United States
Environmental Protection
Agency

ProUCL Version 5.2.0
User Guide

Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations

Publication #
Release date
www.epa.gov

ProUCL Version 5.2.0
User Guide

Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations

Prepared for:

Felicia Barnett, Director
Office of Research and Development (ORD)
Center for Environmental Solutions and Emergency Response (CESER)
Technical Support Coordination Division (TSCD)
Site Characterization and Monitoring Technical Support Center (SCMTSC)
U.S. Environmental Protection Agency
61 Forsyth Street, Atlanta, GA 30303

Version 5.2.0 prepared by:
Neptune and Company, Inc.
1435 Garrison Street, Suite 201
Lakewood, CO 80215

NOTICE

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

- Versions of the ProUCL software up to version ProUCL 5.1 have been developed by Lockheed Martin, IS&GS - CIVIL under the Science, Engineering, Response and Analytical contract with the U.S. EPA. Improvements included in version 5.2 were made by Neptune and Company, Inc. under the ProUCL and Statistical Support for Site Characterization and Monitor Technical Support Center (SCMTSC) contract with the U.S. EPA and is made available through the U.S. EPA Technical Support Center (TSC) in Atlanta, Georgia (GA).

- Use of any portion of ProUCL that does not comply with the ProUCL Technical 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.

With respect to ProUCL distributed software and documentation, neither the U.S. 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.

ProUCL software is a statistical software package providing statistical methods described in various U.S. EPA guidance documents listed in the Reference section of this document. ProUCL does not describe U.S. EPA policies and should not be considered to represent U.S. EPA policies.

Software Requirements

ProUCL 5.2 has been developed in the Microsoft .NET Framework 4.7.2 using the C# programming language and has been tested on Windows 10 that has this framework pre-installed. ProUCL 5.2 may work on previous versions of the Windows operating system, but it has not been tested on them. The downloadable .NET Framework 4.7.2 files can also be obtained from the following websites:

https://dotnet.microsoft.com/download/dotnet-framework/net472

Installation Instructions when Downloading ProUCL 5.2 from the EPA Web Site

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.

- Download the file ProUCLInstall.msi from the EPA Web site and save to a temporary location. Note: You can delete this file when the installation is complete.
- Double click the ProUCLInstall.msi file and follow the installation instructions provided by the install wizard.
- After installation is complete, 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.

Creating a Shortcut for ProUCL 5.2 on Desktop or Pin to Taskbar

- To create a shortcut of the ProUCL program on your desktop, go to your ProUCL directory in the "Program Files" directory and right click on the executable program (filename is "ProUCL.exe") and select "Create shortcut" from the pop-up menu. Send the shortcut to desktop. The ProUCL icon will now be displayed on your desktop. This shortcut will point to the ProUCL directory consisting of all files required to execute ProUCL 5.2.

- To pin ProUCL to Taskbar, open ProUCL. This will trigger a ProUCL ison to be displayed on the Taskbar icon at the bottom of the computer display window. Right click this icon and click the "Pin to Taskbar" option in the pop-up menu. When pinned, the ProUCL icon will be displayed as a shortcut on the taskbar even when the program is closed.

Caution: Because all files in your ProUCL directory are needed to execute the ProUCL software, you need to generate a shortcut using the process described above. Simply dragging the ProUCL executable file from Window Explorer onto your desktop will not work successfully (an error message will appear) as all files needed to run the software are not available on your desktop. Your shortcut should point to the directory path with all required ProUCL files.

ProUCL 5.2

Software ProUCL version 5.2.0 (ProUCL 5.2), its earlier versions: ProUCL version 3.00.01, 4.00.02, 4.00.04, 4.00.05, 4.1.00, 4.1.01, and ProUCL 5.0.00, 5.1.002 and associated Facts Sheet, User Guides and Technical Guides (e.g., EPA 2010a, 2010b, 2013a, 2013b) can be downloaded from the following EPA website:

https://www.epa.gov/land-research/proucl-software

Recordings of ProUCL webinars offered in 2020, which were conducted on ProUCL 5.1 but are still wholly applicable to version 5.2 can be downloaded from:

ProUCL Utilization 2020: Part 1: ProUCL A to Z
https://clu-in.org/conf/tio/ProUCLAtoZ1/

ProUCL Utilization 2020: Part 2: Trend Analysis
https://clu-in.org/conf/tio/ProUCLAtoZ2/

