancel selected options.

- Click on the OK button to continue or on the Cancel button to cancel the Site versus Background Comparison.

Output for Two-Sample Gehan Test (with Nondetects)

Area of Concern Data: Site

Background Data: Background

Raw Statistics

	Site	Background
Number of Valid Samples	10	10
Number of Non-Detect Data	2	4
Number of Detect Data	8	6
Minimum Non-Detect	4	4
Maximum Non-Detect	35	25
Percent Non detects	20.00%	40.00%
Minimum Detected	2	1
Maximum Detected	43	27
Mean of Detected Data	23.63	12.17
Median of Detected Data	22.5	11
SD of Detected Data	14.74	9.642

Site vs Background Gehan Test

H0: Mu of Site or AOC >= Mu of background

Gehan z Test Value	1.769
Critical z (0.05)	-1.645
P-Value	0.962

Conclusion with Alpha = 0.05

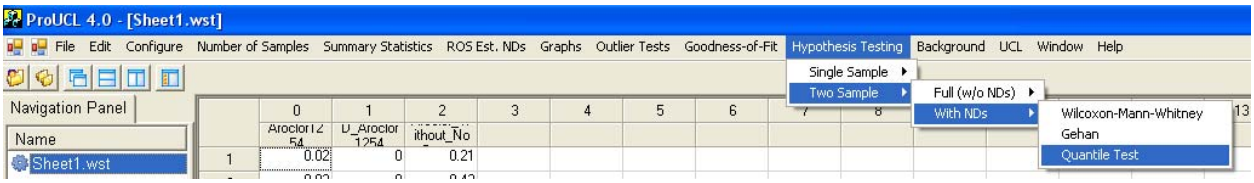
Do Not Reject H0, Conclude Site >= Background

P-Value >= alpha (0.05)

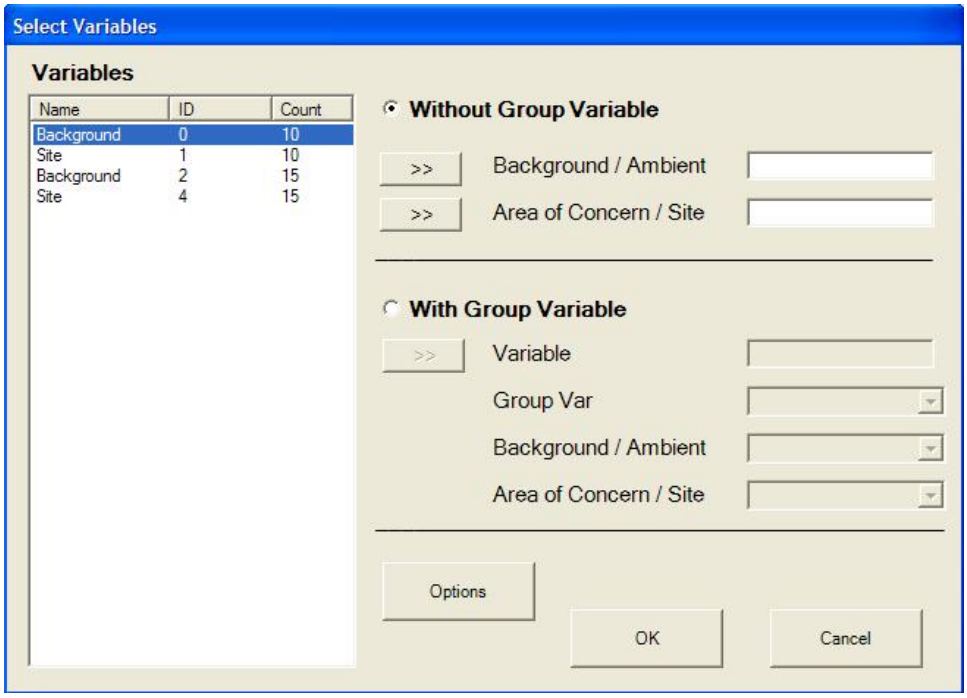
9.2.2.3 Two-Sample Quantile Test for Data Sets with Nondetects

Quantile test as described in EPA (1994) has been included in ProUCL 4.0. The detailed power of the test with many ND values is not well studied. The conclusion of this test should also be supplemented with graphical displays. The use of the Gehan test is preferred when the data set may consist of many NDs with multiple detection limits.

1. Click **Hypothesis Testing ► Two Sample ► With NDs ► Quantile Test**



2. The Select Variables Screen will appear.

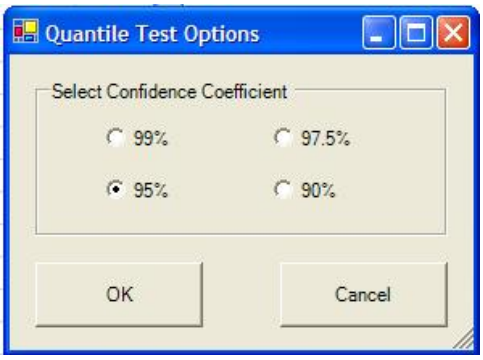


- Select variable (variables) from the Select Variables screen.

- **Without Group Variable**: This option is used when the data values of the variable (COPC) for the site and the background are given in separate columns.

- **With Group Variable**: This option is used when data values of the variable (COPC) for the site and the background are given in the same column. The values are separated into different populations (groups) by the values of an associated Group Variable. When using this option, the user should select a group variable by clicking the arrow next to the **Group Var** option for a drop-down list of available variables. The user selects an appropriate group variable representing groups such as Background and AOC.

- When the Options button is clicked, the following window will be shown.



- Choose the **Confidence** level; the default choice is **95%**.
- Click on OK button to continue or on Cancel button to cancel the option.

- Click on the OK button to continue or on the Cancel button to cancel the Site versus Background Comparison.

Output for Two-Sample Quantile Test (with Nondetects)

Area of Concern Data: Site

Background Data: Background

Raw Statistics

	Site	Background
Number of Valid Samples	15	15
Number of Non-Detect Data	8	12
Number of Detect Data	7	3
Minimum Non-Detect	5	3
Maximum Non-Detect	300	25
Percent Non detects	53.33%	80.00%
Minimum Detected	11	8
Maximum Detected	200	22
Mean of Detected Data	74.43	15
Median of Detected Data	70	15
SD of Detected Data	68.42	7

Quantile Test

H0: Site Concentration <= Background Concentration (Form 1)

Approximate R Value (0.05)	4
Approximate K Value (0.05)	4
Number of Site Observations in 'R' Largest	4

Non-Detect Values in the 'R' Largest - Cannot complete Quantile Test

Chapter 10

Background Statistics

This chapter illustrates the computations of various parametric and nonparametric statistics and upper limits that can be used as estimates of background threshold values (BTVs) and other not-to-exceed values. The BTV estimation methods are available for all data sets with and without nondetect (ND) observations. The details of those methods are given in Chapter 5 (full data sets without NDs) and Chapter 6 (data sets with NDs) of the revised background document for CERCLA sites (EPA, 2002). Technical details can also be found in the Technical Guide associated with ProUCL 4.0. For each selected variable, this option computes various upper limits such as UPLs, UTLs, and upper percentiles to estimate the background threshold values (BTVs) and other compliance limits that are used in site versus background evaluations.

As before, two choices for data sets are available to compute background statistics:

- Full – computes background statistics for a Full data set without any NDs.
- With NDs – computes background statistics for a data set with nondetected as wells as detected values. Multiple detection limits are allowed.

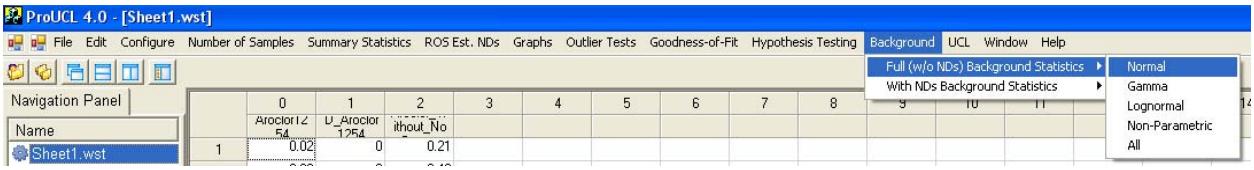
The user specifies the confidence level (probability) associated with each interval estimate. The reasonable confidence level as incorporated in ProUCL 4.0 represents a number in the interval [0.5, 1), 0.5 inclusive. The default choice is 0.95.

For data sets with and without NDs, ProUCL 4.0 can compute the following statistics that can be used as estimates of BTVs and not-to-exceed values.

- Parametric and nonparametric upper percentiles.
- Parametric and nonparametric upper prediction limits (UPLs) for a single observation, future or next k (≥ 1) observations, mean of next k observations. Here future k, or next k observations may also represent k observations from another population (e.g., site) different from the sampled (background) population (used to compute UPLs, UTLs).
- Parametric and nonparametric upper tolerance Limits (UTLs).
- Nonparametric IQR-based upper limits.

10.1 Background Statistics for Full Data Sets without Nondetects

1. Click **Background ► Full (w/o NDs) Background Statistics**

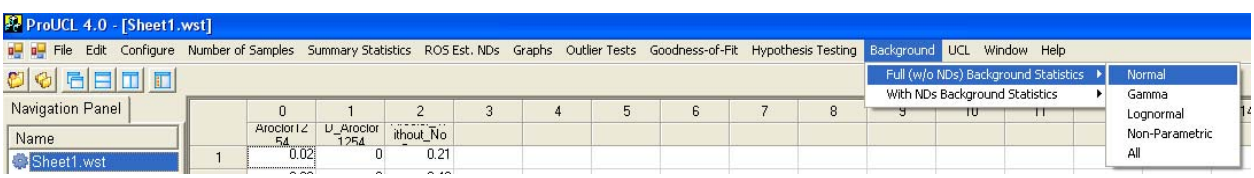


2. Select Full Background Statistics.

- To compute the background statistics assuming the normal distribution, click on Normal from the drop-down menu list.
- To compute the background statistics assuming the gamma distribution, click on Gamma from the drop-down menu list.
- To compute the background statistics assuming the lognormal distribution, click on Lognormal from the drop-down menu list.
- To compute the background statistics using distribution-free nonparametric methods, click on Non-Parametric from the drop-down menu list.
- To compute and see all background statistics available in ProUCL 4.0, click on the All option from the drop-down menu list.

10.1.1 Normal or Lognormal Distribution

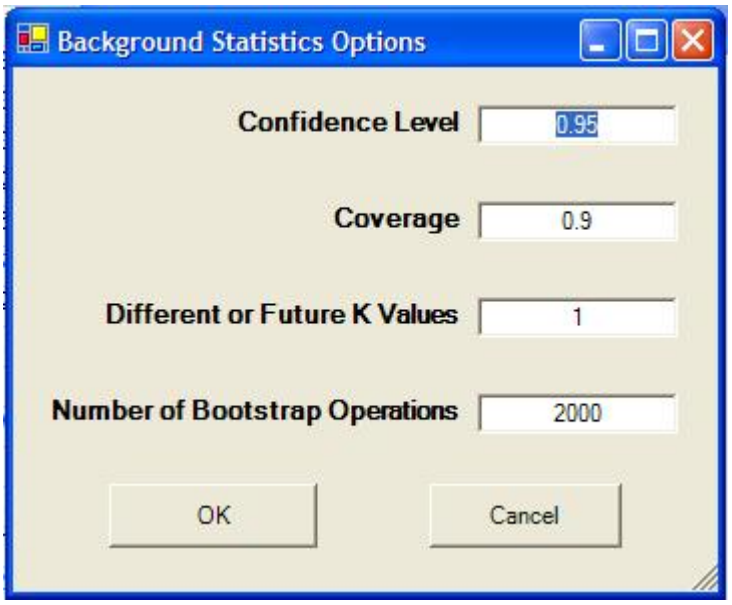
1. Click **Background ► Full (w/o NDs) Background Statistics ► Normal or Lognormal**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable to obtain a drop-down** list of available variables and select an appropriate group variable.

- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** coefficient (for a percentile) needed to compute UTLs. **Coverage** represents a number in the interval (0.0, 1). The default choice is **0.9**. Remember, a UTL is an upper confidence limit (e.g., with confidence level = 0.95) for a 90% (e.g., with coverage = 0.90) percentile.
  - Specify the **Different or Future K Values**. The default choice is **1**. It is noted that when K = 1, the resulting interval will be a UPL for a single future (or site) observations. In the example shown above, a value of K = 1 has been used.
  - Specify the **Number of Bootstrap Operations (resamples)**. The default choice is **2000**.
  - Click on OK button to continue or on Cancel button to cancel this option.

- Click on OK button to continue or on Cancel button to cancel the Background Statistics Options.

Output Screen for Normal Distribution (Full)

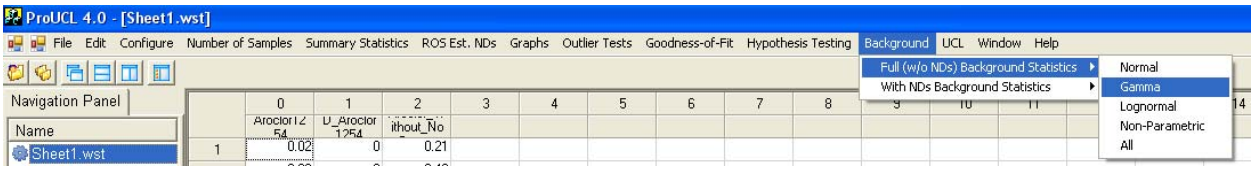
Cadmium	
Raw Statistics	
Number of Valid Samples	33
Number of Unique Samples	28
Minimum	0.094
Maximum	21
Second Largest	6.967
Mean	3.024
First Quantile	0.35
Median	2.5
Third Quantile	3.8
SD	3.751
Coefficient of Variation	1.24
Skewness	3.607
Normal Distribution Test	
Shapiro Wilk Test Statistic	0.642
5% Shapiro Wilk Critical Value	0.931
Data not Normal at 5% Significance Level	
Normal Distribution Test	
Shapiro Wilk Test Statistic	0.642
5% Shapiro Wilk Critical Value	0.931
Data not Normal at 5% Significance Level	
Background Statistics Assuming Normal Distribution	
90% Percentile (z)	7.83
95% Percentile (z)	9.193
99% Percentile (z)	11.75
95% UTL with 90% Coverage	9.549
95% UPL (t)	9.472
Note: UPL (or upper percentile for gamma distributed data) represents a preferred estimate of BTV	

Output Screen for Lognormal Distribution (Full)

Log-Transformed Statistics	
Number of Valid Samples	33
Number of Unique Samples	28
Minimum	-2.364
Maximum	3.045
Second Largest	1.941
Mean	0.459
First Quantile	1.653
Median	0.916
Third Quantile	0.709
SD	1.308
Lognormal Distribution Test	
Shapiro Wilk Test Statistic	0.909
5% Shapiro Wilk Critical Value	0.931
Data not Lognormal at 5% Significance Level	
Background Statistics Assuming Lognormal Distribution	
90% Percentile (z)	8.458
95% Percentile (z)	13.6
99% Percentile (z)	33.18
95% UPL	15
95% UTL with 90% Coverage	15.4
Background Statistics Assuming Lognormal Distribution	
90% Percentile (z)	8.458
95% Percentile (z)	13.6
99% Percentile (z)	33.18
95% UPL	15
95% UTL with 90% Coverage	15.4
Some Nonparametric Background Statistics	
95% Chebyshev UPL	19.62
95% Bootstrap BCA UTL with 90% Coverage	6.737
95% Percentile Bootstrap UTL with 90% Coverage	6.967
Note: UPL (or upper percentile for gamma distributed data) represents a preferred estimate of BTV	

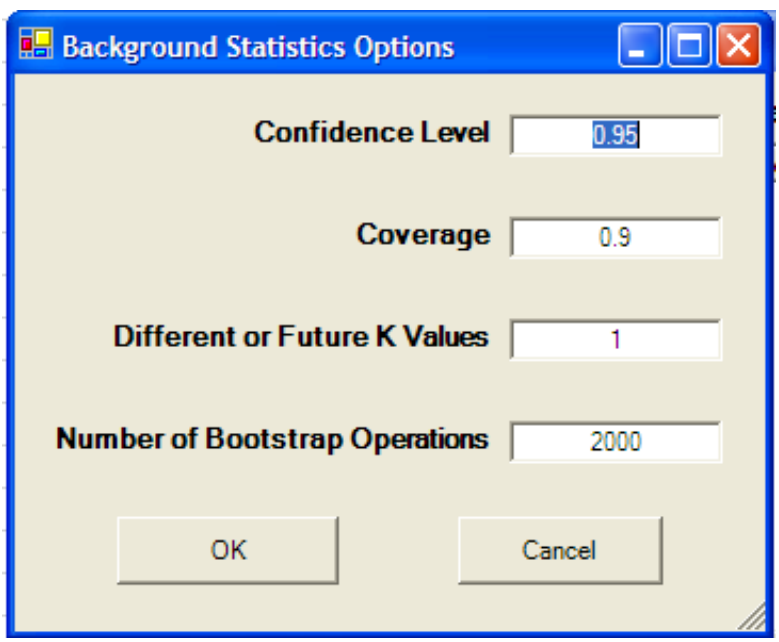
10.1.2 Gamma Distribution

1. Click **Background ► Full (w/o NDs) Background Statistics ► Gamma**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** level; a number in interval (0.0, 1). Default choice is **0.9**.
  - Specify the next **K.** The default choice is **1.**
  - Specify the **Number of Bootstrap Operations**. The default choice is **2000**.
  - Click on OK button to continue or on Cancel button to cancel the option.

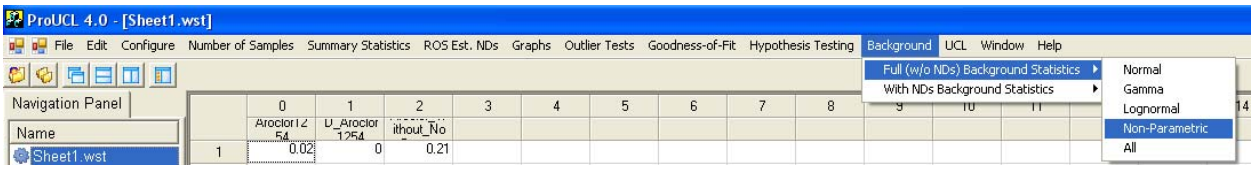
- Click on OK button to continue or on Cancel button to cancel the Background Statistics Options.

Output Screen for Gamma Distribution (Full)

Arsenic	
Raw Statistics	
Number of Valid Observations	24
Number of Distinct Observations	18
Minimum	0.66
Maximum	5.9
Second Largest	3.7
Mean	2.148
First Quartile	1.2
Median	2.05
Third Quartile	2.55
SD	1.159
Gamma Distribution Test	
k hat	4.153
Theta hat	0.517
nu hat	199.3
k star	3.662
Theta star	0.587
MLE of Mean	2.148
MLE of Standard Deviation	1.123
nu star	175.8
95% Percentile of Chisquare (2k)	14.54
A-D Test Statistic	0.341
5% A-D Critical Value	0.748
K-S Test Statistic	0.114
5% K-S Critical Value	0.179
Data appear Gamma Distributed at 5% Significance Level	
Background Statistics Assuming Gamma Distribution	
90% Percentile	3.654
95% Percentile	4.264
99% Percentile	5.574
95% Wilson Hilferty (WH) Approx. Gamma UPL	4.345
95% Hawkins Wixley (HW) Approx. Gamma UPL	4.397
95% Wilson Hilferty (WH) Approx. Gamma UTL with 90% Coverage	4.526
95% Hawkins Wixley (HW) Approx. Gamma UTL with 90% Coverage	4.592
Nonparametic Background Statistics	
95% Chebyshev UPL	7.306
95% BCA Bootstrap UTL with 90% Coverage	4.82
95% Bootstrap (%) UTL with 90% Coverage	5.02
Note: UPL represents a preferred estimate of BTV	

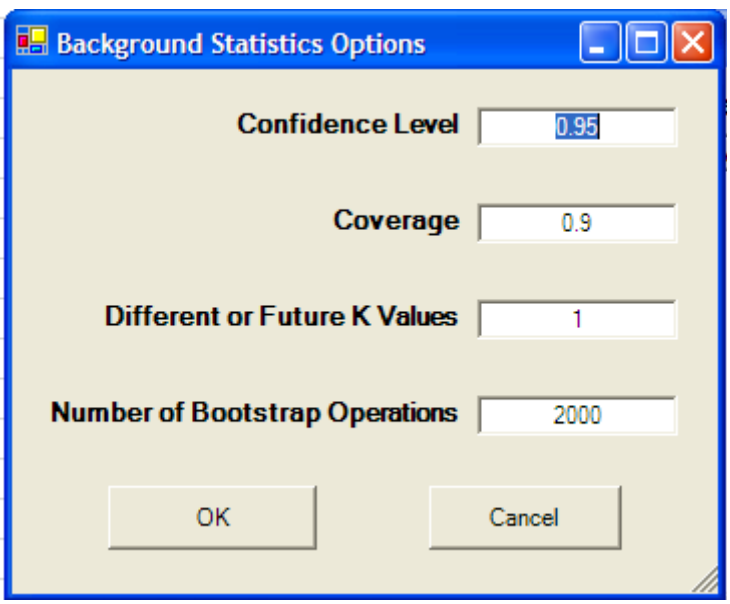
10.1.3 Nonparametric Methods

1. Click **Background ► Full (w/o NDs) Background Statistics ► Non-Parametric**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** level; a number in the interval (0.0, 1). Default choice is **0.9**.
  - Specify the next **K.** The default choice is **1.**
  - Specify the **Number of Bootstrap Operations**. The default choice is **2000**.
  - Click on the OK button to continue or on the Cancel button to cancel the option.

- Click on OK button to continue or on Cancel button to cancel the Background Statistics Options.

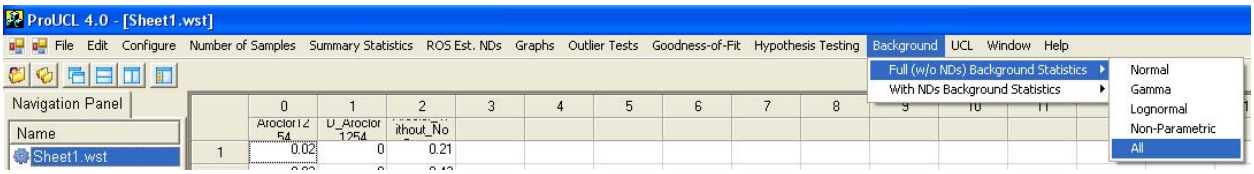
Output Screen for Nonparametric Option (Full)

Cadmium	
Some Non-Parametric Statistics	
Number of Valid Samples	33
Number of Unique Samples	28
Minimum	0.094
Maximum	21
Second Largest	6.967
Mean	3.024
First Quantile	0.35
Median	2.5
Third Quantile	3.8
SD	3.751
Variance	14.07
Coefficient of Variation	1.24
Skewness	3.607
Mean of Log-Transformed data	0.459
SD of Log-Transformed data	1.308

Data Follow Appr. Gamma Distribution at 5% Significance Level)	
Non-Parametric Background Statistics	
90% Percentile	5.27
95% Percentile	6.468
99% Percentile	16.37
95% UTL with 90% Coverage	
Order Statistic	32
Achieved CC	0.969
UTL	6.967
95% BCA Bootstrap UTL with 90% Coverage	6.737
95% Percentile Bootstrap UTL with 90% Coverage	6.967
95% UPL	11.18
95% Chebyshev UPL	19.62
Upper Limit Based upon IQR	8.975
Note: UPL (or upper percentile for gamma distributed data) represents a preferred estimate of BTV	

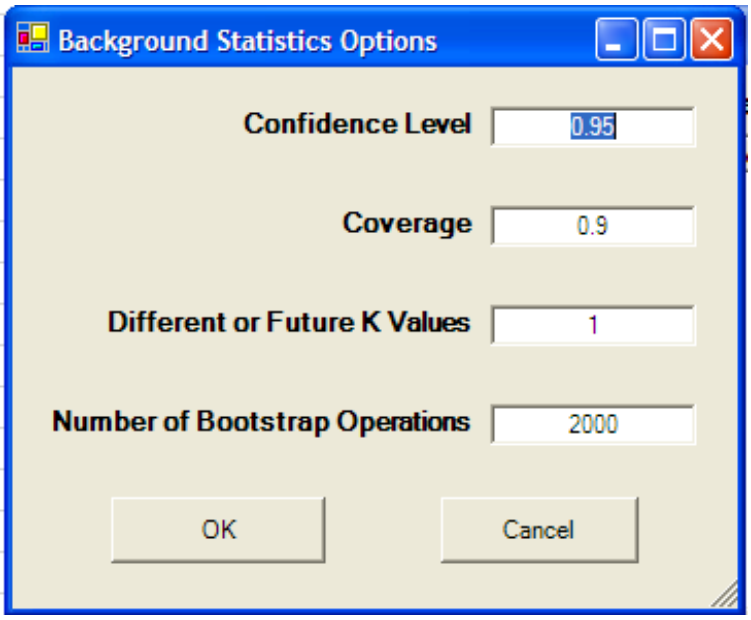
10.1.4 All Statistics Option

1. Click **Background ► Full Background Statistics ► All**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** level; a number in the interval (0.0, 1). Default is **0.9**.
  - Specify the next **K**. The default choice is **1**.
  - Specify the **Number of Bootstrap Operations**. The default choice is **2000.**
  - Click on OK button to continue or on Cancel button to cancel the option.

- Click on OK button to continue or on Cancel button to cancel the Background Statistics Options.

Output Screen for All Statistics Option (Full)

Arsenic

General Statistics

Total Number of Observations	24	Number of Distinct Observations	18

Raw Statistics		Log-Transformed Statistics	
Minimum	0.66	Minimum	-0.416
Maximum	5.9	Maximum	1.775
Second Largest	3.7	Second Largest	1.308
First Quartile	1.2	First Quartile	0.182
Median	2.05	Median	0.718
Third Quartile	2.55	Third Quartile	0.936
Mean	2.148	Mean	0.639
SD	1.159	SD	0.51
Coefficient of Variation	0.54		
Skewness	1.554		

Background Statistics

Normal Distribution Test		Lognormal Distribution Test	
Shapiro Wilk Test Statistic	0.875	Shapiro Wilk Test Statistic	0.977
Shapiro Wilk Critical Value	0.916	Shapiro Wilk Critical Value	0.916
Data not Normal at 5% Significance Level		Data appear Lognormal at 5% Significance Level	

Assuming Normal Distribution		Assuming Lognormal Distribution	
95% UTL with 90% Coverage	4.297	95% UTL with 90% Coverage	4.879
95% UPL (t)	4.176	95% UPL (t)	4.628
90% Percentile (z)	3.634	90% Percentile (z)	3.645
95% Percentile (z)	4.055	95% Percentile (z)	4.388
99% Percentile (z)	4.846	99% Percentile (z)	6.213

Gamma Distribution Test		Data Distribution Test
k star	3.662	Data appear Gamma Distributed at 5% Significance Level
Theta Star	0.587	
MLE of Mean	2.148	
MLE of Standard Deviation	1.123	
nu star	175.8	

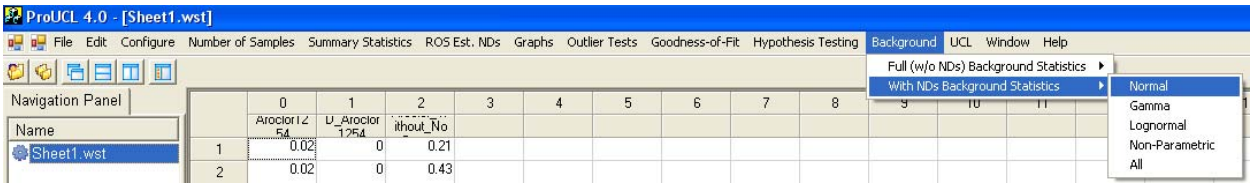
		Nonparametric Statistics	
A-D Test Statistic	0.341		
5% A-D Critical Value	0.748	90% Percentile	3.7
K-S Test Statistic	0.114	95% Percentile	5.35
5% K-S Critical Value	0.179	99% Percentile	5.9
Data appear Gamma Distributed at 5% Significance Level			

Assuming Gamma Distribution			
		95% UTL with 90% Coverage	3.7
90% Percentile	3.654	95% Percentile Bootstrap UTL with 90% Coverage	5.02
95% Percentile	4.264	95% BCA Bootstrap UTL with 90% Coverage	5.02
99% Percentile	5.574	95% UPL	5.35
		95% Chebyshev UPL	7.306
95% WH Approx. Gamma UPL	4.345	Upper Threshold Limit Based upon IQR	4.575
95% HW Approx. Gamma UPL	4.397		
95% WH Approx. Gamma UTL with 90% Coverage	4.526		
95% HW Approx. Gamma UTL with 90% Coverage	4.592		

Note: UPL represents a preferred estimate of BTV

10.2 Background Statistics with NDs

1. Click **Background ► With NDs Background Statistics**

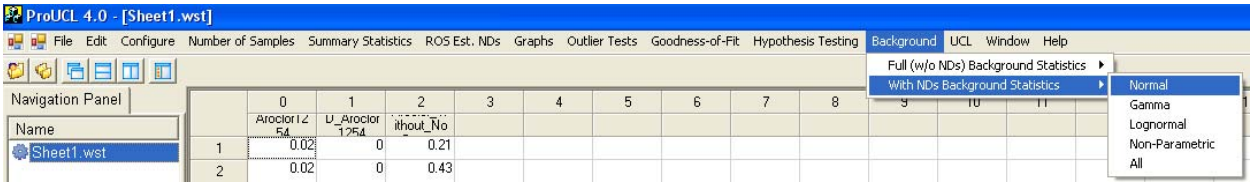


2. Select the With NDs Background Statistics option.

- To compute the background statistics assuming the normal distribution, click on Normal from the drop-down menu list.
- To compute the background statistics assuming the gamma distribution, click on Gamma from the drop-down menu list.
- To compute the background statistics assuming the lognormal distribution, click on Lognormal from the drop-down menu list.
- To compute the background statistics using distribution-free methods, click on Non-Parametric from the drop-down menu list.
- To compute all available background statistics in ProUCL 4.0, click on the All option from the drop-down menu list.

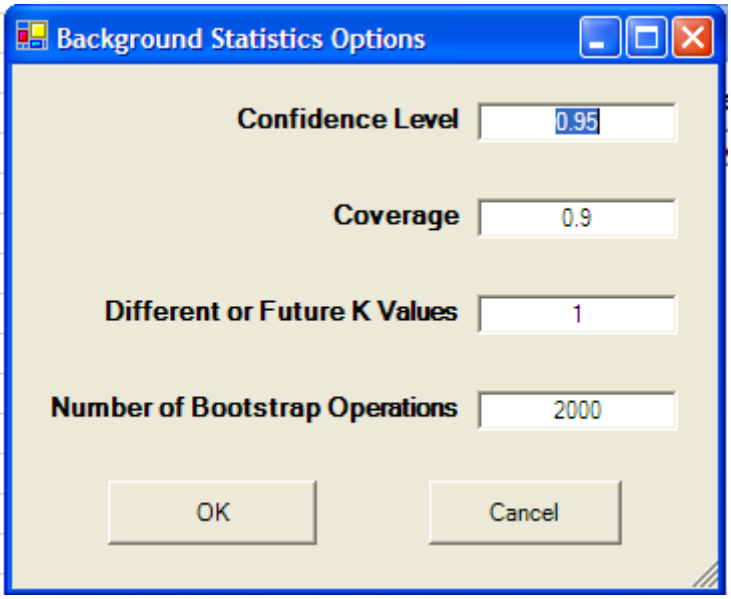
10.2.1 Normal or Lognormal Distribution

1. Click **Background ► With NDs Background Statistics ► Normal or Lognormal**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



- Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
- Specify the **Coverage** level; a number in the interval (0.0, 1). Default choice is **0.9**.
- Specify the next **K**. The default choice is **1**.
- Specify the **Number of Boostrap Operations.** The default choice is **2000.**
- Click on the OK button to continue or on the Cancel button to cancel the option.

- Click on OK button to continue or on Cancel button to cancel the Background Statistics Options.

Output Screen for Normal Distribution (with NDs)

Arsenic

Raw Statistics	
Total Number of Data	24
Number of Non-Detect Data	13
Number of Detected Data	11
Minimum Detected	0.5
Maximum Detected	3.2
Percent Non-Detects	54.17%
Minimum Non-detect	0.9
Maximum Non-detect	2
Mean of Detected Data	1.236
SD of Detected Data	0.965

Normal Distribution Test with Detected Values Only	
Shapiro Wilk Test Statistic	0.777
5% Shapiro Wilk Critical Value	0.85

Data not Normal at 5% Significance Level

Normal Distribution Test with Detected Values Only	
Shapiro Wilk Test Statistic	0.777
5% Shapiro Wilk Critical Value	0.85

Data not Normal at 5% Significance Level

Background Statistics Assuming Normal Distribution

DL/2 Substitution Method	
Mean	1.002
SD	0.699
95% UTL 90% Coverage	2.296
95% UPL (t)	2.224
90% Percentile (z)	1.897
95% Percentile (z)	2.151
99% Percentile (z)	2.627

Note: DL/2 is not a recommended method.

Maximum Likelihood Estimate (MLE) Method
MLE Method is Not Appliable for This Data

Maximum Likelihood Estimate (MLE) Method	
MLE Method is Not Appliable for This Data	
Kaplan-Meier (KM) Method	
Mean	0.949
SD	0.713
SE of Mean	0.165
95% UTL with 90% Coverage	2.27
95% KM UPL (t)	2.196
95% KM Chebyshev UPL	4.121
90% Percentile (z)	1.863
95% Percentile (z)	2.122
99% Percentile (z)	2.608
Note: UPL (or upper percentile for gamma distributed data) represents a preferred estimate of BTV. For an Example: KM-UPL may be used when multiple detection limits are present	

Output Screen for Lognormal Distribution (with NDs)

Arsenic

Log-Transformed Statistics	
Number of Valid Samples	24
Number of Unique Samples	10
Minimum	-0.693
Maximum	1.163
Second Largest	1.03
Mean	0.215
First Quantile	0.693
Median	0.203
Third Quantile	-0.693
SD	0.574

Lognormal Distribution Test	
Shapiro Wilk Test Statistic	0.906
5% Shapiro Wilk Critical Value	0.916

Data not Lognormal at 5% Significance Level

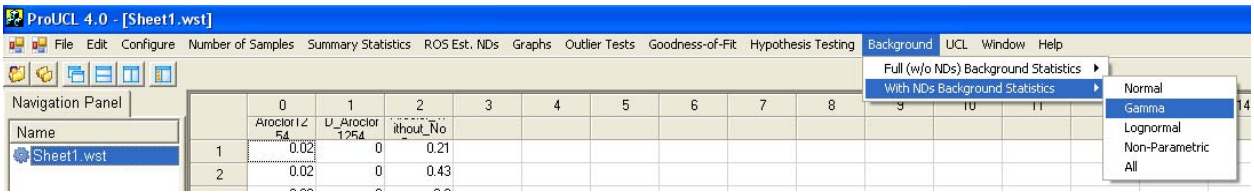
Background Statistics Assuming Lognormal Distribution	
90% Percentile (z)	2.587
95% Percentile (z)	3.186
99% Percentile (z)	4.711
95% UPL	3.383
95% UTL with 90% Coverage	3.59

Some Nonparametric Background Statistics	
95% Chebyshev UPL	4.825
95% Bootstrap BCA UTL with 90% Coverage	1.8
95% Percentile Bootstrap UTL with 90% Coverage	3.04

Note: UPL (or upper percentile for gamma distributed data) represents a preferred estimate of BTV

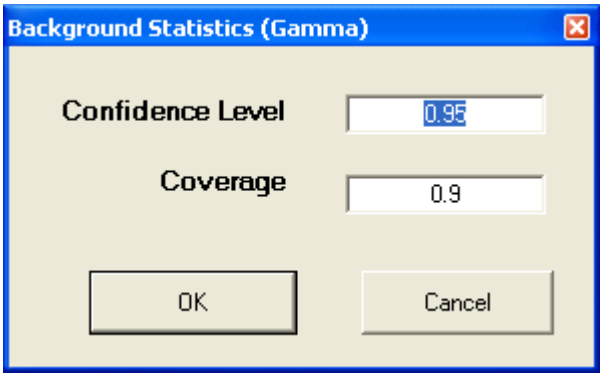
10.2.2 Gamma Distribution

1. Click **Background ► With NDs Background Statistics ► Gamma**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** level; a number in the interval (0.0, 1). Default choice is **0.9**.
  - Click on the OK button to continue or on the Cancel button to cancel option.

- Click on OK button to continue or on Cancel button to cancel the Background Statistics.

Output Screen for Gamma Distribution (with NDs)

X

Raw Statistics	
Total Number of Data	27
Number of Non-Detect Data	6
Number of Detected Data	21
Minimum Detected	28
Maximum Detected	118
Percent Non-Detects	22.22%
Minimum Non-detect	32
Maximum Non-detect	82
Mean of Detected Data	60.05
SD of Detected Data	21.02

Gamma Distribution Test with Detected Values Only	
k star	8.271
Theta star	7.26
nu star	347.4
95% Percentile of Chisquare (2k)	27
A-D Test Statistic	0.439
5% A-D Critical Value	0.743
K-S Test Statistic	0.172
5% K-S Critical Value	0.189

Data appear Gamma Distributed at 5% Significance Level

Background Statistics Assuming Gamma Distribution	
Gamma ROS Statistics with Extrapolated Data	
Mean	57.17
Median	54
SD	20.26
k Star	8.275
Theta Star	6.909
nu Star	446.8
95% Percentile of Chisquare (2k)	27.01
90% Percentile	83.66
95% Percentile	93.3
99% Percentile	113.2
95% Wilson Hilferty (WH) Approx. Gamma UPL	94.27
95% Hawkins Wixley (HW) Approx. Gamma UPL	94.74
ilson Hilferty (WH) Approx. Gamma UTL with 90% Coverage	96.26
wkins Wixley (HW) Approx. Gamma UTL with 90% Coverage	96.82
Kaplan Meier (KM) Method	
Mean	54.86
SD	21.25
SE of Mean	4.271
95% UTL 90% Coverage	93.34
95% KM Chebyshev UPL	149.2
95% KM UPL (t)	91.76
90% Percentile	82.09
95% Percentile (z)	89.81
99% Percentile (z)	104.3

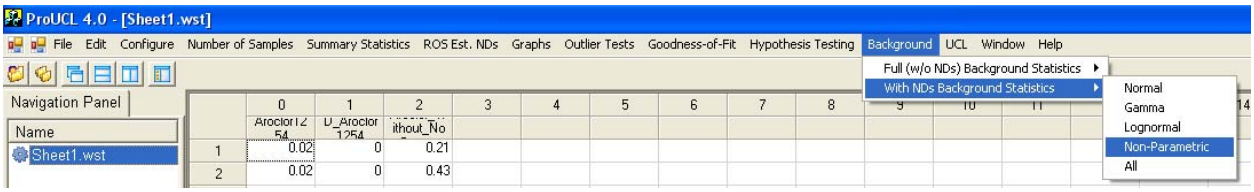
Note: UPL represents a preferred estimate of BTV

For an Example:

KM-UPL may be used when multiple detection limits are present

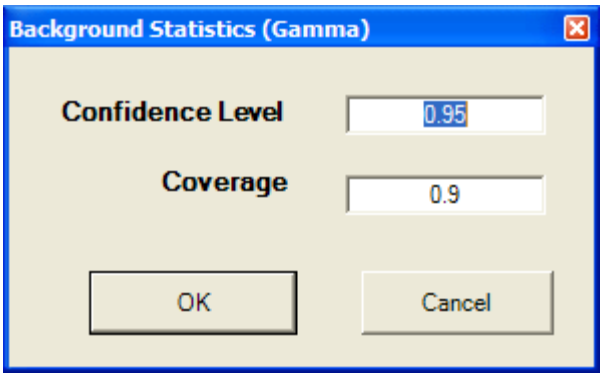
10.2.3 Nonparametric Methods (with NDs)

1. Click **Background ► With NDs Background Statistics ► Non-Parametric**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** level; a number in interval (0.0, 1). Default choice is **0.9**.
  - Click on the OK button to continue or on the Cancel button to cancel the option.

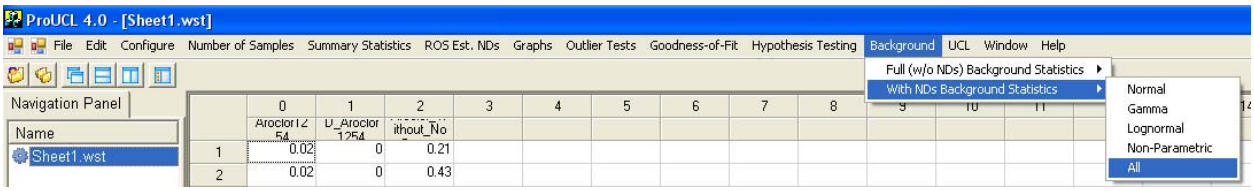
- Click on OK button to continue or on Cancel button to cancel the Background Statistics.

Output Screen for Nonparametric Methods (with NDs)

Arsenic	
Total Number of Data	24
Number of Non-Detect Data	13
Number of Detected Data	11
Minimum Detected	0.5
Maximum Detected	3.2
Percent Non-Detects	54.17%
Minimum Non-detect	0.9
Maximum Non-detect	2
Mean of Detected Data	1.236
SD of Detected Data	0.965
Mean of Log-Transformed Detected Data	-0.0255
SD of Log-Transformed Detected Data	0.694
Data Follow Appr. Gamma Distribution at 5% Significance Level	
Nonparametric Background Statistics	
95% UTL with 90% Coverage	
Order Statistic	23
Achieved CC	92.02
UTL	2.8
Largest Non-detect at Order	22
Kaplan-Meier (KM) Method	
Mean	0.949
SD	0.713
Standard Error of Mean	0.165
95% UTL 90% Coverage	2.27
95% KM Chebyshev UPL	4.121
95% KM UPL (t)	2.196
90% KM Percentile (z)	1.863
95% KM Percentile (z)	2.122
99% KM Percentile (z)	2.608
Note: UPL (or upper percentile for gamma distributed data) represents a preferred estimate of BTV. For an Example: KM-UPL may be used when multiple detection limits are present	

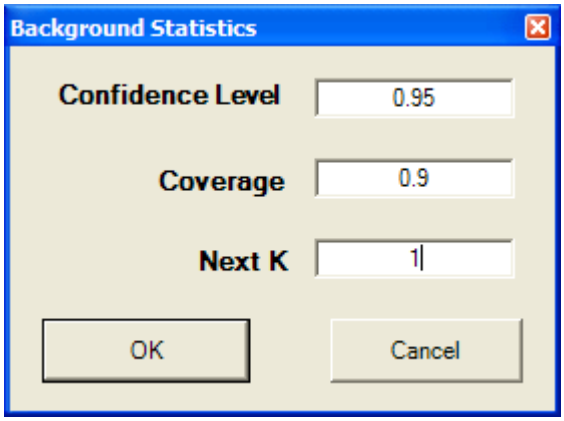
10.2.4 All Statistics Option

1. Click **Background ► With NDs Background Statistics ► All**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Coverage** level; a number in the interval (0.0, 1). Default choice is **0.9**.
  - Specify the **Next K**. The default choice is **1**.
  - Click on the OK button to continue or on the Cancel button to cancel the option.

- Click on OK button to continue or on Cancel button to cancel the Background Statistics.