ProUCL Utilization 2020: Part 3: Background Level Calculations
https://clu-in.org/conf/tio/ProUCLAtoZ3/

Relevant literature used in the development of various ProUCL versions can be downloaded from:
https://www.epa.gov/land-research/proucl-software

Contact Information for all Versions of ProUCL

Since 1999, the ProUCL software has been developed under the direction of the Technical Support Center (TSC). As of November 2007, the direction of the TSC is transferred from Brian Schumacher to Felicia Barnett. Therefore, any comments or questions concerning all versions of ProUCL software should be addressed to:

Felicia Barnett, Director
ORD Site Characterization and Monitoring Technical Support Center (SCMTSC)
Superfund and Technology Liaison, Region 4
U.S. Environmental Protection Agency
61 Forsyth Street SW, Atlanta, GA 30303-8960

barnett.felicia@epa.gov
(404)562-8659
Fax: (404) 562-8439

QUICK START GUIDE

The ProUCL Window

The look and feel of ProUCL 5.2 is similar to that of ProUCL 5.1/5.0; and they share the same names for modules and drop-down menus. ProUCL 5.2 uses a pull-down menu structure, similar to a typical Windows program. Some of the screen shots within this guide will have ProUCL 5.1 or 5.0 in their titles as those screen shots have not been re-generated and replaced, however their functionality should be identical. With that in mind it is important to note that the existing limitations of ProUCL are also still present. If the user wishes to complete multivariate trend analysis or is unsatisfied with the level of customization available in the graphical production options, users should consult a statistician. The screen shown below appears when the program is executed (

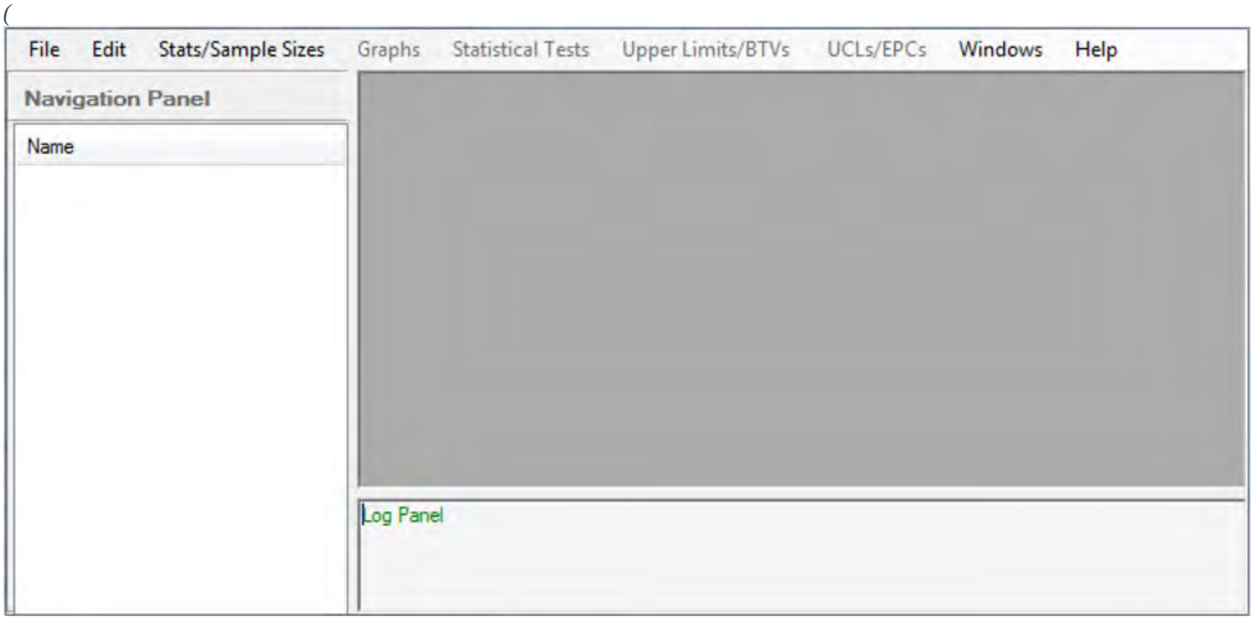

Figure 1. The screen that appears when the program is executed.

).