Output Screen for All Statistics Option (with NDs)

X

General Statistics			
Number of Valid Data	27	Number of Detected Data	21
Number of Distinct Detected Data	17	Number of Non-Detect Data	6
		Percent Non-Detects	22.22%

Raw Statistics		Log-transformed Statistics	
Minimum Detected	28	Minimum Detected	3.332
Maximum Detected	118	Maximum Detected	4.771
Mean of Detected	60.05	Mean of Detected	4.042
SD of Detected	21.02	SD of Detected	0.329
Minimum Non-Detect	32	Minimum Non-Detect	3.466
Maximum Non-Detect	82	Maximum Non-Detect	4.407

Data with Multiple Detection Limits	Single Detection Limit Scenario	
Note: Data have multiple DLs - Use of KM Method is recommended	Number treated as Non-Detect with Single DL	24
For all methods (except KM, DL/2, and ROS Methods),	Number treated as Detected with Single DL	3
Observations < Largest ND are treated as NDs	Single DL Non-Detect Percentage	88.89%

Background Statistics			
Normal Distribution Test with Detected Values Only		Lognormal Distribution Test with Detected Values Only	
Shapiro Wilk Test Statistic	0.901	Shapiro Wilk Test Statistic	0.972
5% Shapiro Wilk Critical Value	0.908	5% Shapiro Wilk Critical Value	0.908
Data not Normal at 5% Significance Level		Data appear Lognormal at 5% Significance Level	
Assuming Normal Distribution		Assuming Lognormal Distribution	
DL/2 Substitution Method		DL/2 Substitution Method	
Mean	52.43	Mean (Log Scale)	3.855
SD	23.82	SD (Log Scale)	0.484
95% UTL 90% Coverage	95.56	95% UTL 90% Coverage	113.3
95% UPL (t)	93.8	95% UPL (t)	109.4
90% Percentile (z)	82.95	90% Percentile (z)	87.74
95% Percentile (z)	91.6	95% Percentile (z)	104.6
99% Percentile (z)	107.8	99% Percentile (z)	145.4

Maximum Likelihood Estimate(MLE) Method		Log ROS Method	
Mean	37.25	Mean in Original Scale	55.04
SD	36.93	SD in Original Scale	21.06
95% UTL with 90% Coverage	104.1	95% UTL with 90% Coverage	98.32
		95% BCA UTL with 90% Coverage	97.7
		95% Bootstrap (%) UTL with 90% Coverage	102.6
95% UPL (t)	101.4	95% UPL (t)	95.77
90% Percentile (z)	84.58	90% Percentile (z)	81.47
95% Percentile (z)	97.99	95% Percentile (z)	92.69
99% Percentile (z)	123.2	99% Percentile (z)	118.1
Gamma Distribution Test with Detected Values Only		**Data Distribution Test with Detected Values Only**	
k star (bias corrected)	8.271	Data appear Gamma Distributed at 5% Significance Level	
Theta Star	7.26		
nu star	347.4		
A-D Test Statistic	0.439	**Nonparametric Statistics**	
5% A-D Critical Value	0.743	Kaplan-Meier (KM) Method	
K-S Test Statistic	0.172	Mean	54.86
5% K-S Critical Value	0.189	SD	21.25
Data appear Gamma Distributed at 5% Significance Level		SE of Mean	4.271
		95% KM UTL with 90% Coverage	93.34
Assuming Gamma Distribution		95% KM Chebyshev UPL	149.2
Gamma ROS Statistics with Extrapolated Data		95% KM UPL (t)	91.76
Mean	57.17	90% Percentile (z)	82.09
Median	54	95% Percentile (z)	89.81
SD	20.26	99% Percentile (z)	104.3
k star	8.275		
Theta star	6.909	**Gamma ROS Limits with Extrapolated Data**	
Nu star	446.8	95% Wilson Hilferty (WH) Approx. Gamma UPL	94.27
95% Percentile of Chisquare (2k)	27.01	95% Hawkins Wixley (HW) Approx. Gamma UPL	94.74
		95% WH Approx. Gamma UTL with 90% Coverage	96.26
90% Percentile	83.66	95% HW Approx. Gamma UTL with 90% Coverage	96.82
95% Percentile	93.3		
99% Percentile	113.2		
Note: UPL represents a preferred estimate of BTV			
For an Example: KM-UPL may be used when multiple detection limits are present			
Note: DL/2 is not a recommended method.			

Chapter 11

Computing Upper Confidence Limits (UCLs) of Mean

The UCL computation module of ProUCL 4.0 represents an update of the UCL module of ProUCL 3.0. The detailed theory and formulae used to compute gamma and lognormal statistics are given by Land (1971, 1975), Gilbert (1987), Singh, Singh, and Engelhardt (1997, 1999), Singh et al. (2002a), Singh et al. (2002b), and Singh and Singh (2003).

Several parametric and nonparametric UCL computation methods for data sets with NDs have been incorporated in ProUCL 4.0. Methods such as the Kaplan-Meier (KM) and regression on order statistics (ROS) methods as incorporated in ProUCL 4.0 can handle multiple detection limits. For details regarding the distributions and methods available in ProUCL 4.0, refer to the ProUCL 4.0 Technical Guide and Singh, Maichle, and Lee (USEPA, 2006). Recommendations for the computations of UCLs for data sets with NDs have been made based upon the findings of the simulation experiments performed by Singh, Maichle, and Lee (USEPA, 2006).

In ProUCL 4.0, two choices are available to compute UCL statistics:

- Full – Computes UCLs for full data sets without any nondetected values.

- With NDs - Computes UCLs for data sets that have detected as well as BDL observations. It is pointed out that it is not desirable to use statistical methods as incorporated in ProUCL 4.0 on data sets consisting of all nondetect values. Discussion about the detection sampling frequency is provided in Chapter 1 of this User Guide. Some of the available methods can handle multiple detection limits. The program provides a message to the user about the use of an appropriate method when multiple detection limits may be present.

- For full data sets without NDs and also for data sets with NDs, the following options and choices are available to compute UCLs of the population mean.

    - The user specifies the confidence level; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is 0.95.
    - The program computes several nonparametric UCLs using the central limit theorem (CLT), Chebyshev inequality, jackknife, and bootstrap re-sampling methods.
    - For the bootstrap method, the user can select the number of bootstrap runs (re-samples). The default choice for the number of bootstrap runs is 2000.
    - The user is responsible for selecting an appropriate choice for the data distribution: normal, gamma, lognormal, or nonparametric. It is desirable that user determines data distribution using the Goodness-of-Fit test option prior to using the UCL option. The UCL option informs the user if data are normal, gamma, lognormal, or a non-discernable distribution. Program computes statistics depending on the user selection.
    - For data sets, which are not normal, one may try the gamma UCL next. The program will offer you advice if you chose the wrong UCL option.
    - For data sets, which are neither normal nor gamma, one may try the lognormal UCL. The program will offer you advice if you chose the wrong UCL option.

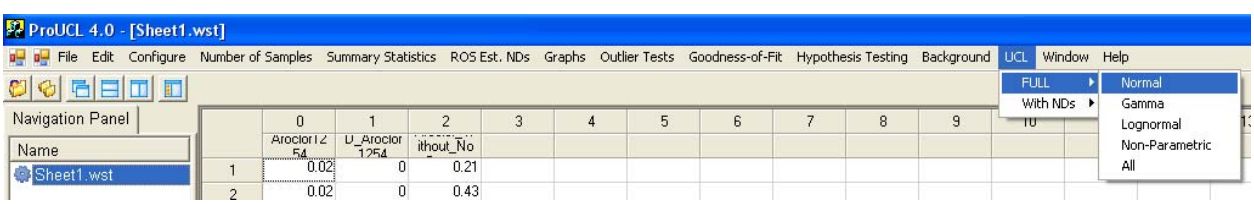
- Data sets that are not normal, gamma, or lognormal are classified as distribution-free nonparametric data sets. The user may use nonparametric UCL option for such data sets. The program will offer you advice if you chose the wrong UCL option.
- The program also provides the **All** option. By selecting this option, the UCLs are computed using most of the relevant methods available in ProUCL 4.0. The program informs the user about the distribution of the underlying data set, and offers advice regarding the use of an appropriate UCL.
- For lognormal data sets, ProUCL can compute only a 90% or a 95% Land's statistic-based H-UCL of the mean. For all other methods, ProUCL can compute a UCL for any confidence coefficient in the interval [0.5,1.0), 0.5 inclusive.
- If you have selected a distribution, then ProUCL will provide a recommended UCL computation method for 0.95, confidence coefficient. Even though ProUCL can compute UCLs for confidence coefficients in the interval [0.5, 1.0), recommendations are provided only for 95% UCL; as EPC term is estimated by a 95% UCL of the mean.

Note: *It is recommended that the user identify a few low probability outlying observations that may be present in the data set. Outliers distort many statistics of interest including summary statistics, data distributions, test statistics, UCLs, and estimates of BTVs. Decisions based upon distorted statistics may be misleading and incorrect. The objective is to compute relevant statistics and estimates based upon the majority of the data set(s) representing the dominant population(s). Those few low probability outlying observations require separate attention and investigation. The project team should decide about the proper disposition (to include or not to include) of outliers before computing the statistics to estimate the EPC terms and BTVs. In order to determine and compare the improper and unbalanced influence of outliers on UCLs and background statistics, the project team may want to compute statistics using data sets with outliers and without outliers.*

11.1 UCLs for Full Data Sets

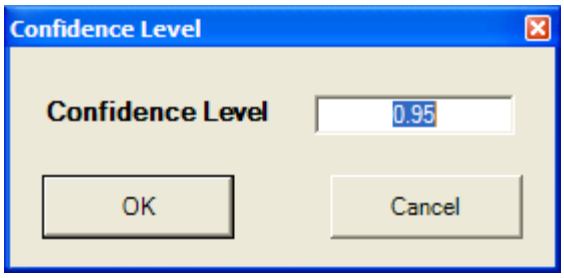
11.1.1 Normal Distribution (Full Data Sets without NDs)

1. Click **UCL ► Full ► Normal**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If needed, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



- Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
- Click on OK button to continue or on Cancel button to cancel the option.

- Click on OK button to continue or on Cancel button to cancel the UCL computation option.

Output Screen for Normal Distribution (Full Data without NDs)

	Normal UCL Statistics for Full Data Sets
User Selected Options	
From File	C:\Narain\ProUCL-Data\Data\Aroclor 1254.wst
Full Precision	OFF
Confidence Coefficient	95%

Aroclor_Without_NonDetects	
Number of Valid Samples	44
Number of Unique Samples	41
Minimum	0.21
Maximum	19000
Mean	1532
Median	94.5
SD	3355
Variance	11255595
Coefficient of Variation	2.19
Skewness	3.756
Shapiro Wilk Test Statistic	0.526
5% Shapiro Wilk Critical Value	0.944
Data not Normal at 5% Significance Level	

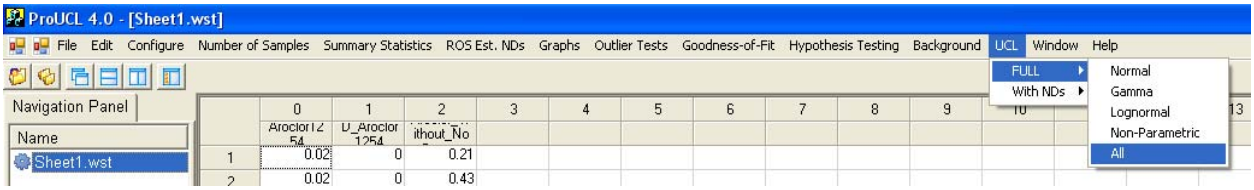
95% UCL (Assuming Normal Distribution)	
Student's-t LCL	681.6
Student's-t UCL	2382

Data appear Lognormal (0.05)

May want to try Lognormal UCLs

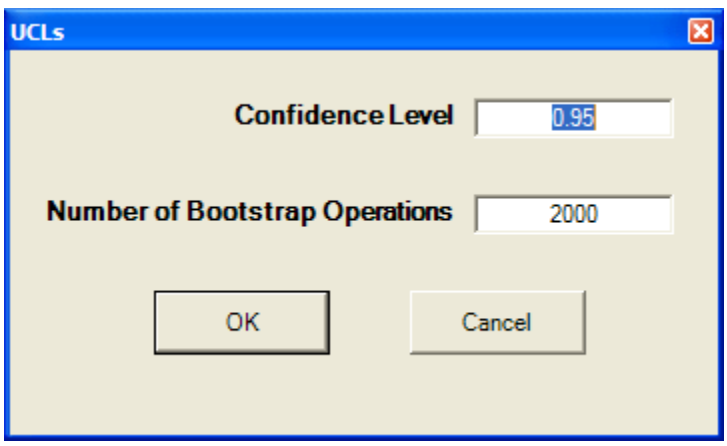
11.1.2 Gamma, Lognormal, Nonparametric, All Statistics Option (Full Data without NDs)

1. Click **UCL ► Full ► Gamma, Lognormal, Non-Parametric, or All**



2. The **Select Variables Screen** (Chapter 3) will appear.

- Select a variable (variables) from the Select Variables screen.
- If desired, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable.
- When the option button is clicked, the following window will be shown.



  - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
  - Specify the **Number of Bootstrap Operations (runs)**. Default choice is **2000**.
  - Click on OK button to continue or on Cancel button to cancel the UCLs option.

- Click on OK button to continue or on Cancel button to cancel the selected UCL computation option.

Output Screen for Gamma Distribution (Full)

	Gamma UCL Statistics for Full Data Sets
User Selected Options	
From File	C:\Narain\ProUCL-Data\Data\Aroclor 1254.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000

Aroclor_Without_NonDetects

Number of Valid Samples	44
Number of Unique Samples	41
Minimum	0.21
Maximum	19000
Mean	1532
Median	94.5
Standard Deviation	3355
Variance	11255595
k star (bias corrected)	0.247
Theta Star	6208
nu star	21.72
Approximate Chi Square Value (.05)	12.12
Adjusted Level of Significance	0.0445
Adjusted Chi Square Value (.05)	11.88

Anderson-Darling Test Statistic	1.01
Anderson-Darling Critical Value	0.884
Kolmogorov-Smirnov Test Statistic	0.169
Kolmogorov-Smirnov Critical Value	0.146

Data not Gamma distributed at 5% Significance Level

95% UCLs (Adjusted for Skewness)	
95% Adjusted-CLT UCL	2670
95% Modified-t UCL	2430
95% Non-Parametric UCLs	
95% Bootstrap-t UCL	3010
95% Hall's Bootstrap UCL	5380
95% Gamma UCLs(Assuming Gamma Distribution)	
95% Approximate Gamma UCL	2744
95% Approximate Gamma LCL	991
95% Adjusted Gamma UCL	2800
95% Adjusted Gamma LCL	976.8

Data appear Lognormal (0.05)

May want to try Lognormal UCLs

Output Screen for Lognormal Distribution (Full)

	Lognormal UCL Statistics for Full Data Sets
User Selected Options	
From File	C:\Narain\ProUCL-Data\Data\Aroclor 1254.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000
Aroclor_Without_NonDetects	
Number of Valid Samples	44
Number of Unique Samples	41
Minimum of log data	-1.561
Maximum of log data	9.852
Mean of log data	4.474
SD of log data	3.163
Variance of log data	10
Shapiro Wilk Test Statistic	0.948
Shapiro Wilk 5% Critical Value	0.944
Data appear Lognormal at 5% Significance Level	
95% UCL (Assuming Normal Distribution)	
95% Student's-t UCL	2382
ML Estimates Assuming Lognormal Distribution	
Mean	13038
SD	1938235
Coefficient of Variation	148.7
Skewness	3285837
Median	87.7
80% Quantile	1256
90% Quantile	5050
95% Quantile	15935
99% Quantile	137549
MVU Estimate of Median	78.25
MVU Estimate of Mean	7626
MVU Estimate of SD	165089
MVU Estimate of Standard Error of Mean	5553
UCLs (Assuming Lognormal Distribution)	
95% H-UCL	87513
95% Chebyshev (MVUE) UCL	31832
97.5% Chebyshev (MVUE) UCL	42307
99% Chebyshev (MVUE) UCL	62881
Potential UCL to Use	
Use 99% Chebyshev (MVUE) UCL	62881
Recommended UCL exceeds the maximum observation	

Output for Nonparametric Methods (Full Data without NDs)

	Nonparametric UCL Statistics for Full Data Sets
User Selected Options	
From File	C:\Narain\ProUCL-Data\Data\Aroclor 1254.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000

Aroclor_Without_NonDetects

Number of Valid Samples	44
Number of Unique Samples	41
Minimum	0.21
Maximum	19000
Mean	1532
Median	94.5
SD	3355
Variance	11255595
Coefficient of Variation	2.19
Skewness	3.756
Mean of log data	4.474
SD of log data	3.163

Non-Parametric UCLs	
95% CLT UCL	135.1
95% Jackknife UCL	136.1
95% Standard Bootstrap UCL	134.4
95% Bootstrap-t UCL	144.5
95% Hall's Bootstrap UCL	139.7
95% Percentile Bootstrap UCL	136.1
95% BCA Bootstrap UCL	141.3
95% Chebyshev(Mean, Sd) UCL	167.3
97.5% Chebyshev(Mean, Sd) UCL	189.7
99% Chebyshev(Mean, Sd) UCL	233.7
Potential UCL to Use	
Use 95% Student's-t UCL	136.1
Or 95% Modified-t UCL	136.8

Output Screen for All Statistics Option (Full Data without NDs)

	General UCL Statistics for Full Data Sets
User Selected Options	
From File	C:\Narain\ProUCL-Data\Data\Aroclor 1254.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000

Aroclor_Without_NonDetects

General Statistics

Number of Valid Samples	44	Number of Unique Samples	41

Raw Statistics		**Log-transformed Statistics**	
Minimum	0.21	Minimum of Log Data	-1.561
Maximum	19000	Maximum of Log Data	9.852
Mean	1532	Mean of log Data	4.474
Median	94.5	SD of log Data	3.163
SD	3355		
Coefficient of Variation	2.19		
Skewness	3.756		

Relevant UCL Statistics

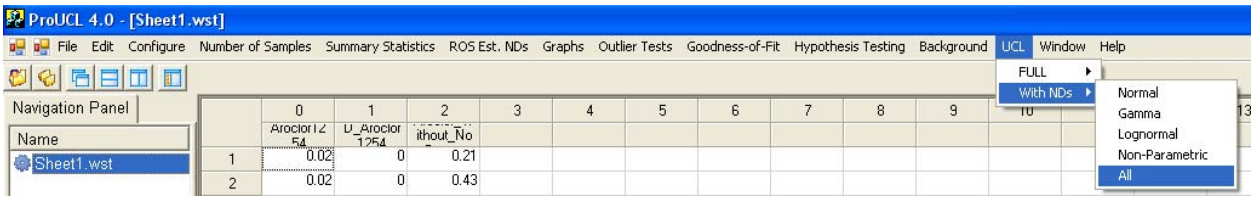
Normal Distribution Test		**Lognormal Distribution Test**	
Shapiro Wilk Test Statistic	0.526	Shapiro Wilk Test Statistic	0.948
Shapiro Wilk Critical Value	0.944	Shapiro Wilk Critical Value	0.944
Data not Normal at 5% Significance Level		Data appear Lognormal at 5% Significance Level	
Assuming Normal Distribution		**Assuming Lognormal Distribution**	
95% Student's-t UCL	2382	95% H-UCL	87513
95% UCLs (Adjusted for Skewness)		95% Chebyshev (MVUE) UCL	31832
95% Adjusted-CLT UCL	2670	97.5% Chebyshev (MVUE) UCL	42307
95% Modified-t UCL	2430	99% Chebyshev (MVUE) UCL	62881

Gamma Distribution Test		**Data Distribution Tests**	
k star (bias corrected)	0.247	Data appear Lognormal at 5% Significance Level	
Theta Star	6208		
nu star	21.72		
Approximate Chi Square Value (.05)	12.12	**Nonparametric Statistics**	
Adjusted Level of Significance	0.0445	95% CLT UCL	2364
Adjusted Chi Square Value	11.88	95% Jackknife UCL	2382
		95% Standard Bootstrap UCL	2352
Anderson-Darling Test Statistic	1.01	95% Bootstrap-t UCL	3035
Anderson-Darling 5% Critical Value	0.884	95% Hall's Bootstrap UCL	5521
Kolmogorov-Smirnov Test Statistic	0.169	95% Percentile Bootstrap UCL	2404
Kolmogorov-Smirnov 5% Critical Value	0.146	95% BCA Bootstrap UCL	2727
Data not Gamma Distributed at 5% Significance Level		95% Chebyshev(Mean, Sd) UCL	3737
		97.5% Chebyshev(Mean, Sd) UCL	4690
Assuming Gamma Distribution		99% Chebyshev(Mean, Sd) UCL	6564
95% Approximate Gamma UCL	2744		
95% Adjusted Gamma UCL	2800		
Potential UCL to Use		Use 99% Chebyshev (MVUE) UCL	62881
Recommended UCL exceeds the maximum observation			

Note: *Once again, it should be noted that the number of valid samples represents the total number of samples minus (-) the missing values (if any). The number of unique or distinct samples simply represents number of distinct observations. The information about the number of distinct samples is useful when using bootstrap methods. Specifically, it is not desirable to use bootstrap methods on data sets with only a few (< 4-5) distinct values.*

11.2 UCL for Data Sets with NDs

1. Click **UCL ► With NDs**



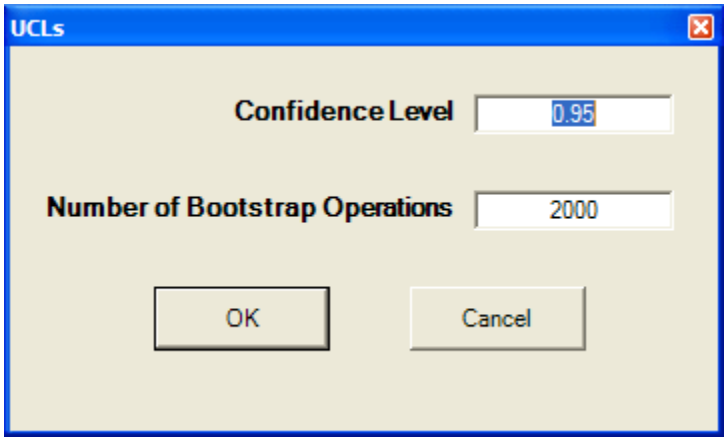
2. Choose the **Normal, Gamma, Lognormal, Non-Parametric,** or **All** option.

3. The **Select Variables Screen** (Chapter 3) will appear.

   - Select a variable (variables) from the Select Variables screen.

   - If desired, select a group variable by clicking the arrow below the **Group by variable** to obtain a drop-down list of available variables, and select a proper group variable. The selection of this option will compute the relevant statistics separately for each group that may be present in the data set.

   - When the option button is clicked, the following window will be shown.



   - Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
   - Specify the **Number of Bootstrap Operations**. The default choice is **2000**.

   - Click on OK button to continue or on Cancel button to cancel the UCLs option.

- Click on OK button to continue or on Cancel button to cancel the selected UCL computation option.

Output Screen for Normal Distribution (with NDs)

	Normal UCL Statistics for Data Sets with Non-Detects
User Selected Options	
From File	D:\example.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000
Arsenic	
Total Number of Data	20
Number of Non-Detect Data	3
Number of Detected Data	17
Minimum Detected	5
Maximum Detected	9.2
Percent Non-Detects	15.00%
Minimum Non-detect	4.3
Maximum Non-detect	4.5
Mean of Detected Data	6.126
SD of Detected Data	1.15
Note: Data have multiple DLs - Use of KM Method is recommended	
For all methods (except KM, DL/2, Robust ROS, and Gamma ROS)	
the Largest DL value is used for all NDs	

Output Screen for Normal Distribution (with NDs) – Continued

Normal Distribution Test with Detected Values Only	
Shapiro Wilk Test Statistic	0.794
5% Shapiro Wilk Critical Value	0.892
Data not Normal at 5% Significance Level	
DL/2 Substitution Method	
Mean	5.54
SD	1.779
95% DL/2 (t) UCL	6.228
Maximum Likelihood Method	
Mean	5.774
SD	1.348
95% MLE (t) UCL	6.295
95% MLE (Tiku) UCL	6.301
Kaplan Meier(KM) Method	
Mean	5.958
SD	1.104
Standard Error of Mean	0.255
95% KM (t) UCL	6.398
95% KM (z) UCL	6.376
95% KM (BCA) UCL	6.393
95% KM (Percentile Bootstrap) UCL	6.388
Data do not follow a Discernable Distribution (0.05)	
May want to try Non-Parametric UCLs	

Output Screen for Gamma Distribution (with NDs)

	Gamma UCL Statistics for Data Sets with Non-Detects
User Selected Options	
From File	D:\example.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000

Arsenic

Total Number of Data	20
Number of Missing Values	20
Number of Non-Detect Data	3
Number of Detected Data	17
Minimum Detected	5
Maximum Detected	9.2
Percent Non-Detects	15.00%
Minimum Non-detect	4.3
Maximum Non-detect	4.5
Mean of Detected Data	6.126
Median of Detected Data	5.8
SD of Detected Data	1.15
k Star of Detected Data	28.99
Theta Star of Detected Data	0.211
Nu Star of Detected Data	985.5

Note: Data have multiple DLs - Use of KM Method is recommended
For all methods (except KM, DL/2, Robust ROS, and Gamma ROS),
the Largest DL value is used for all NDs

Output Screen for Gamma Distribution (with NDs) – Continued

Gamma Distribution Test with Detected Values Only	
A-D Test Statistic	1.027
5% A-D Critical Value	0.737
K-S Test Statistic	0.214
5% K-S Critical Value	0.209
Data not Gamma Distributed at 5% Significance Level	
Gamma ROS Statistics with Extrapolated Data	
Minimum	3.516
Maximum	9.2
Mean	5.804
Median	5.625
SD	1.325
k Star	18.24
Theta Star	0.318
Nu Star	729.8
95% Percentile of Chisquare (2k)	51.58
AppChi2	668.1
95% Gamma Approximate UCL	6.34
95% Gamma Adjusted UCL	6.384
Kaplan Meier(KM) Method	
Mean	5.958
SD	1.104
Standard Error of Mean	0.255
95% KM (t) UCL	6.398
95% KM (BCA) UCL	6.455
95% KM (Percentile Bootstrap) UCL	6.405
95% KM (Chebyshev) UCL	7.067
Data do not follow a Discernable Distribution (0.05)	
May want to try Nonparametric UCLs	

Output Screen for Lognormal Distribution (with NDs)

	Lognormal UCL Statistics for Data Sets with Non-Detects		
User Selected Options			
From File	D:\example.wst		
Full Precision	OFF		
Confidence Coefficient	95%		
Number of Bootstrap Operations	2000		
Arsenic			
General Statistics			
Number of Valid Samples	20	Number of Detected Data	17
Number of Unique Samples	15	Number of Non-Detect Data	3
		Percent Non-Detects	15.00%
Raw Statistics			
Maximum Detected	9.2	Maximum Non-detect	4.5
Mean of Detected Data	6.126	SD of Detected Data	1.15
Log-Transformed Statistics			
Mean of Detected Data	1.798	SD of Detected Data	0.168
Note: Data have multiple DLs - Use of KM Method is recomm		Number treated as Non-Detect	3
For all methods (except KM, DL/2, Robust ROS, and Gamma		Number treated as Detected	17
those Observations < Largest ND are treated as NDs		Single DL Non-Detect Percentage	15.00%
Lognormal Distribution Test with Detected Values Only			
Shapiro Wilk Test Statistic	0.854	5% Shapiro Wilk Critical Value	0.892
Data not Lognormal at 5% Significance Level			

Note: *Once again, it should be noted that the number of valid samples represents total number of samples minus (-) the missing values (if any). The number of unique or distinct samples simply represents number of distinct observations. The information about the number of distinct samples is useful when using bootstrap methods. Specifically, it is not desirable to use bootstrap methods on data sets with only a few (< 4-5) distinct values.*

Output Screen for Lognormal Distribution (with NDs) – Continued

DL/2 Substitution Method			
Mean	5.54	95% H-Stat (DL/2) UCL	6.522
SD	1.779		
Mean (in Log Scale)	1.648		
SD (in Log Scale)	0.398		
Robust ROS Method			
Mean	5.831	95% Percentile Bootstrap UCL	6.318
SD	1.28	95% BCA Bootstrap UCL	6.357
Mean (in Log Scale)	1.742		
SD (in Log Scale)	0.207		
Kaplan Meier (KM) Method			
Mean	5.958	95% KM (t) UCL	6.398
SD	1.104	95% KM (BCA) UCL	6.43
SE of Mean	0.255	95% KM (% Bootstrap) UCL	6.373
		95% KM (Chebyshev) UCL	7.067
		97.5% KM (Chebyshev) UCL	7.547
		99% KM (Chebyshev) UCL	8.49
Potential UCL to Use		Data do not follow a Discernable Distribution (0.05)	
		May want to try Nonparametric UCLs	

Output Screen for Nonparametric Methods (with NDs)

	Nonparametric UCL Statistics for Data Sets with Non-Detects
User Selected Options	
From File	D:\example.wst
Full Precision	OFF
Confidence Coefficient	95%
Number of Bootstrap Operations	2000

Arsenic

Total Number of Data	20
Number of Non-Detect Data	3
Number of Detected Data	17
Minimum Detected	5
Maximum Detected	9.2
Percent Non-Detects	15.00%
Minimum Non-detect	4.3
Maximum Non-detect	4.5
Mean of Detected Data	6.126
Median of Detected Data	5.8
Variance of Detected Data	1.323
SD of Detected Data	1.15
CV of Detected Data	0.188
Skewness of Detected Data	1.783
Mean of Detected log data	1.798
SD of Detected Log data	0.168

Note: Data have multiple DLs - Use of KM Method is recommended
For all methods (except KM, DL/2, Robust ROS, and Gamma ROS),
the Largest DL value is used for all NDs

Output for Nonparametric Methods (with NDs) – Continued

Nonparametric Test with Detected Values Only	
Data do not follow a Discernable Distribution (0.05)	
Winsorization Method	0.168
Mean	5.738
SD	0.624
95% Winsor (t) UCL	5.985
Kaplan Meier (KM) Method	
Mean	5.958
SD	1.104
Standard Error of Mean	0.255
95% KM (t) UCL	6.398
95% KM (z) UCL	6.376
95% KM (BCA) UCL	6.478
95% KM (Percentile Bootstrap) UCL	6.438
95% KM (Chebyshev) UCL	7.067
97.5% KM (Chebyshev) UCL	7.547
99% KM (Chebyshev) UCL	8.49
Potential UCL to Use	
95% KM (Chebyshev) UCL	7.067

Output Screen for All Statistics Option (with NDs)

	General UCL Statistics for Data Sets with Non-Detects		
User Selected Options			
From File	D:\example.wst		
Full Precision	OFF		
Confidence Coefficient	95%		
Number of Bootstrap Operations	2000		
Arsenic			
	General Statistics		
Number of Valid Samples	20	Number of Detected Data	17
Number of Unique Samples	15	Number of Non-Detect Data	3
		Percent Non-Detects	15.00%
Raw Statistics		**Log-transformed Statistics**	
Minimum Detected	5	Minimum Detected	1.609
Maximum Detected	9.2	Maximum Detected	2.219
Mean of Detected	6.126	Mean of Detected	1.798
SD of Detected	1.15	SD of Detected	0.168
Minimum Non-Detect	4.3	Minimum Non-Detect	1.459
Maximum Non-Detect	4.5	Maximum Non-Detect	1.504
Note: Data have multiple DLs - Use of KM Method is recommended		Number treated as Non-Detect	3
For all methods (except KM, DL/2, Robust ROS, and Gamma ROS),		Number treated as Detected	17
those Observations < Largest ND are treated as NDs		Single DL Non-Detect Percentage	15.00%
	UCL Statistics		
Normal Distribution Test with Detected Values Only		**Lognormal Distribution Test with Detected Values Only**	
Shapiro Wilk Test Statistic	0.794	Shapiro Wilk Test Statistic	0.854
5% Shapiro Wilk Critical Value	0.892	5% Shapiro Wilk Critical Value	0.892
Data not Normal at 5% Significance Level		**Data not Lognormal at 5% Significance Level**	

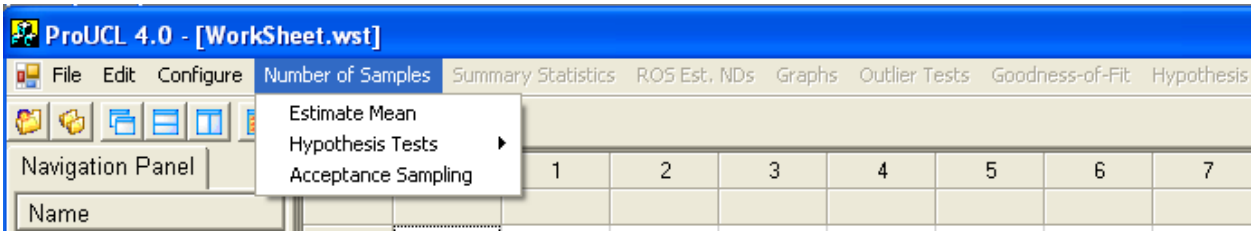
Note: *Once again, it should be noted that the number of valid samples represents the total number of samples minus (-) the missing values (if any). The number of unique or distinct samples simply represents number of distinct observations. The information about the number of distinct samples is useful when using bootstrap methods. Specifically, it is not desirable to use bootstrap methods on data sets with only a few (< 4-5) distinct values.*

Output Screen for All Statistics Option (with NDs) – Continued

Assuming Normal Distribution		**Assuming Lognormal Distribution**	
DL/2 Substitution Method		DL/2 Substitution Method	
Mean	5.54	Mean	1.648
SD	1.779	SD	0.398
95% DL/2 (t) UCL	6.228	95% H-Stat (DL/2) UCL	6.522
Maximum Likelihood Estimate(MLE) Method		Robust ROS Method	
Mean	5.774	Mean in Log Scale	1.742
SD	1.348	SD in Log Scale	0.207
95% MLE (t) UCL	6.295	Mean in Original Scale	5.831
95% MLE (Tiku) UCL	6.301	SD in Original Scale	1.28
		95% Percentile Bootstrap UCL	6.314
		95% BCA Bootstrap UCL	6.37
Gamma Distribution Test with Detected Values Only		**Nonparametic Test with Detected Values Only**	
k star (bias corrected)	28.99	**Data do not follow a Discernable Distribution (0.05)**	
Theta Star	0.211		
nu star	985.5		
A-D Test Statistic	1.027	**Nonparametric Statistics**	
5% A-D Critical Value	0.737	Kaplan-Meier (KM) Method	
K-S Test Statistic	0.737	Mean	5.958
5% K-S Critical Value	0.209	SD	1.104
Data not Gamma Distributed at 5% Significance Level		SE of Mean	0.255
		95% KM (t) UCL	6.398
Assuming Gamma Distribution		95% KM (z) UCL	6.376
Gamma ROS Statistics using Extrapolated Data		95% KM (jackknife) UCL	6.387
Minimum	3.516	95% KM (bootstrap t) UCL	6.673
Maximum	9.2	95% KM (BCA) UCL	6.43
Mean	5.804	95% KM (percentile) UCL	6.398
Median	5.625	95% KM (Chebyshev) UCL	7.067
SD	1.325	97.5% KM (Chebyshev) UCL	7.547
k star	18.24	99% KM (Chebyshev) UCL	8.49
Theta star	0.318		
Nu star	729.8	**Potential UCLs to Use**	
AppChi2	668.1	95% KM (Chebyshev) UCL	7.067
95% Gamma Approximate UCL	6.34		
95% Adjusted Gamma UCL	6.384		

Chapter 12

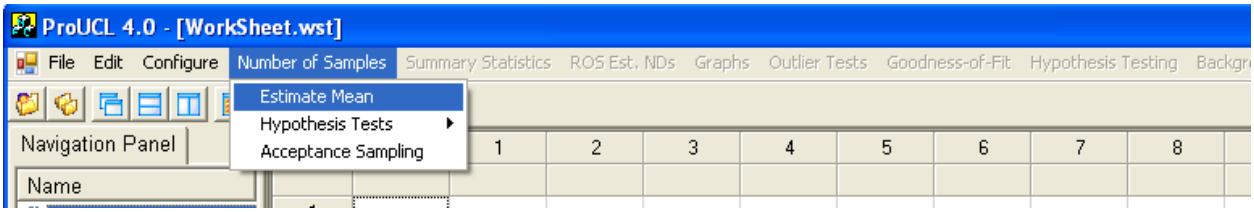
Sample Sizes for User Specified Performance Measures



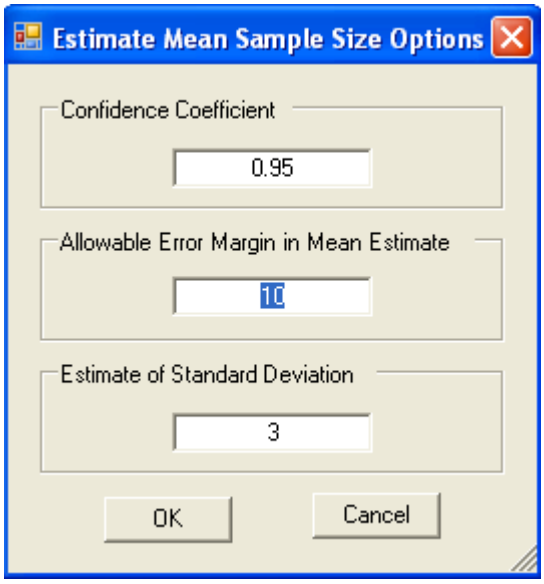
One of the most frequent problems in the application of statistical theory to practical applications including environmental projects is the determination of appropriate sample sizes for sampling of reference/background areas and survey units (e.g., potentially impacted site areas, remediation units, areas of concern, decision units) needed to make defensible and cost effective decisions based upon those sampled data. The sample size determination formulae for estimation of the population mean (or some other parameters) depend upon certain decision parameters including the confidence coefficient, (1-α) and the specified error margin, Δ from the unknown true population mean, μ. Similarly, for hypotheses testing approaches, sample size determination formulae depend upon pre-specified values of the decision parameters selected while defining and describing the data quality objectives (DQOs) associated with an environmental project. The decision parameters (DQOs) associated with hypotheses testing approaches include Type I (false positive error rate, α) and Type II (false negative error rate, β=1-power) error rates; and the allowable width, Δ of the gray region. For values of the parameter of interest (e.g., mean, proportion) lying in the gray region, the consequences of committing the two types of errors described above are not significant from both human health and cost effectiveness point of view. Both parametric (assuming normality) and nonparametric (distribution free) sample size determination formulae as described in EPA guidance documents (EPA 2000, 2002c, 2006, 2009) have been incorporated in ProUCL 4.00.05. The details can be found in "Supplement to ProUCL 4.0 Technical Guide – Determining Minimum Sample Sizes for User Specified Performance Measures."

12.1 Estimation of Mean

1. Click **Number of Samples ► Estimate Mean**



2. The following options window is shown.



- Specify the **Confidence Coefficient**. Default is **0.95**.
- Specify the **Estimate of standard deviation**. Default is **3**
- Specify the **Allowable Error Margin in Mean Estimate**. Default is **10**.
- Click on OK button to continue or on Cancel button to cancel the options.

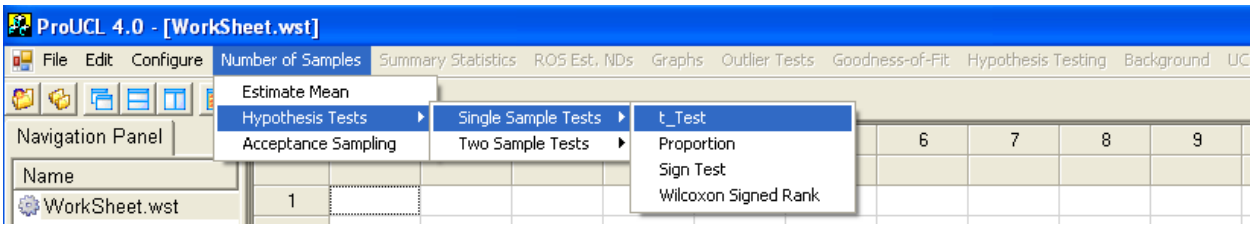
Output Screen for Sample sizes for estimation of mean (CC = 95%, sd = 25, error margin = 10)

	Sample Size for Estimation of Mean
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:12:37 PM
User Selected Options	
Confidence Coefficient	95%
Allowable Error Margin	10
Estimate of Standard Deviation	25
	Approximate Minimum Sample Size
95% Confidence Coefficient:	26

12.2 Sample Sizes for Single Sample Hypothesis Tests

12.2.1 Sample Size for Single Sample t-Test

1. Click **Number of Samples ► Hypothesis Tests ► Single Sample Tests ► t-Test**



2. The following options window is shown.

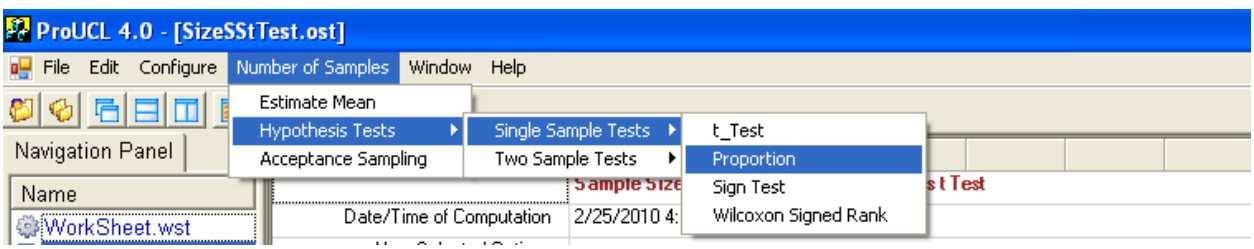
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on OK button to continue or on Cancel button to cancel the options.

Output Screen for sample sizes for single sample t-test (α = 0.05, β = 0.2, s = 10.41, Δ = 10 – Example from EPA 2006 (page 49))

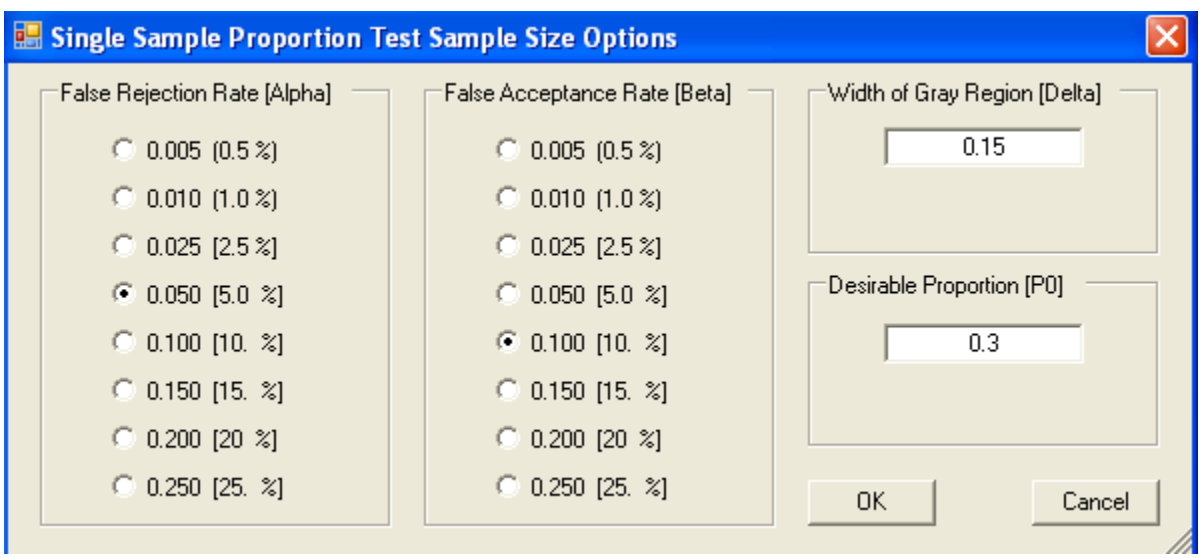
	Sample Sizes for Single Sample t Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:41:58 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.2
Width of Gray Region [Delta]	10
Estimate of Standard Deviation	10.41
	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	9
Two Sided Alternative Hypothesis:	11

12.2.2 Sample Size for Single Sample Proportion Test

1. Click **Number of Samples ► Hypothesis Tests ► Single Sample Tests ► Proportion**



2. The following options window is shown.



- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.

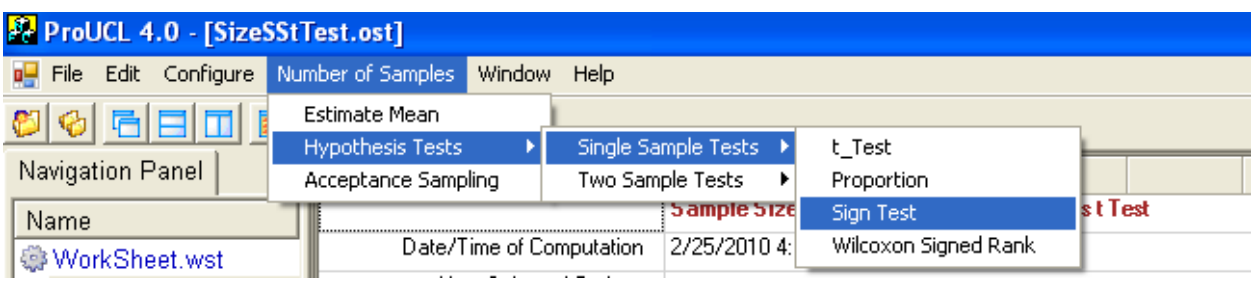
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Desirable Proportion (P0)**. Default is **0.3.**
- Specify the **Width of the Gray Region (Delta)**. Default is **0.15**.
- Click on OK button to continue or on Cancel button to cancel the options.

Output Screen for sample sizes for single sample proportion test (α = 0.05, β = 0.2, P0 = 0.2, Δ = 0.05 - Example from EPA 2006 (page 59))

	Sample Sizes for Single Sample Proportion Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:50:52 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.2
Width of Gray Region [Delta]	0.05
Proportion/Action Level [P0]	0.2
	Approximate Minimum Sample Size
Right Sided Alternative Hypothesis:	419
Left Sided Alternative Hypothesis:	368
Two Sided Alternative Hypothesis:	max(471, 528)

12.2.3 Sample Size for Single Sample Sign Test

1. Click **Number of Samples ► Hypothesis Tests ► Single Sample Tests ► Sign Test**



2. The following options window is shown.

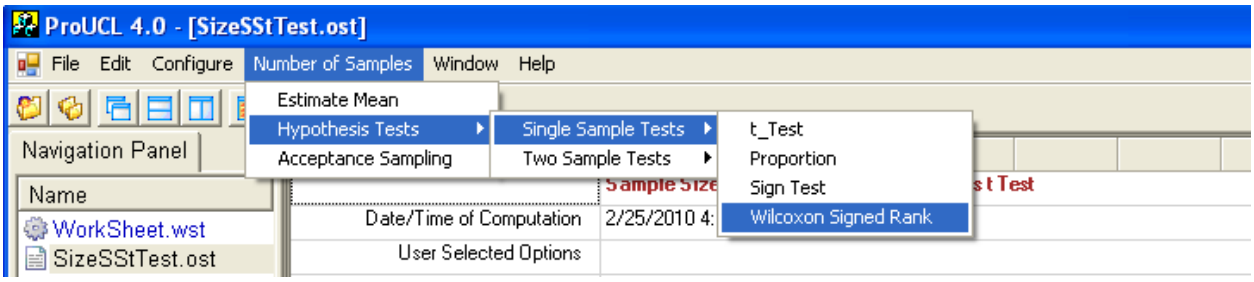
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on OK button to continue or on Cancel button to cancel the options.

Output Screen for sample sizes for single sample sign test (default options)

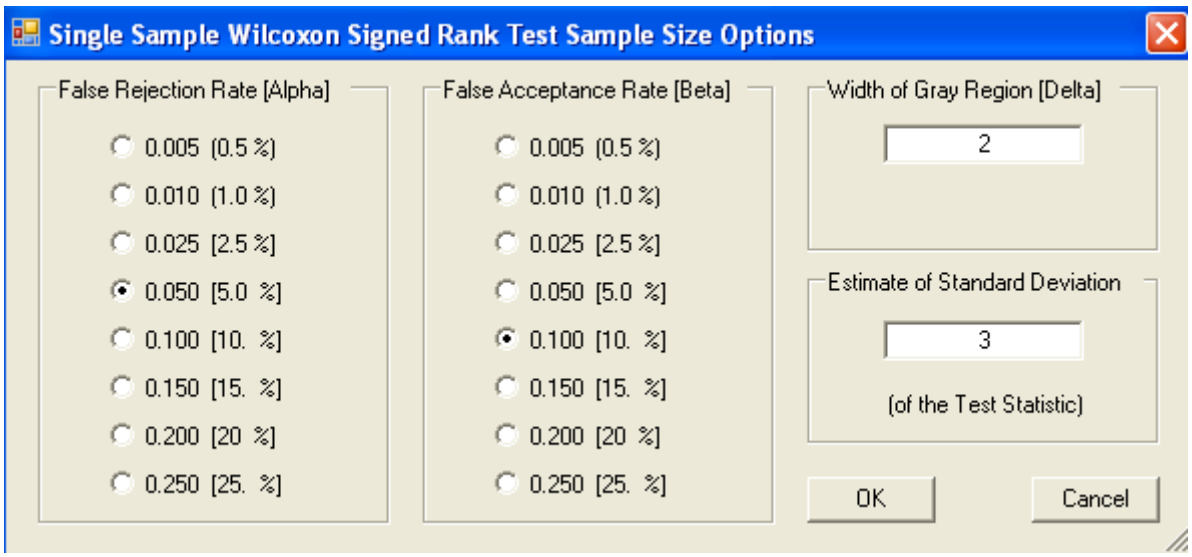
	Sample Sizes for Single Sample Sign Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:15:27 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.1
Width of Gray Region [Delta]	2
Estimate of Standard Deviation	3
	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	35
Two Sided Alternative Hypothesis:	43

12.2.4 Sample Size for Single Sample Wilcoxon Signed Rank Test

1. Click **Number of Samples ► Hypothesis Tests ► Single Sample Tests ► Wilcoxon Signed Rank**



2. The following options window is shown.



- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation of WSR Test Statistic**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on OK button to continue or on Cancel button to cancel the options.

Output Screen for sample sizes for single sample WSR test (α = 0.1, β = 0.2, sd = 130, Δ = 100 - Example from EPA 2006 (page 65))

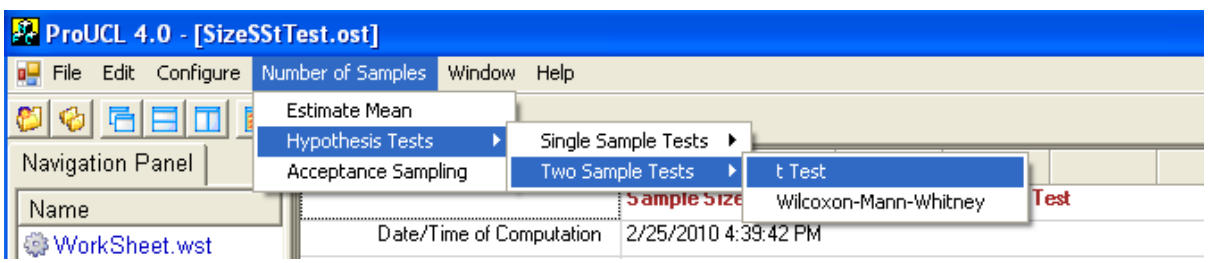
	Sample Sizes for Single Sample Wilcoxon Signed Rank Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 1:13:58 PM
User Selected Options	
False Rejection Rate [Alpha]	0.1
False Acceptance Rate [Beta]	0.2
Width of Gray Region [Delta]	100
Estimate of Standard Deviation	130

	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	10
Two Sided Alternative Hypothesis:	14

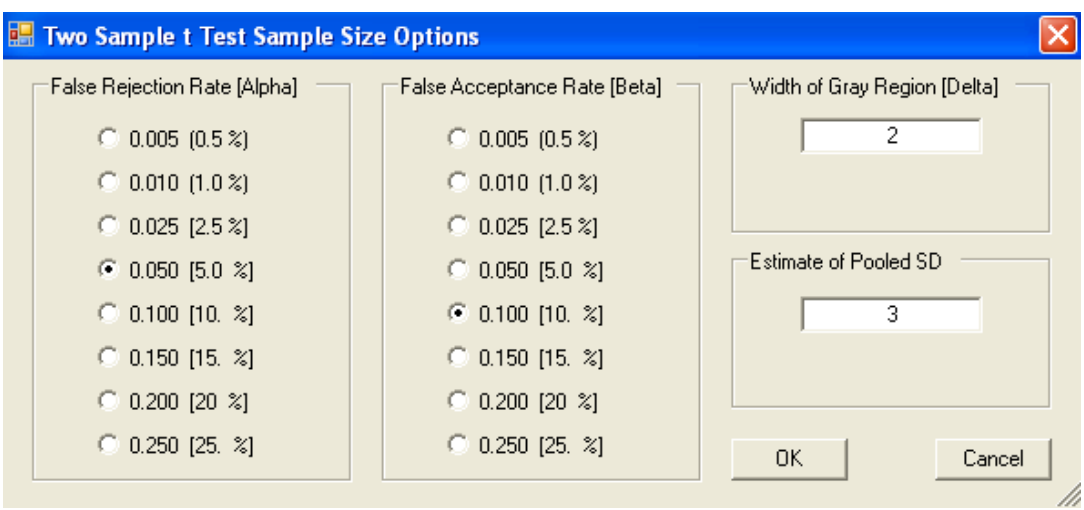
12.3 Sample Sizes for Two Sample Hypothesis Tests

12.3.1 Sample Size for Two Sample t-Test

1. Click **Number of Samples ► Hypothesis Tests ► Two Sample Tests ► t-Test**



2. The following options window is shown.



- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.

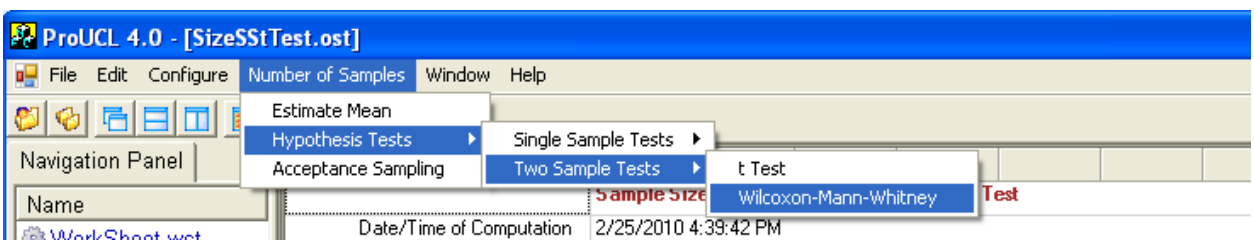
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on OK button to continue or on Cancel button to cancel the options.

Output Screen for sample sizes for two sample t-test (α = 0.05, β = 0.2, s_p = 1.467, Δ = 2.5 - Example from EPA 2006 (page 68))

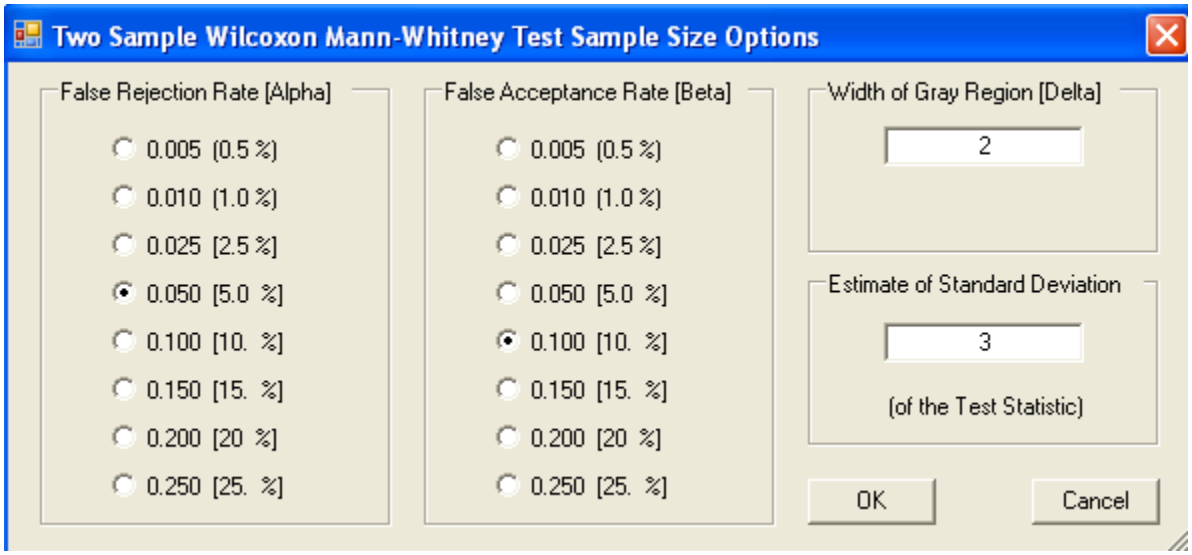
	Sample Sizes for Two Sample t Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 1:17:57 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.2
Width of Gray Region [Delta]	2.5
Estimate of Pooled SD	1.467
	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	5
Two Sided Alternative Hypothesis:	7

12.3.2 Sample Size for Two Sample Wilcoxon Mann-Whitney Test

1. Click **Number of Samples ► Hypothesis Tests ► Two Sample Tests ► Wilcoxon Mann-Whitney Test**



2. The following options window is shown.



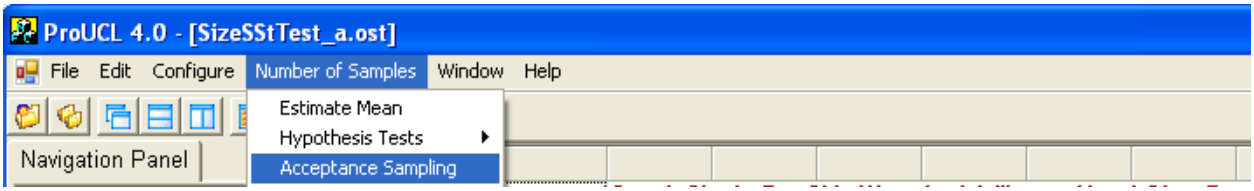
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation of WMW Test Statistic**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on OK button to continue or on Cancel button to cancel the options.

Output Screen for sample sizes for single sample WMW test (default options)

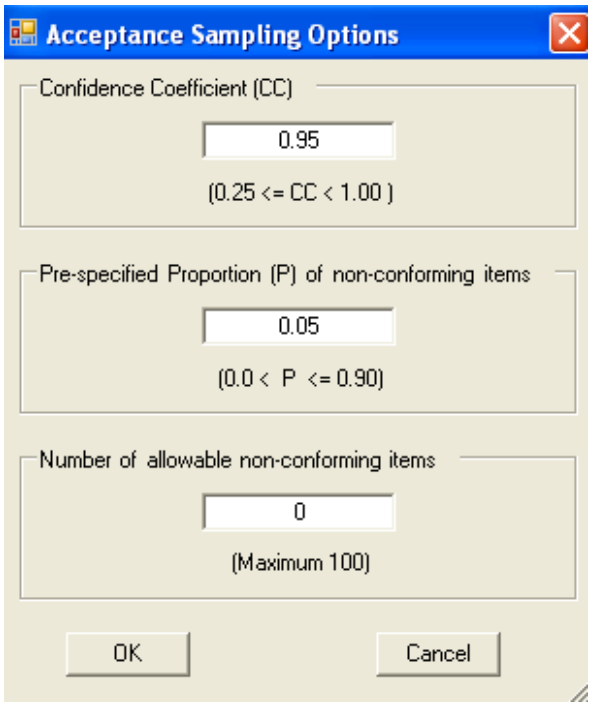
	Sample Sizes for Two Sample Wilcoxon-Mann-Whitney Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:18:47 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.1
Width of Gray Region [Delta]	2
Estimate of Standard Deviation	3
	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	46
Two Sided Alternative Hypothesis:	56

12.4 Sample Sizes for Acceptance Sampling

1. Click **Number of Samples ► Acceptance Sampling**



2. The following options window is shown.



- Specify the **Confidence Coefficient.** Default is **0.95**.
- Specify the **Pre-specified proportion of non-conforming items**. Default is **0.05**.
- Specify the **Number of allowable non-conforming items**. Default is **0.**
- Click on OK button to continue or on Cancel button to cancel the options.

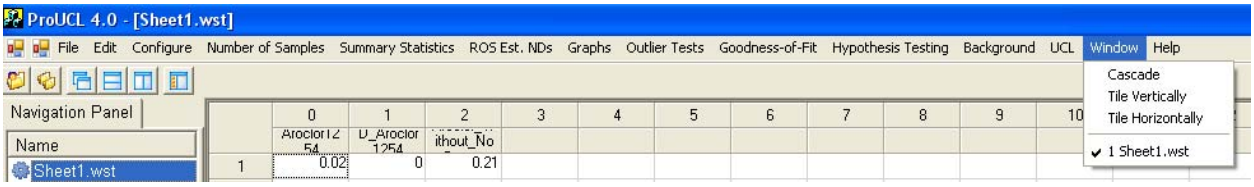
Output Screen for sample sizes for acceptance sampling (default options)

	Acceptance Sampling for Pre-specified Proportion of Non-conforming Items
	Based on Specified Values of Decision Parameters/DQOs
Date/Time of Computation	2/26/2010 12:20:34 PM
User Selected Options	
Confidence Coefficient	0.95
Pre-specified proportion of non-conforming items in the lot	0.05
Number of allowable non-conforming items in the lot	0

	Approximate Minimum Sample Size				
Exact Binomial/Beta Distribution	59				
Approximate Chisquare Distribution (Tukey-Scheffe)	59				

Chapter 13

Windows



Click on the Window menu to reveal the drop-down options shown above.

The following Window drop-down menu options are available:

- Cascade option: arranges windows in a cascade format. This is similar to a typical Windows program option.

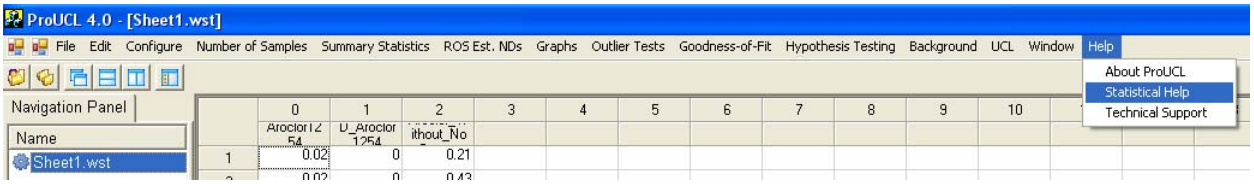
- Tile option: resizes each window vertically or horizontally and then displays all open windows. This is similar to a typical Windows program option.

- The drop-down options list also includes a list of all open windows with a check mark in front of the active window. Click on any of the windows listed to make that window active. This is especially useful if you have more than 20 windows open, as the navigation panel only holds the first 20 windows.

Chapter 14

Help

When the Help menu is clicked, the following window will appear.



Three options are available under Help menu:

- About ProUCL: This option provides a brief description of ProUCL 4.0, and all improvements made compared to ProUCL 3.0.

- Statistical Help: This option executes an online help directory. This option provides information about the various algorithms and formulae (with references) used in the development of ProUCL 4.0. More information on the various topics covered under Statistical Help is provided below.

- Technical Support: This option will provide contact information for primary technical support via phone and e-mail.

Statistical Help provides online help notes for the methods and options available in ProUCL 4.0. A screen (shown below), listing the topics containing help notes, appears after clicking on the Help menu.

The online help directory is divided into the following submenus:

- Getting Started: This chapter provides basic information on the software, including software installation to various menu displays.

- Summary Statistics: This chapter provides information and examples on procedures for simple classical summary statistics for data sets with and without nondetects.

- ROS Estimates of NDs: This chapter briefly describes the estimation (extrapolation) of nondetects using regression on order statistics (ROS) for normal, lognormal, and gamma distribution.

- Graphs: This chapter provides information and examples on the graphical displays that ProUCL 4.0 can produce: box plots, histograms, and Multi Q-Q plots.

- Outlier Tests: This chapter provides information and examples for two classical outlier tests available in ProUCL 4.0: Dixon's and Rosner's tests for data sets with and without outliers.

- Goodness-of-Fit: This chapter provides information and examples for several goodness-of-fit tests available in ProUCL 4.0 for data sets with and without NDs.

- Background Statistics: This chapter provides information and examples for the computation of Background Statistics needed to estimate the BTVs and not-to-exceed values. These statistics are sometimes used to compare point-by-point site data (not more than 4 to 5 site samples) with the BTVs.

- Hypotheses Testing: This chapter provides brief descriptions (with examples) of the various single sample and two-sample hypotheses testing approaches as incorporated in ProUCL 4.0.

    - Single Sample Hypotheses Testing: This chapter provides brief description and examples of single sample hypotheses testing approaches that are useful to compare site concentrations with cleanup standards, compliance limits, or not-to-exceed limits. The minimum sample size requirements for site data are briefly discussed in Chapter 1 of this User Guide.
    - Two-sample hypotheses are used to compare site data with background data (or upgradient and downgradient wells) provided enough site and background data are available. The minimum sample size requirements for site and background data are briefly discussed in Chapter 1 of this User Guide.

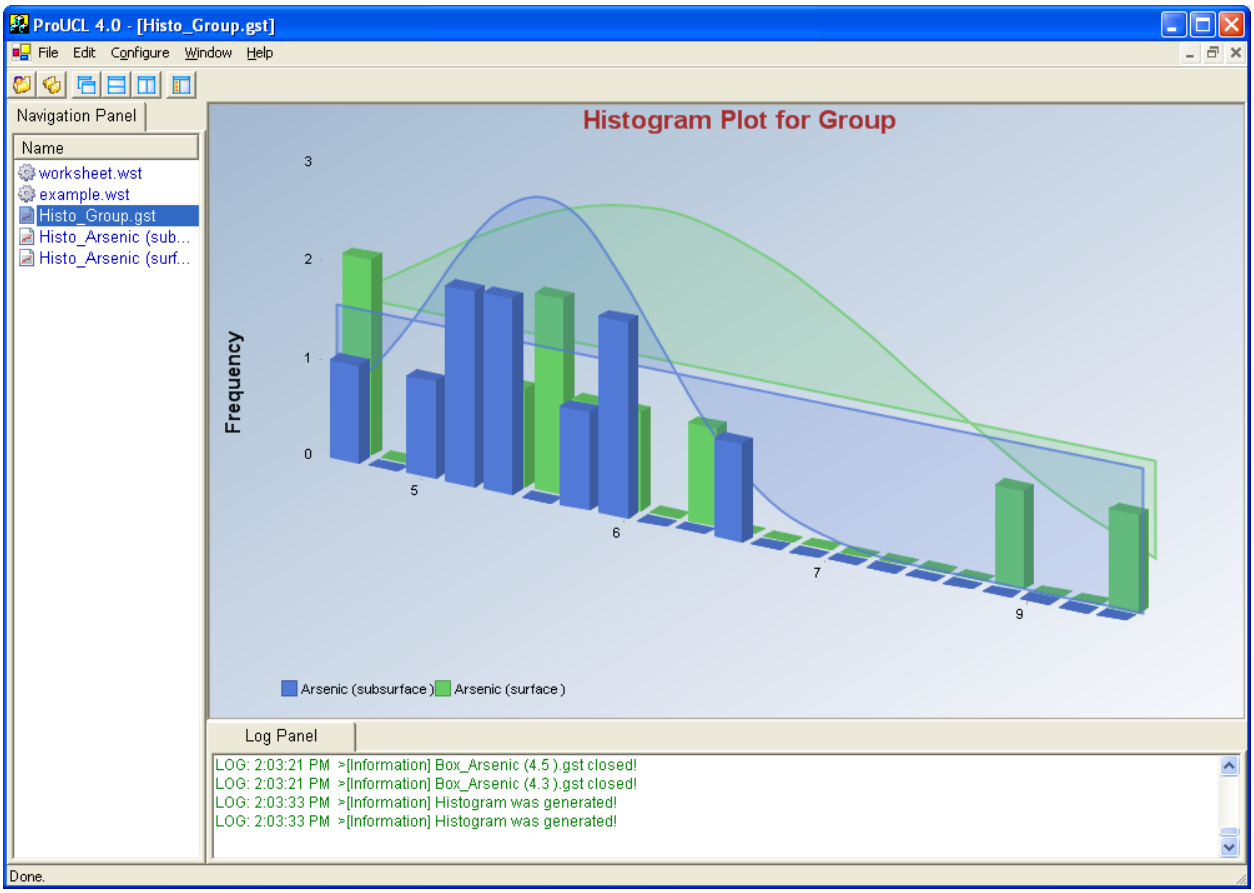
- Upper Confidence Limits (UCLs): This option provides information and examples for the various UCL computation methods as incorporated in ProUCL 4.0.

Chapter 15

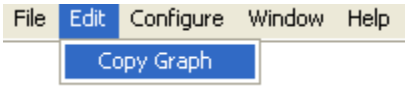
Handling the Output Screens and Graphs

Copying Graphs

1. Click the graph you want to copy or save in the **Navigation Panel**. The graphs can be saved using the copy option.



2. Click **Edit ► Copy Graph**.



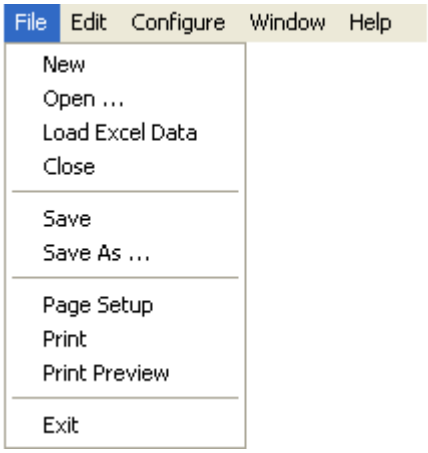
3. Once the user has clicked "Copy Graph," the graph is ready to be imported (pasted) into most Microsoft Office applications (Word, Excel, and PowerPoint have been tested) by clicking **Edit ► Paste** in these Microsoft applications.

4. It is important to note that the graph cannot be saved as its own file and must be imported into an application to be saved. This will save the graph, but the overall file attribute and properties will be that of the application in which the graph was saved in. For example, if the graph was saved within Microsoft Word, the graph will be saved in a document with a .doc extension.

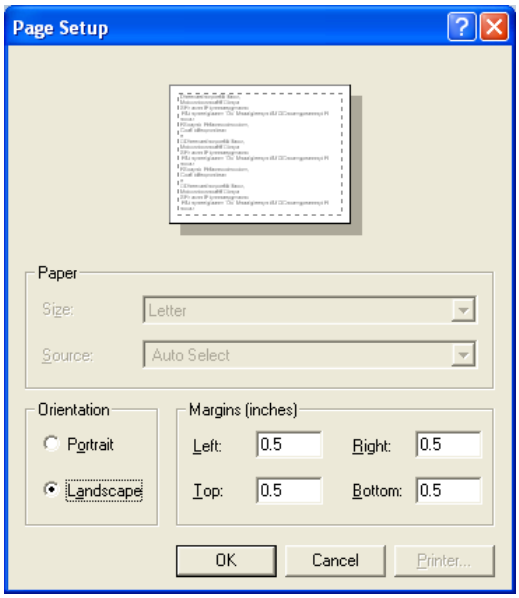
Printing Graphs

1. Click the graph you want to print in the **Navigation Panel**.

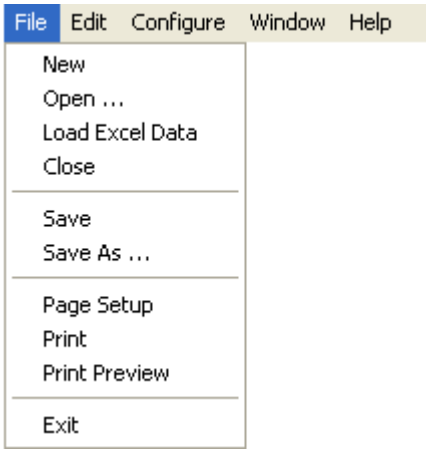
2. Click **File ► Page Setup**.



3. Check the radio button next to **Portrait** or **Landscape**, and click **OK**. In some cases, with larger headings and captions, it may be desirable to use the Landscape printing option.

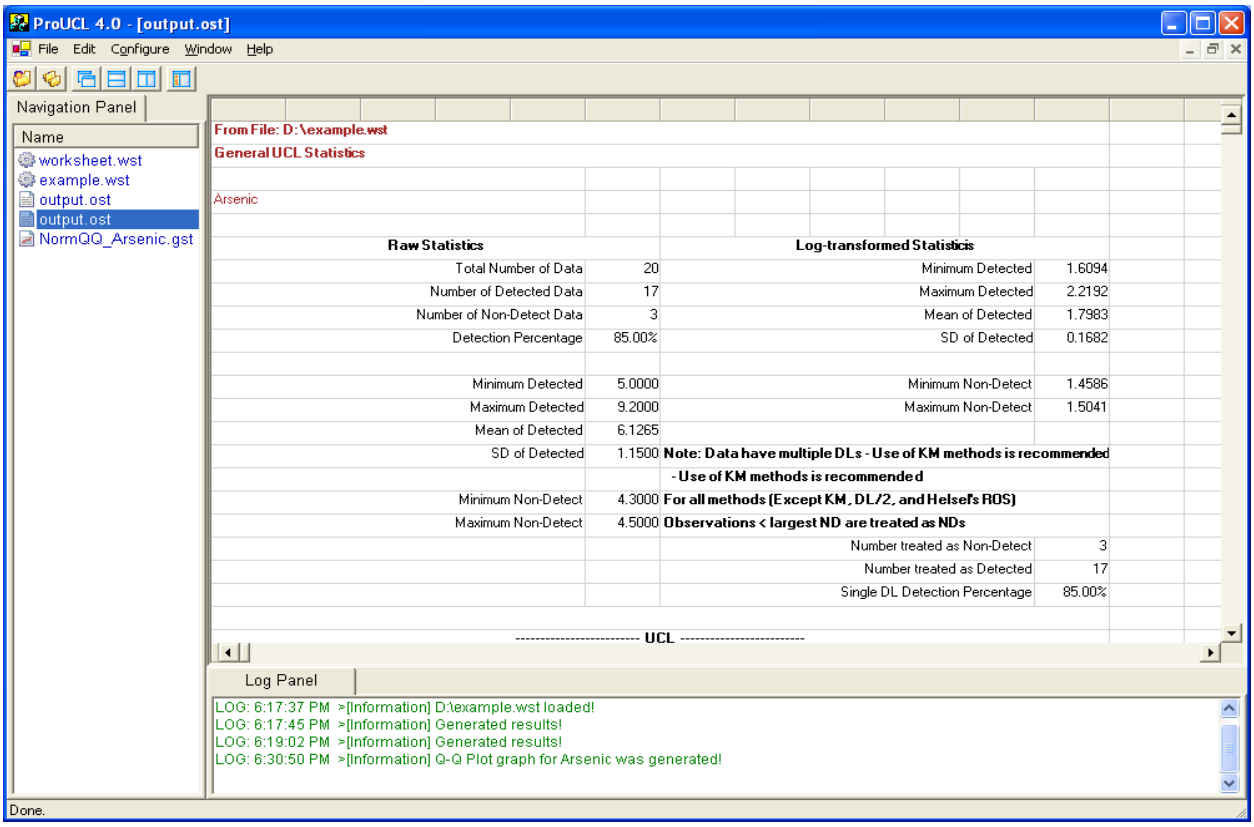


4. Click **File ► Print** to print the graph, and **File ► Print Preview** to preview the graph before printing.

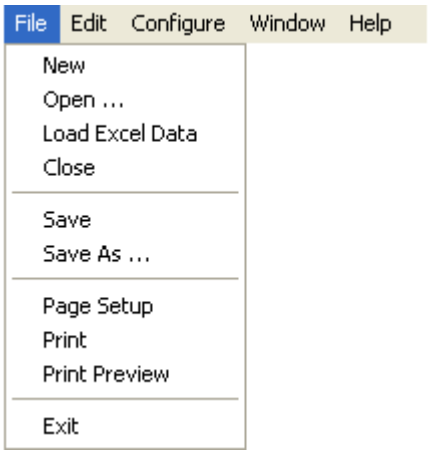


Printing Non-graphical Outputs

1. Click the output you want to copy or print in the **Navigation Panel**.



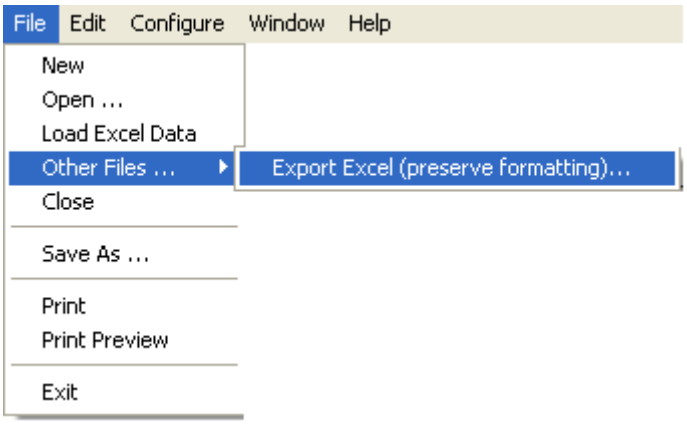
2. Click **File ► Print**.



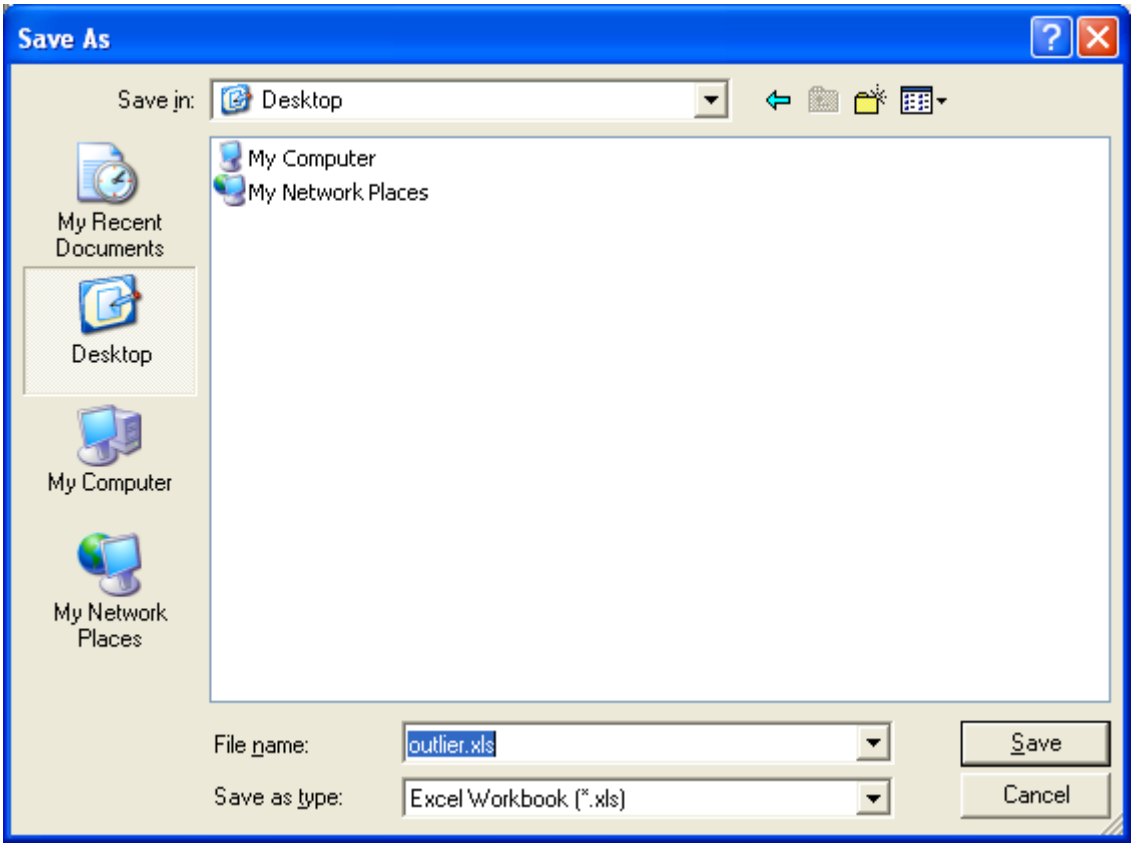
Saving Output Screens as Excel Files

1. Click on the output you want to save in the **Navigation Panel List**.

2. Click **File ► Other Files… ►**Export **Excel (preserve formatting).**



3. Enter the desired file name you want to use, and click **Save**, and save the file in the desired folder using your browser.



Chapter 16

Recommendations to Compute a 95% UCL (Estimate of EPC Term) of the Population Mean, μ_1, Using Symmetric and Positively Skewed Full Data Sets without any Nondetects

This chapter describes the recommendations for the computation of a 95% *UCL* of the unknown population arithmetic mean, μ_1, of a contaminant data distribution based upon full data sets without any nondetect observations. These recommendations are based upon the findings of Singh, Singh, and Engelhardt (1997, 1999); Singh *et al*. (2002a); Singh, Singh, and Iaci (2002b); and Singh and Singh (2003). These recommendations are applicable to full data sets without censoring and nondetect (ND) observations. Recommendations to compute *UCL95* based upon data sets with NDs are summarized in Chapter 16.

For skewed parametric as well as nonparametric data sets, there is no simple solution to compute a *95% UCL* of the population mean, μ_1. Contrary to the general conjecture, Singh *et al.* (2002a); Singh, Singh, and Iaci (2002b); and Singh and Singh (2003) noted that the *UCLs* based upon the skewness adjusted methods, such as the Johnson's modified t and Chen's adjusted-*CLT* do not provide the specified coverage (e.g., 95 %) to the population mean even for mildly to moderately skewed (e.g., $\hat{\sigma}$ the *sd* of log-transformed data in interval [0.5, 1.0)) data sets for samples of size as large as 100. The coverage of the population mean by the skewness-adjusted *UCLs* becomes much smaller than the specified coverage of 0.95 for highly skewed data sets, where skewness is defined as a function of σ or $\hat{\sigma}$ (*sd* of log-transformed data).

It is noted that even though, the simulation results for highly skewed data sets of small sizes suggest that the bootstrap t and Hall's bootstrap methods do approximately provide the adequate coverage to the population mean, sometimes in practice these two bootstrap methods yield erratic inflated values (orders of magnitude higher than the other *UCL* values) when dealing with individual highly skewed data sets of small sizes. This is especially true when potential outliers may be present in the data set. ProUCL 4.0 provides warning messages whenever the recommendations are made regarding the use the bootstrap t method or Hall's bootstrap method.

16.1 Normally or Approximately Normally Distributed Data Sets

- For normally distributed data sets, a *UCL* based upon the Student's t-statistic provides the optimal *UCL* of the population mean. Therefore, for normally distributed data sets, one should always use a 95% *UCL* based upon the Student's t-statistic.

- The *95% UCL* of the mean based upon Student's t-statistic may also be used when the *Sd,* s_y of the log-transformed data, is less than 0.5, or when the data set approximately follows a normal distribution. Typically, a data set is approximately normal when the normal Q-Q plot displays a linear pattern (without outliers and jumps of significant magnitude) and the resulting correlation coefficient is high (e.g., 0.95 or higher). The jumps and breaks in a Q-Q plot (even with a high correlation coefficient) suggest the presence of multiple populations in the data set under study.

- Student's t-*UCL* may also be used when the data set is symmetric (but possibly not normally distributed). A measure of symmetry (or skewness) is $\hat{k}_3$. A value of $\hat{k}_3$ close to zero (e.g., if the absolute value of the skewness is roughly less than 0.2 or 0.3) suggests approximate symmetry. The approximate symmetry of a data distribution can also be judged by looking at the histogram of the data set.

16.2 Gamma Distributed Skewed Data Sets

In practice, many skewed data sets can be modeled both by a lognormal distribution and a gamma distribution, especially when the sample size is smaller than 70-100. As well known, the 95% *H-UCL* of the mean based upon a lognormal model often behaves in an erratic manner. Specifically, 95% *H-UCL* often results in an unjustifiably large and impractical 95% *UCL* value when the sample size is small (e.g., n <20, 50, ..) and skewness is high. Moreover, it is also observed that a 95% UCL based upon Land H-statistic becomes even smaller than the sample arithmetic mean. This is especially true for mildly skewed to moderately skewed data sets of large sizes (e.g., > 50-70). In such cases, a gamma model, G(k, θ), may be used to compute a reliable 95% *UCL* of the unknown population mean, μ_1.

- Many skewed data sets follow a lognormal as well as a gamma distribution. It should be noted that the population means based upon the two models could differ significantly. A lognormal model based upon a highly skewed (e.g., $\hat{\sigma} \geq 2.5$) data set will have an unjustifiably large and impractical population mean, μ_1, and its associated *UCL*. The gamma distribution is better suited to model positively skewed environmental data sets.

- One should always first check if a given skewed data set follows a gamma distribution. If a data set does follow a gamma distribution or an approximate gamma distribution, one should compute a 95% UCL based upon a gamma distribution. Use of highly skewed (e.g., $\hat{\sigma} \geq 2.5$-3.0) lognormal distributions should be avoided. For such highly skewed lognormally distributed data sets that cannot be modeled by a gamma or an approximate gamma distribution, nonparametric *UCL* computation methods based upon the Chebyshev inequality may be used.

- The five bootstrap methods do not perform better than the two gamma *UCL* computation methods. It is noted that the performances (in terms of coverage probabilities) of the bootstrap t and Hall's bootstrap methods are very similar. Out of the five bootstrap methods, bootstrap t and Hall's bootstrap methods perform the best (with coverage probabilities for population mean closer to the nominal level of 0.95). This is especially true when the skewness is quite high (e.g., $\hat{k} < 0.1$) and the sample size is small (e.g., $n <$ 10-15). Whenever the use of Hall's *UCL* or bootstrap t *UCL* is recommended, an informative warning message about their use is also provided.

- Contrary to the conjecture, the bootstrap BCA method does not perform better than the Hall's method or the bootstrap t method. The coverage for the population mean, μ_1, provided by the BCA method is much lower than the specified 95% coverage. This is especially true when the skewness is high (e.g., $\hat{k} < 1$) and the sample size is small (e.g., Singh and Singh (2003), and Singh, Singh, and Iaci (2002b)).

- From the results presented in Singh, Singh, and Iaci (2002b), and in Singh and Singh (2003), it is concluded that for data sets which follow a gamma distribution, a 95% *UCL* of the mean should be computed using the adjusted gamma *UCL* when the shape parameter, k, is: $0.1 \leq k < 0.5$, and for values of $k \geq 0.5$, a 95% *UCL* can be computed using an approximate gamma *UCL* of the mean, μ_1.

- For highly skewed gamma distributed data sets with $k < 0.1$, the bootstrap t *UCL* or Hall's bootstrap (Singh and Singh 2003) may be used when the sample size is smaller than 15, and the adjusted gamma *UCL* should be used when the sample size starts approaching or exceeding 15. The small sample size requirement increases as the skewness increases (that is, as k decreases, the required sample size, n, increases).

- The bootstrap t and Hall's bootstrap methods should be used with caution as these methods may yield erratic, unreasonably inflated, and unstable *UCL* values, especially in the presence of outliers. In case Hall's bootstrap and bootstrap t methods yield inflated and erratic *UCL* results, the 95% *UCL* of the mean should be computed based upon the adjusted gamma 95% *UCL*. ProUCL 4.0 prints out a warning message associated with the recommended use of the *UCLs* based upon the bootstrap t method or Hall's bootstrap method. The recommendations for gamma distribution are summarized in Table 15-1.

Table 16-1. Computation of a UCL95 of the Unknown Mean, μ_1, of a Gamma Distribution

$\hat{k}$	Sample Size, n	Recommendation
$\hat{k} \geq 0.5$	For all n	Approximate gamma 95% *UCL*
$0.1 \leq \hat{k} < 0.5$	For all n	Adjusted gamma 95% *UCL*
$\hat{k} < 0.1$	$n < 15$	95% *UCL* based upon bootstrap t or Hall's bootstrap method*
$\hat{k} < 0.1$	$n \geq 15$	Adjusted gamma 95% *UCL* if available, otherwise use approximate gamma 95% *UCL*

*If bootstrap t or Hall's bootstrap methods yield erratic, inflated, and unstable *UCL* values (which often happens when outliers are present), the *UCL* of the mean should be computed using the adjusted gamma *UCL*.

***Note**: Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL. These recommendations are based upon the results of the simulation studies summarized in Singh, Singh, and Iaci (2002) and Singh and Singh (2003). For additional insight, the user may want to consult a statistician.*

16.3 Lognormally Distributed Skewed Data Sets

For lognormally, LN (μ, σ^2), distributed data sets, the H-statistic-based *H-UCL* provides specified 0.95, coverage for the population mean for all values of σ. However, the H-statistic often results in unjustifiably large *UCL* values that do not occur in practice. This is especially true when the skewness is high (e.g., $\sigma >$

2.0). The use of a lognormal model unjustifiably accommodates large and impractical values of the mean concentration and its *UCLs*. The problem associated with the use of a lognormal distribution is that the population mean, μ_1, of a lognormal model becomes impractically large for larger values of σ, which in turn results in an inflated *H-UCL* of the population mean, μ_1. Since the population mean of a lognormal model becomes too large, none of the other methods, except for the *H-UCL,* provides the specified 95% coverage for that inflated population mean, μ_1. This is especially true when the sample size is small and the skewness is high. For extremely highly skewed data sets (with $\sigma > 2.5$-3.0) of smaller sizes (e.g., < 70-100), the use of a lognormal distribution-based *H-UCL* should be avoided (e.g., see Singh *et al*. (2002a) and Singh and Singh (2003)). Therefore, alternative *UCL* computation methods, such as the use of a gamma distribution or the use of a *UCL* based upon nonparametric bootstrap methods or the Chebyshev inequality-based methods, are desirable. All skewed data sets should first be tested for a gamma distribution. For lognormally distributed data (that cannot be modeled by gamma distribution), methods summarized in Table 15-2 may be used to compute a 95% *UCL* of mean.

ProUCL can compute an *H-UCL* for samples of sizes up to 1000. For highly skewed lognormally distributed data sets of smaller sizes, alternative methods to compute a 95% *UCL* of the population mean, μ_1, are summarized in Table 15-2. Since skewness is a function of σ (or $\hat{\sigma}$), the recommendations for the computation of the *UCL* of the population mean are also summarized in terms of $\hat{\sigma}$ and the sample size, n. Here, $\hat{\sigma}$ is an *MLE* of σ, and is given by the *Sd* of log-transformed data. Note that Table 15-2 is applicable only to the computation of a 95% *UCL* of the population mean based upon lognormally distributed data sets without nondetect observations.

Due to unstable and erratic behavior of H-UCL, the users are discouraged to use lognormal distribution to estimate the EPC terms. ProUCL computes and outputs H-statistic based UCLs for historical reasons only. H-statistic often results in unstable (both high and low) values of UCL95 as shown in examples 1 and 2 below. It is therefore recommended to avoid the use of H-statistic based 95% UCLs. Use of nonparametric methods is preferred to compute UCL95 for skewed data sets which do not follow a gamma distribution.

Table 16-2. Computation of a 95% UCL of Mean, μ_1 of a Lognormal Population

$\hat{\sigma}$	Sample Size, *n*	Recommendation
$\hat{\sigma} < 0.5$	For all *n*	Student's t, modified t, or *H-UCL*
$0.5 \le \hat{\sigma} < 1.0$	For all *n*	H-UCL
$1.0 \le \hat{\sigma} < 1.5$	$n < 25$	95% Chebyshev (Mean, Sd) *UCL*
	$n \ge 25$	H-UCL
$1.5 \le \hat{\sigma} < 2.0$	$n < 20$	99% Chebyshev (Mean, Sd) *UCL*
	$20 \le n < 50$	95% Chebyshev (Mean, Sd) *UCL*
	$n \ge 50$	H-UCL
$1.5 \le \hat{\sigma} < 2.0$	$n < 20$	99% Chebyshev (Mean, Sd) *UCL*
	$20 \le n < 50$	97.5% Chebyshev (Mean, Sd) *UCL*
	$50 \le n < 70$	95% Chebyshev (Mean, Sd) *UCL*
	$n \ge 70$	H-UCL
$2.5 \le \hat{\sigma} < 3.0^{**}$	$n < 30$	99% Chebyshev (Mean, Sd)
	$30 \le n < 70$	97.5% Chebyshev (Mean, Sd) *UCL*

	$70 \le n < 100$	95% Chebyshev (Mean, Sd) *UCL*
	$n \ge 100$	H-UCL
$3.0 \le \hat{\sigma} \le 3.5^{**}$	$n < 15$	Hall's bootstrap method*
	$15 \le n < 50$	99% Chebyshev (Mean, Sd)
	$50 \le n < 100$	97.5% Chebyshev (Mean, Sd) *UCL*
	$100 \le n < 150$	95% Chebyshev (Mean, Sd) *UCL*
	$n \ge 150$	H-UCL
$\hat{\sigma} > 3.5^{**}$	For all n	Use nonparametric methods*

*If Hall's bootstrap method yields an erratic or unrealistically large UCL value, then the UCL of the mean may be computed based upon the Chebyshev inequality.

*** For highly skewed data sets with $\hat{\sigma}$ exceeding 3.0, 3.5, it is suggested to pre-process the data. It is very likely that the data consist of outliers and/or come from multiple populations. The population partitioning methods may be used to identify mixture populations present in the data set. For defensible conclusions, the decision statistics such as UCL95 should be computed separately for each of the identified sub-population.*

***Note**: Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL. These recommendations are based upon the results of the simulation studies summarized in Singh, Singh, and Iaci (2002) and Singh and Singh (2003). For additional insight, the user may want to consult a statistician.*

Based upon the results of the research conducted to evaluate the appropriateness of applicability of lognormal distribution based estimates of EPC term (Singh, et. al., 1997, 1999, 2002, 2003), the developers of ProUCL 4.0 had suggested to avoid the use of lognormal distribution to estimate the EPC term. However, computations of various lognormal distribution based statistics were included in ProUCL 4.00.05 for historical reasons and for the sake of comparison for interested practitioners. For Lognormally distributed data set, the developers again suggest and recommend the use of nonparametric distribution free methods (Efron, 1981, 1982 and Efron and Tibshirani, 1993) to estimate environmental parameters of interest (e.g., EPC, BTV), especially when skewness is high, that is when *Sd* of log-transformed data starts exceeding 1.5. As a result, the developers have revised recommendations and decision tables given in ProUCL Technical Guide and User Guide. Those revisions have been incorporated in this updated ProUCL 4.00.05 version.

16.4 Data Sets without a Discernable Skewed Distribution – Nonparametric Methods for Skewed Data Sets

The use of gamma and lognormal distributions as discussed here cover a wide range of skewed data distributions. For skewed data sets which are neither gamma nor lognormal, one can use a nonparametric Chebyshev *UCL* or Hall's bootstrap *UCL* (for small samples) of the mean to estimate the EPC term.

- For skewed nonparametric data sets with negative and zero values, use a *95% Chebyshev (Mean, Sd) UCL* for the population mean, μ_1.

For all other nonparametric data sets with only positive values, the following method may be used to estimate the EPC term.

- For mildly skewed data sets with $\hat{\sigma} \leq 0.5$, one can use the Student's t-statistic or modified t-statistic to compute a 95% *UCL* of mean, μ_1.

- For nonparametric moderately skewed data sets (e.g., σ or its estimate, $\hat{\sigma}$, in the interval (0.5, 1]), one may use a *95% Chebyshev (Mean, Sd) UCL* of the population mean, μ_1.

- For nonparametric moderately to highly skewed data sets (e.g., $\hat{\sigma}$ in the interval (1.0, 2.0]), one may use a *99% Chebyshev (Mean, Sd) UCL* or a *97.5% Chebyshev (Mean, Sd) UCL* of the population mean, μ_1, to obtain an estimate of the EPC term.

- For highly skewed to extremely highly skewed data sets with $\hat{\sigma}$ in the interval (2.0, 3.0], one may use Hall's *UCL* or a *99% Chebyshev (Mean, Sd) UCL* to compute the EPC term.

- Extremely skewed nonparametric data sets with σ exceeding 3 provide poor coverage. For such highly skewed data distributions, none of the methods considered provide the specified 95% coverage for the population mean, μ_1. The coverage provided by the methods decrease as σ increases. For such data sets of sizes less than 30, a 95% *UCL* can be computed based upon Hall's bootstrap method or bootstrap t method. Hall's bootstrap method provides the highest coverage (but less than 0.95) when the sample size is small. It is noted that the coverage for the population mean provided by Hall's method (and bootstrap t method) does not increase much as the sample size, *n*, increases. However, as the sample size increases, coverage provided by the *99% Chebyshev (Mean, Sd) UCL* method also increases. Therefore, for larger samples, a *UCL* should be computed based upon the *99% Chebyshev (Mean, Sd)* method. This large sample size requirement increases as $\hat{\sigma}$ increases. These recommendations are summarized in Table 15-3.

Table 16-3. Computation of a 95% *UCL* of the Unknown Mean, μ_1, Based Upon a Skewed Data Set (with all Positive Values) without a Discernable Distribution, Where $\hat{\sigma}$ is the *sd* of Log-transformed Data

$\hat{\sigma}$	*Sample Size, n*	*Recommendation*
$\hat{\sigma} < 0.5$	For all *n*	Student's t, modified t, or *H-UCL*
$0.5 \leq \hat{\sigma} < 1.0$	For all *n*	95% Chebyshev *(Mean, Sd) UCL*
$1.0 \leq \hat{\sigma} < 1.5$	For all *n*	95% Chebyshev *(Mean, Sd) UCL*
$1.5 \leq \hat{\sigma} < 2.0$	$n < 20$	99% Chebyshev *(Mean, Sd) UCL*
	$20 \leq n$	95% Chebyshev *(Mean, Sd) UCL*
$2.0 \leq \hat{\sigma} < 2.5$	$n < 10$	Hall's bootstrap method
	$10 \leq n < 20$	99% Chebyshev *(Mean, Sd) UCL*
	$20 \leq n < 50$	97.5% Chebyshev *(Mean, Sd) UCL*
	$50 \leq n$	95% Chebyshev *(Mean, Sd) UCL*
$2.5 \leq \hat{\sigma} < 3.0$	$n < 10$	Hall's bootstrap method
	$10 \leq n < 30$	99% Chebyshev (Mean, Sd)
	$30 \leq n < 70$	97.5% Chebyshev *(Mean, Sd) UCL*

	$70 \le n$	95% Chebyshev *(Mean, Sd) UCL*
$3.0 \le \hat{\sigma} \le 3.5^{**}$	$n < 15$	Hall's bootstrap method*
	$15 \le n < 50$	99% Chebyshev(Mean, Sd)
	$50 \le n < 100$	97.5% Chebyshev *(Mean, Sd) UCL*
	$100 \le n$	95% Chebyshev *(Mean, Sd) UCL*
$\hat{\sigma} > 3.5^{**}$	For all n	Use nonparametric methods*

*If Hall's bootstrap method yields an erratic and unstable UCL value (e.g., happens when outliers are present), a *UCL* of the population mean may be computed based upon the *99% Chebyshev (Mean, Sd)* method.

*** For highly skewed data sets with $\hat{\sigma}$ exceeding 3.0, 3.5, it is suggested to pre-process the data. It is very likely that the data consist of outliers and/or come from multiple populations. The population partitioning methods may be used to identify mixture populations present in the data set. For defensible conclusions, the decision statistics such as UCL95 should be computed separately for each of the identified sub-population.*

***Note**: Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL. These recommendations are based upon the results of the simulation studies summarized in Singh, Singh, and Iaci (2002) and Singh and Singh (2003). For additional insight, the user may want to consult a statistician.*

Table 16-4. Recommended UCL95 Computation Methods for Full-Uncensored Data Sets without Nondetect Observations

Skewness	Sample Size	95% Student t	95% modified t	95% H-UCL	95% Hall's Bootstrap	95% Chebyshev (MVUE)	97.5% Chebyshev (MVUE)	99% Chebyshev (MVUE)	95% Chebyshev (Mean, Sd)	97.5% Chebyshev (Mean, Sd)	99% Chebyshev (Mean, Sd)	95% Approx. Gamma	95% Bootstrap t	95% Adjusted Gamma
Normal or Approximate Normal (with $\hat{\sigma}$ < 0.5) Distribution														
	All *n*	●	●											
Gamma Distribution														
$\hat{k}$ < 0.1	*n* < 15				●								●	
	n ≥ 15											●		●
0.1 ≤ $\hat{k}$ < 0.5	All *n*													●
$\hat{k}$ ≥ 0.5	All *n*											●		
Lognormal Distribution														
$\hat{\sigma}$ < 0.5	All *n*	●	●	●										
0.5 ≤ $\hat{\sigma}$ < 1.0	All *n*			●										
1.0 ≤ $\hat{\sigma}$ < 1.5	*n* < 25								●					
	n ≥ 25			●										
1.5 ≤ $\hat{\sigma}$ < 2.0	*n* < 20										●			
	20 ≤ *n* < 50								●					
	n ≥ 50			●										
2.0 ≤ $\hat{\sigma}$ < 2.5	*n* < 20										●			
	20 ≤ *n* < 50									●				
	50 ≤ *n* < 70								●					
	n ≥ 70			●										
2.5 ≤ $\hat{\sigma}$ < 3.0	*n* < 30										●			
	30 ≤ *n* < 70									●				
	70 ≤ *n* < 100								●					
	n ≥ 100			●										
3.0 ≤ $\hat{\sigma}$ ≤ 3.5	*n* < 15				●									
	15 ≤ *n* < 50										●			
	50 ≤ *n* < 100									●				
	100 ≤ *n* < 150								●					
	n ≥ 150			●										

Table 16-4. Recommended UCL95 Computation Methods for Full-Uncensored Data Sets without Nondetect Observations-Continued

	Sample Size	95% Student t	95% modified t	95% H-UCL	95% Hall's Bootstrap	95% Chebyshev (MVUE)	97.5% Chebyshev (MVUE)	99% Chebyshev (MVUE)	95% Chebyshev (Mean, Sd)	97.5% Chebyshev (Mean, Sd)	99% Chebyshev (Mean, Sd)	95% Approx. Gamma	95% Bootstrap t	95% Adjusted Gamma
Nonparametric Distribution Free Methods														
$\hat{\sigma} \le 0.5$	All *n*	●	●											
$0.5 < \hat{\sigma} \le 1.0$	All *n*								●					
$1.0 < \hat{\sigma} \le 1.5$	All *n*								●					
$1.0 < \hat{\sigma} \le 2.0$	$n < 20$										●			
	$n \ge 20$								●					
$2.0 < \hat{\sigma} \le 2.5$	$n < 10$				●									
	$10 \le n < 20$										●			
	$20 \le n < 50$									●				
	$50 \le n$								●					
$2.5 < \hat{\sigma} \le 3.0$	$n < 10$				●									
	$10 \le n < 30$										●			
	$30 \le n < 70$									●				
	$70 \le n$								●					
$3.0 < \hat{\sigma} \le 3.5$	$n < 15$				●									
	$15 \le n < 50$										●			
	$50 \le n < 100$									●				
	$100 \le n$								●					
$\hat{\sigma} > 3.5$	$n < 100$				●									
	$n \ge 100$										●			

Note: *Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL. These recommendations are based upon the results of the simulation studies summarized in Singh, Singh, and Iaci (2002) and Singh and Singh (2003). For additional insight, the user may want to consult a statistician.*

16.5 Should the Maximum Observed Concentration be Used as an Estimate of the EPC Term?

This topic has been discussed earlier in Chapter 1. It is included here only for the convenience of the user. In practice, a typical user tends to use the maximum sample value as an estimate of the EPC term. This is especially true when the sample size is small or the data are highly skewed. The discussion and suggestions as described in Chapter 1 apply to both Chapters 15 and 16. Singh and Singh (2003) studied the max test (using the maximum observed value as an estimate of the EPC term) in their simulation study. Previous (e.g., RAGS document (EPA, 1992)) use of the maximum observed value has been recommended as a default option to estimate the EPC term when a 95% *UCL* (e.g., the *H-UCL*) exceeded the maximum value. Only two 95% *UCL* computation methods, namely the Student's t-*UCL* and Land's *H-UCL,* were used previously to estimate the EPC term (e.g., EPA, 1992). ProUCL 4.0 can compute a 95% *UCL* of the mean using several methods based upon the normal, gamma, lognormal, and "nonparametric" distributions. Furthermore, since the EPC term represents the average exposure contracted by an individual over an exposure area (EA) during a long period of time, the EPC term should be estimated by using an average value (such as an appropriate 95% *UCL* of the mean) and not by the maximum observed concentration.

Singh and Singh (2003) also noted that for skewed data sets of small sizes (e.g., < 10-20), the max test does not provide the specified 95% coverage to the population mean, and for larger data sets, it overestimates the EPC term which may require unnecessary further remediation. For the distributions considered, the maximum value is not a sufficient statistic for the unknown population mean. The use of the maximum value as an estimate of the EPC term ignores most (except for the maximum value) of the information contained in a data set. It is, therefore, not desirable to use the maximum observed value as an estimate of the EPC term representing average exposure by an individual over an EA.

It is recommended that the maximum observed value NOT be used as an estimate of the EPC term. For the sake of interested users, ProUCL displays a warning message when the recommended 95% *UCL* (e.g., Hall's bootstrap *UCL*) of the mean exceeds the observed maximum concentration. When a 95% *UCL* exceeds the maximum observed value, ProUCL recommends the use of an alternative *UCL* method based upon a 97.5% or 99% Chebyshev *UCL*.

It should also be noted that for highly skewed data sets, the sample mean indeed can even exceed the upper 90%, or higher, etc., percentiles, and consequently, a 95% *UCL* of the mean can exceed the maximum observed value of a data set. This is especially true when one is dealing with lognormally distributed data sets of small sizes. For such highly skewed data sets which cannot be modeled by a gamma distribution, a 95% *UCL* of the mean should be computed using an appropriate nonparametric method as summarized in Table 15-4.

Chapter 17

Recommendations to Compute a 95% UCL of the Population Mean, μ_1, Using Data Sets with Nondetects with Multiple Detection Limits

This chapter summarizes the recommendations based on the simulation experiments conducted by Singh, Maichle, and Lee (USEPA, 2006) to compare the performances of the UCL computation methods based upon data sets with BDLs and multiple detection limits (DLs). ProUCL 4.0 suggests the use of appropriate UCLs based upon the findings of Singh, Maichle, and Lee (USEPA, 2006). For convenience, the recommended UCL95 computation methods have been tabulated in Table 16 as functions of the sample size, skewness, and censoring intensity. General observations and recommendations regarding the difficulties associated with data sets with NDs are described first.

17.1 General Recommendations and Suggestions

- In practice, it is not easy to verify (perform goodness-of-fit) the distribution of a left-censored data set with NDs. Therefore, emphasis is given on the use of nonparametric UCL95 computation methods, which can also be used to handle multiple detection limits.

- This is specifically true when the percentage (%) of nondetects exceeds 40%-50%.

- Most of the parametric MLE methods assume that there is only one detection limit. It should also be noted that the MLEs behave in an unstable manner when the % of NDs exceeds 40%-50%. Moreover, as mentioned before, it is hard to verify and justify the conclusion of a GOF test for data sets with nondetects in excess of 40%-50%.

- Therefore, for data sets with many nondetects (> 40%-50%), it is suggested to use nonparametric methods to estimate the various environmental parameters (BTVs, EPC terms) of interest and to perform site versus background comparisons.

- In practice, a left-censored data set often has multiple detection limits. For such methods, the KM method can be used. ProUCL 4.0 provides UCL computation methods that can be used on data sets with multiple detection limits including the DL/2 method, KM method, and robust ROS method.

- As mentioned earlier, for reliable and accurate results, it is suggested that the user make sure that the data set under study represents a single statistical population (e.g., background reference area, or an AOC) and not a mixture population (e.g., clean and contaminated site areas).

- It is recommended to identify all of the potential outliers and study them separately. The computation of the statistics such as UCL95 and background statistics should be based upon the majority of the data set representing a single dominant population. Decisions about the appropriate disposition (include or not include) of outliers should be made by all interested members of the project team. When in doubt, it is suggested to compute and

compare all relevant statistics and estimates based upon data sets with and without the outliers. This step will exhibit the undue influence of outliers on the statistics of interest such as UCL95 and UPL95.

- Simple classical outlier identification methods (Dixon test and Rosner test) are also available in ProUCL 4.0. More effective robust outlier procedures (e.g., Rousseeuw and Leroy (1987) and Singh and Nocerino (1995)) are available in Scout (1999).

- In case a data set represents a mixture sample (from two or more populations), one should partition the mixture sample into component sub-samples (e.g., Singh, Singh, and Flatman (1994)).

- Avoid the use of transformations (to achieve symmetry) while computing the upper limits for various environmental applications, as all remediation, cleanup, background evaluation decisions, and risk assessment decisions have to be made using statistics in the original scale. Also, it is more accurate and easier to interpret the results computed in the original scale. The results and statistics computed in the original scale do not suffer from an unknown amount of transformations bias.

- Specifically, avoid the use of a lognormal model even when the data appear to be lognormally distributed. Its use often results in incorrect and unrealistic statistics of no practical purpose or importance or significance. Several variations of estimation methods (e.g., robust ROS and FP-ROS on log-transformed data, delta lognormal method) on log-transformed data have been developed and used by the practitioners. This has caused some confusion among the users of the statistical methods dealing with environmental data sets. The proper use of a lognormal distribution (e.g., how to properly back-transform UCL of mean in the log-scale to obtain a UCL of mean in original scale) is not clear to many users, which in turn may result in the incorrect use and computation of an estimate (= UCL95) of the population mean.

- The parameter in the transformed space may not be of interest to make cleanup decisions. The cleanup and remediation decisions are often made in the original raw scale; therefore, the statistics (e.g., UCL95) computed in transformed space need to be back-transformed in the original scale. It is not clear to a typical user how to back-transform results in the log-scale or any other scale obtained using a Box-Cox (BC)-type transformation to original raw scale. The transformed results often suffer from significant amount of transformation bias.

- The question now arises-how one should back-transform results from a log-space (or any other transformed space) to the original space? Unfortunately, no defensible guidance is available in the environmental literature to address this question. Moreover, the back-transformation formula will change from transformation to transformation (BC-type transformations), and the bias introduced by such transformations will remain unknown. Therefore, in cases when a data set in the "raw" scale cannot be modeled by a parametric distribution, it is desirable to use nonparametric methods (many available in ProUCL 4.0) rather than testing or estimating a parameter in the transformed space.

- On page (78) of Helsel (2005), the use of the robust ROS MLE method (Kroll, C.N. and J.R. Stedinger (1996)) has been suggested to compute summary statistics. In this hybrid

method, MLEs are computed using log-transformed data. Using the regression model as given by equation (3-21) of Section 3, the MLEs of the mean (used as intercept) and *sd* (used as slope) in the log-scale are used to extrapolate the NDs in the log-scale. Just like in the robust ROS method, all of the NDs are transformed back in the original scale by exponentiation. This results in a full data set in the original scale. One may then compute the mean and *sd* using the full data set. The estimates thus obtained are called robust ROS ML estimates (Helsel (2005), and Kroll and Stedinger (1996)). However, the performance of such a hybrid estimation method is not well known. Moreover, for higher censoring levels, the MLE methods sometimes behave in an unstable manner, especially when dealing with moderately skewed to highly skewed data sets (e.g., with $\sigma > 1.0$).

- It should be noted that the performance of this hybrid method is unknown.
- It is not clear why this method is called a robust method.
- The stability of the MLEs obtained using the log-transformed data is doubtful, especially for higher censoring levels.
- The BCA and (% bootstrap) UCLs based upon this method will fail to provide the adequate coverage for the population mean for moderately skewed to highly skewed data sets.

- The DL/2 (t) UCL method does not provide adequate coverage (for any distribution and sample size) for the population mean, even for censoring levels as low as 5%, 10%, 15%. This is contrary to the conjecture and assertion (e.g., EPA (2000)) often made that the DL/2 method can be used for lower (≤ 20%) censoring levels. The coverage provided by the DL/2 (t) method deteriorates fast as the censoring intensity increases.

This DL/2 (t) UCL method is not recommended by the authors and developers of ProUCL 4.0; it is included only for comparison or research purposes.

- The KM method is a preferred method as it can handle multiple detection limits. Moreover, the nonparametric UCL95 methods (KM (BCA), KM (z), KM (%), KM (t)) based upon the KM estimates provide good coverages for the population mean (e.g., Helsel (2005) and Singh et al. (2006)).

- For a symmetric distribution (approximate normality), several UCL95 methods provide good coverage (~95%) for the population mean, including the Winsorization mean, Cohen's MLE (t), Cohen's MLE (Tiku), KM (z), KM (t), KM (%) and KM (BCA) (e.g., Helsel (2005) and Singh et al. (2006)).

17.2 Recommended UCL95 Methods for Normal (Approximate Normal) Distribution

- *For normal and approximately normal (e.g., symmetric or with sd, $\hat{\sigma} < 0.5$) distribution*: The most appropriate UCL95 computation methods for normal or approximately normal distributions are the KM (t) or KM (%) methods. For symmetric distributions, both of these methods perform equally well on left-censored data sets for all censoring levels (% nondetects) and sample sizes.

17.3 Recommended UCL95 Methods for Gamma Distribution

- Highly skewed gamma distributions, G(k, θ), with shape parameter, k ≤ 1:
  - Use the nonparametric KM (Chebyshev) UCL95 method for censoring levels < 30%.
  - Use the nonparametric KM (BCA) UCL95 method for censoring levels in the interval [30%, 50%).
  - Use the nonparametric KM (t) UCL95 method for censoring levels ≥ 50%.

- Moderately skewed gamma distributions, G(k, θ), with shape parameter, 1< k ≤ 2:
  - For censoring level < 10%, use the KM (Chebyshev) UCL95 method.
  - For higher censoring levels [10%, 25%), use the KM (BCA) UCL95 method.
  - For censoring levels in [25%, 40%), use the KM (%) UCL95 method.
  - For censoring levels ≥ 40%, use the KM (t) UCL95 method.

- Mildly skewed gamma distributions, G(k, θ), with k > 2:
  - Use the KM (BCA) UCL95 method for lower censoring levels (< 20%).
  - For censoring levels in the interval [20%, 40%), use the KM (%) UCL95.
  - For censoring ≥ 40%, use the KM (t) UCL95 computation method.

17.4 Recommended UCL95 Methods for Lognormal Distribution

- Mildly skewed data sets with $\hat{\sigma} \leq 1$:
  - For censoring levels (< 20%) and sample of sizes less than 50-70, use the KM (Chebyshev) UCL95.
  - For censoring levels (< 20%) and samples of sizes greater than 50-70, use the KM (BCA) UCL95.
  - For censoring levels in the interval [20%, 40%) and all sample sizes, use the KM (BCA) UCL95.
  - For censoring level ≥ 40%, use the KM (%) or KM (t) UCL95 method.

- Data sets with $\hat{\sigma}$ in the interval (1, 1.5]:
  - For censoring levels < 50% and samples of sizes < 40, use the 97.5% KM (Chebyshev) UCL.
  - For censoring levels < 50% , samples of sizes ≥ 40, use 95% KM (Chebyshev) UCL.
  - For censoring levels ≥ 50%, use the KM (BCA) UCL95 for samples of all sizes.

- Highly skewed data sets with $\hat{\sigma}$ in the interval (1.5, 2]:
  - For sample sizes < 40, censoring levels <50%, use 99% KM (Chebyshev) UCL.
  - For sample sizes ≥ 40, censoring levels < 50%, use 97.5% KM (Chebyshev) UCL.
  - For samples of sizes < 40-50 and censoring levels ≥ 50%, use the 97.5% KM (Chebyshev) UCL.
  - For samples of sizes ≥ 40-50, and censoring levels ≥ 50%, use the 95% KM (Chebyshev) UCL.

- Use a similar pattern for more highly skewed data sets with $\hat{\sigma}$ > 2.0, 3.0:
  - For extremely highly skewed data sets, an appropriate estimate of the EPC term (in terms of adequate coverage) is given by a UCL based upon the Chebyshev inequality

and KM estimates. The confidence coefficient to be used will depend upon the skewness. For highly skewed data sets, a higher (e.g., > 95%) confidence coefficient may have to be used to estimate the EPC.

- As the skewness increases, the confidence coefficient also increases.
- For such highly skewed distributions (with $\hat{\sigma} > 2.0, 3.0$), for lower sample sizes (e.g., < 50-60), one may simply use 99% KM (Chebyshev) UCL to estimate the population mean, EPC term, and other relevant threshold (e.g., UPL, percentiles) values.
- For sample sizes greater than 60, one may use a 97.5% KM (Chebyshev) UCL as an estimate of the population mean or mass.

17.5 Recommended Nonparametric UCL Methods

- *For symmetric or approximately symmetric distribution-free, nonparametric data sets with $\hat{\sigma} < 0.5$:* Use the same UCL computation methods as for the data sets coming from a normal or an approximate normal (symmetric) population. These methods are summarized above in the normal UCL computation section.

- *For skewed distribution-free, nonparametric data sets with $\hat{\sigma} \geq 0.5$:* Most of the recommended UCL computation methods for a lognormal distribution, as described above in the lognormal UCL section, do not assume the lognormality of the data set. Therefore, the UCL computation methods, as described in the lognormal UCL computation section, can be used on skewed nonparametric data sets that do not follow any of the well-known parametric distributions.

The suggested parametric and nonparametric UCL95 computation methods for data sets with nondetect observations are summarized in Table 16, shown on the next page.

Table 17. Recommended UCL95 Computation Methods for Left-Censored Data Sets with Nondetect Observations

Skewness	Sample Size	% ND	95% KM (t)	95% KM (%)	95% KM (Chebyshev)	97.5% KM (Chebyshev)	99% KM (Chebyshev)	95% KM (BCA)
Normal or Approximate Normal (with $\hat{\sigma}$ < 0.5) Distribution								
$\hat{\sigma}$ < 0.5	All *n*	> 0%	●	●				
Gamma Distribution								
$\hat{k}$ ≤ 1	All *n*	< 30%			●			
	All *n*	[30%, 50%)						●
	All *n*	≥ 50%	●					
1 < $\hat{k}$ ≤ 2	All *n*	< 10%			●			
	All *n*	[10%, 25%)						●
	All *n*	[25%, 40%)		●				
	All *n*	≥ 40%	●					
$\hat{k}$ > 2	All *n*	< 20%						●
	All *n*	[20%, 40%)		●				
	All *n*	≥ 40%	●					
Lognormal Distribution								
$\hat{\sigma}$ ≤ 1.0	*n* ≤ 50-70	< 20%			●			
	n > 50-70							●
	All *n*	[20%, 40%)						●
	All *n*	≥ 40%	●	●				
1 < $\hat{\sigma}$ ≤ 1.5	*n* < 40	< 50%				●		
	n ≥ 40				●			
	All *n*	≥ 50%						●
1.5 < $\hat{\sigma}$ ≤ 2.0	*n* < 40	< 50%					●	
	n ≥ 40					●		
	n < 40-50	≥ 50%				●		
	n ≥ 40-50				●			
$\hat{\sigma}$ > 2.0,3.0	*n* < 50-60	> 0%					●	
	n ≥ 60					●		

Table 17. Recommended UCL95 Computation Methods for Left-Censored Data Sets with Nondetect Observations – Continued

Skewness	Sample Size	% ND	95% KM (t)	95% KM (%)	95% KM (Chebyshev)	97.5% KM (Chebyshev)	99% KM (Chebyshev)	95% KM (BCA)
Symmetric or Approximate Symmetric Non-Discernable Distribution								
$\hat{\sigma}$ < 0.5	All *n*	> 0%	●	●				
Moderately Skewed to Highly Skewed Non-Discernable Distribution								
0.5 ≤ $\hat{\sigma}$ ≤ 1.0	*n* ≤ 50-70	< 20%			●			
	n > 50-70							●
	All *n*	[20%, 40%)						●
	All *n*	≥ 40%	●	●				
1 < $\hat{\sigma}$ ≤ 1.5	n < 40	< 50%				●		
	n ≥ 40				●			
	All *n*	≥ 50%						●
1.5 < $\hat{\sigma}$ ≤ 2.0	n < 40	< 50%					●	
	n ≥ 40					●		
	n < 40-50	≥ 50%				●		
	n ≥ 40-50				●			
$\hat{\sigma}$ > 2.0,3.0	*n* < 50-60	> 0%					●	
	n ≥ 60					●		

Note*: In Table 16, phrase "All n" represents only valid (e.g., n > 3) and recommended (n > 8 to 10) values of the sample size, n. As mentioned throughout the report, it is not desirable to use statistical methods on data sets of small sizes (e.g., with n < 8 to 10). However, it should be noted that the sample size requirements and recommendations (n > 8 to 10) as described in this report are not the limitations of the methods considered in this report. One of the main reasons for the recommendation that the sample size should be at least 8 to 10 is that the estimates and UCLs based upon small data sets, especially with many below detection limit observations (e.g., 30%, 40%, 50%, and more), may not be reliable and accurate enough to draw conclusions for environmental applications. It should be noted that in order to be able to use bootstrap re-sampling methods, it is desirable to have a minimum of 10-15 observations (e.g., n > 10-15). Therefore, the phrase "All n" in Table 16, should be interpreted as that the sample size, n, is least 8 to 10. The software, ProUCL 4.0, will provide appropriate warning messages when a user tries to use a method on data sets of small sizes.*

Also, Hall's bootstrap and bootstrap t methods to compute a UCL based upon a full data set (without nondetects) should be used with caution. These two bootstrap methods may yield erratic and unstable UCL results, especially, when outliers are present. In such cases, it is desirable to use alternative UCL

methods based upon Chebyshev inequality. ProUCL software provides a warning message for erratic UCL results based upon Hall's bootstrap t methods.

***Note**: Suggestions regarding the selection of a 95% UCL of mean are provided to help the user to select the most appropriate 95% UCL. These recommendations are based upon the results of the simulation studies summarized in Singh, Maichle, and Lee (2006). For additional insight, the user may want to consult a statistician.*

Glossary

This glossary defines selected words in this User Guide to describe impractically large *UCL* values of the unknown population mean, m_1. In practice, the *UCLs* based upon Land's H-statistic (*H-UCL*), and bootstrap methods such as the bootstrap t and Hall's bootstrap methods (especially when outliers are present) can become impractically large. The *UCLs* based upon these methods often become larger than the *UCLs* based upon all other methods by several orders of magnitude. Such large *UCL* values are not achievable as they do not occur in practice. Words like "unstable" and "unrealistic" are used to describe such impractically large *UCL* values.

UCL: upper confidence limit of the unknown population mean.

Coverage = Coverage Probability: The coverage probability (e.g., = 0.95) of a *UCL* of the population mean represents the confidence coefficient associated with the *UCL*.

Optimum: An interval is optimum if it possesses optimal properties as defined in the statistical literature. This may mean that it is the shortest interval providing the specified coverage (e.g., 0.95) to the population mean. For example, for normally distributed data sets, the *UCL* of the population mean based upon Student's t-distribution is optimum.

Stable *UCL*: The *UCL* of a population mean is a stable *UCL* if it represents a number of a practical merit, which also has some physical meaning. That is, a stable *UCL* represents a realistic number (e.g., contaminant concentration) that can occur in practice. Also, a stable *UCL* provides the specified (at least approximately, as much as possible, as close as possible to the specified value) coverage (e.g., ~0.95) to the population mean.

Reliable *UCL*: This is similar to a stable *UCL*.

Unstable *UCL* = Unreliable *UCL* = Unrealistic *UCL*: The *UCL* of a population mean is unstable, unrealistic, or unreliable if it is orders of magnitude higher than the other *UCLs* of population mean. It represents an impractically large value that cannot be achieved in practice. For example, the use of Land's H-statistic often results in impractically large inflated *UCL* value. *UCLs* such as the bootstrap t *UCL* and Hall's *UCL* can be inflated by outliers, resulting in an impractically large and unstable value. All such impractically large *UCL* values are called unstable, unrealistic, unreliable, or inflated *UCLs* in this User Guide.

References

Breckenridge, R.P. and Crockett, A.B. (1995). Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites. EPA/540/S-96/500.

Cohen, A. C. Jr. 1991. *Truncated and Censored Samples*. 119, Marcel Dekker Inc. New York, NY, 1991.

Conover, W. J. (1999). Practical Nonparametric Statistics. Third Edition. John Wiley.

Department of Navy (Navy, 1999). Handbook for Statistical Analysis of Environmental Background Data. Prepared by SWDIV and EFA West of Naval Facilities Engineering Command. July 1999.

Department of Navy (Navy, 2002a). Guidance for Environmental Background Analysis. Volume 1 Soil. Naval Facilities Engineering Command. April 2002.

Department of Navy (Navy, 2002b). Guidance for Environmental Background Analysis. Volume 2 Sediment. Naval Facilities Engineering Command. May 2002.

Department of Navy (Navy, 2003). Guidance for Environmental Background Analysis. Volume 3 Groundwater. Naval Facilities Engineering Command. May 2002.

Gerlach, R.W., and Nocerino, J. M. (2003). Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples, EPA/600/R-03/027, Nov, 2003.

Gilbert, R.O. (1987), *Statistical Methods for Environmental Pollution Monitoring.* John Wiley and Sons, New York.

Hahn, J. G., and Meeker, W.Q. (1991). Statistical Intervals. A Guide for Practitioners. John Wiley, New York.

Helsel, D.R., Hirsch, R.M. (1994). Statistical Methods in Water Resources. John Wiley.

Hinton, S.W. 1993. *Δ Lognormal Statistical Methodology Performance.* ES&T Environmental Sci. Technol., Vol. 27, No. 10, pp. 2247-2249.

Hogg, R.V., and Craig, A. (1995). Introduction to Mathematical Statistics; 5th edition. Macmillan.

Johnson, N.L., Kotz, S. and Balakrishnan, N. 1994. *Continuous Univariate Distributions, Vol. 1*. Second Edition. John Wiley, New York.

Kaplan, E.L. and Meier, O. 1958. *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, Vol. 53. 457-481.

Land, C. E. (1971), "Confidence Intervals for Linear Functions of the Normal Mean and Variance," *Annals of Mathematical Statistics*, 42, 1187-1205.

Land, C. E. (1975), "*Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance*," in Selected Tables in Mathematical Statistics, Vol. III, American Mathematical Society, Providence, R.I., 385-419.

Laga, J., and Likes, J. 1975, *Sample Sizes for Distribution-Free Tolerance Intervals* Statistical Papers. Vol. 16, No. 1. 39-56

Massachusetts DEP (1995). Guidance for Disposal Site Risk Characterization. Interim Final Policy WSC/ORS - 94-141. July 1995.

Ohio EPA (2004). Background Calculation Methodology. Division of Emergency and Remedial Response (DERR). DERR-00-RR-039P. June 2004.

ProUCL 3.0. (2004). *A Statistical Software*. National Exposure Research Lab, EPA, Las Vegas Nevada, October 2004.

Scheffe, H., and Tukey, J.W. 1944, *A formula for Sample Sizes for Population Tolerance Limits*. The Annals of Mathematical Statistics. Vol 15, 217.

Schulz, T. W., and Griffin, S. (1999), Estimating Risk Assessment Exposure Point Concentrations when Data are Not Normal or Lognormal. Risk Analysis, Vol. 19, No. 4, 1999. A Data Analysis Program, Technology Support Project. EPA, NERL-LV, Las Vegas, NV 89193-3478.

Shacklette, H.T, and Boerngen, J.G. (1984). *Element Concentrations in Soils an Other Surficial Materials in the Conterminous United States,* U.S. Geological Survey Professional Paper 1270.

Shumway, R.H., R.S. Azari, and M. Kayhanian. 2002. *Statistical Approaches to Estimating Mean Water Quality Concentrations with Detection Limits*. Environmental Science and Technology, Vol. 36, pp. 3345-3353.

Singh, A., Singh, A.K., and G. Flatman (1994). Estimation of Background Levels of Contaminants. Math Geology, Vol. 26, No, 3, pp. 361-388.

Singh, A.K., Singh, A., and Engelhardt, M. 1997. *The Lognormal Distribution in Environmental Applications*. Technology Support Center Issue Paper, 182CMB97, EPA/600/R-97/006.

Singh, A. and Nocerino, J.M. 2002. *Robust Estimation of the Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations,* Vol. 60, pp. 69-86.

Singh, A., Singh, A.K., and Iaci, R.J. 2002. *Estimation of the Exposure Point Concentration Term Using a Gamma Distribution,* EPA/600/R-02/084, October 2002.

Singh, A. and Singh, A.K. 2003. *Estimation of the Exposure Point Concentration Term (95% UCL) Using Bias-Corrected Accelerated (BCA) Bootstrap Method and Several other methods for Normal, Lognormal, and Gamma Distributions*. Draft EPA Internal Report.

Singh, A., Maichle, R., and Lee, S. 2006. *On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations.* EPA/600/R-06/022, March 2006.

Singh, A. and Singh, A.K. 2007. *ProUCL Version 4.0 Technical Guide.* Publication EPA/600/R-07/041. April 2007.

USEPA (1989). Methods for Evaluating the Attainment of Cleanup Standards, Vol. 1, Soils and Solid Media, Publication EPA 230/2-89/042.

USEPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term.* Publication EPA 9285.7-08, May 1992.

USEPA (1992a). Guidance for Data Usability in Risk Assessment. Office of Emergency and Remedial Response Publication. 9285.7-09A. PB92-963356.

USEPA (1992b). Statistical Analysis of Ground-water Monitoring Data at RCRA Facilities. Addendum to Interim Final Guidance. Washington, DC: Office of Solid Waste. July 1992.

USEPA (1996). Soil Screening Guidance: User's Guide. Office of Solid Waste and Emergency Response, Washington, DC. EPA/540/R-96/018, April 1996.

USEPA (1994). Statistical Methods for Evaluating the Attainment of Cleanup Standards. Volume 3: Reference Based Standards for Soils and Solid Media. Office of Policy, Planning, and Evaluation, Washington, DC. EPA/230/R-94/004, June 1994.

USEPA (1997). Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM). EPA 402-R-97-016, NUREG-1575, US EPA, Washington, DC.

USEPA (2000). Guidance for Data Quality Assessment: Practical Methods for Data Analysis, EPA QA/G-9, QA00 Version. EPA 600-R-96-084. Quality Assurance Management Staff, Washington, DC. Available at http://www.epa.gov/quality/qa_docs.html.

USEPA. 2002a. *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*. OSWER 9285.6-10. December 2002.

USEPA. 2002b. *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. EPA 540-R-01-003-OSWER 9285.7-41. September 2002.

USEPA. EPA. 2002c. *RCRA Waste Sampling, Draft Technical Guidance – Planning, Implementation and Assessment*. EPA 530-D-02-002, 2002.

USEPA (2004), *ProUCL Version 3.1, A Statistical Software*, National Exposure Research Lab, EPA, Las Vegas Nevada, October 2004.

USEPA. 2006. *Data Quality Assessment: Statistical Methods for Practitioners,* EPA QA/G-9S. EPA/240/B-06/003. Office of Environmental Information, Washington, D.C. Download from: http://www.epa.gov/quality/qs-docs/g9s-final.pdf

USEPA. 2009. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance*. EPA 530-R-09-007, 2009.

Wyoming DEQ (2005). Establishing Site-Specific Background Metals Concentrations in Soil Under the Voluntary Remediation Program. Fact Sheet #24.

United States
Environmental Protection
Agency

Office of Research
and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

EPA/600/R-07/038
May 2010
www.epa.gov

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE ☐ ; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Recycled/Recyclable
Printed with vegetable-based ink on paper that contains a minimum of 50% post-consumer fiber content processed chlorine free