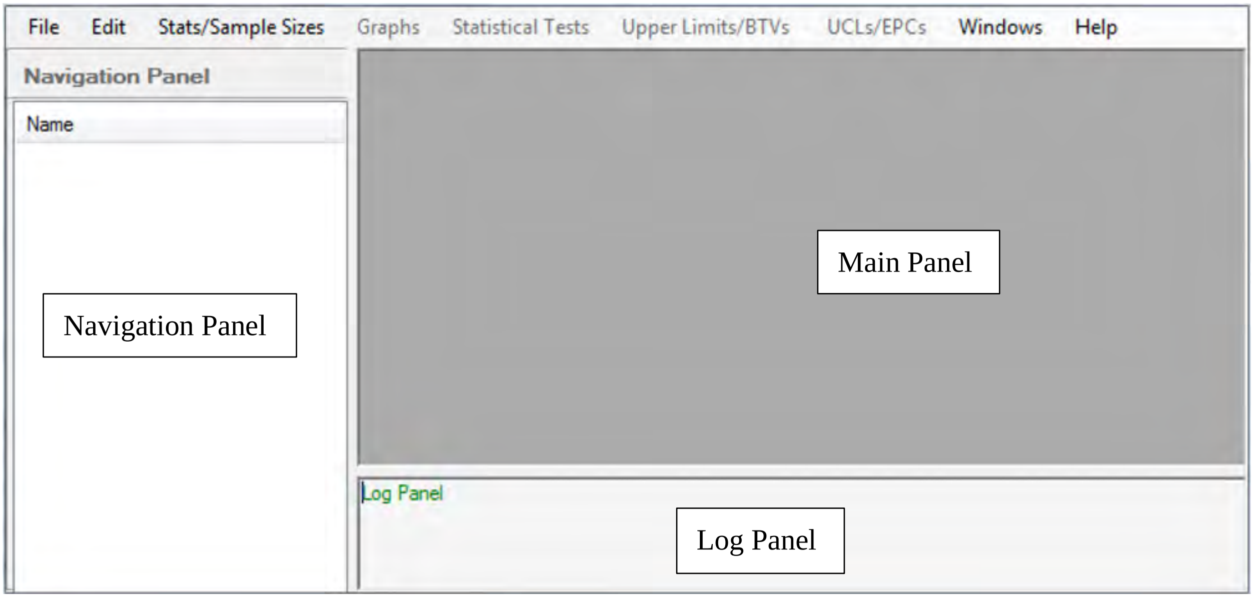

Figure 1. The screen that appears when the program is executed.

The above screen will be the main view users will have for ProUCL 5.2. This screen consists of three main window panels:

- The **MAIN WINDOW** displays data sheets and outputs results from the procedure used.
- The **NAVIGATION PANEL** displays the name of data sets and all generated outputs.
    - The navigation panel can hold up to 40 output files. In order to see more files (data files or generated output files), one can click on Widow Option.
    - In the **NAVIGATION PANEL**, ProUCL assigns self-explanatory names to output files generated using the various modules of ProUCL (**Error! Reference source not found.**). If the same module (e.g., **Time Series Plot**) is used many times, ProUCL identifies them by using letters a, b, c,...and so on as shown below.

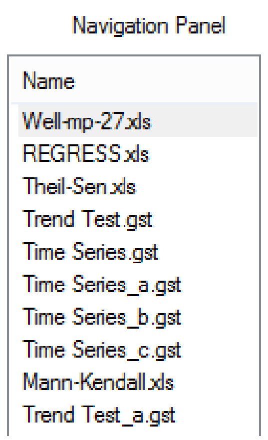

Figure 2. Navigation Panel.

- The user may want to assign names of their choice to these output files when saving them using the "Save" or "Save As" Options.

- The **LOG PANEL** displays transactions in green, warning messages in orange, and errors in red. For an example, when one attempts to run a procedure meant for left-censored data sets on a full-uncensored data set, ProUCL 5.2 will output a warning in orange in this panel.

  - Should both panels be unnecessary, you can choose **Edit ► Configure Display ► Panel ON/OFF (**Error! Reference source not found.**)**.

Turning some panels off gives space to see and print out the statistics of interest. For example, one may want to turn off these panels when multiple variables (e.g., multiple quantile-quantile [Q-Q] plots) are analyzed and goodness-of-fit (GOF) statistics and other statistics may need to be captured for all of the selected variables.

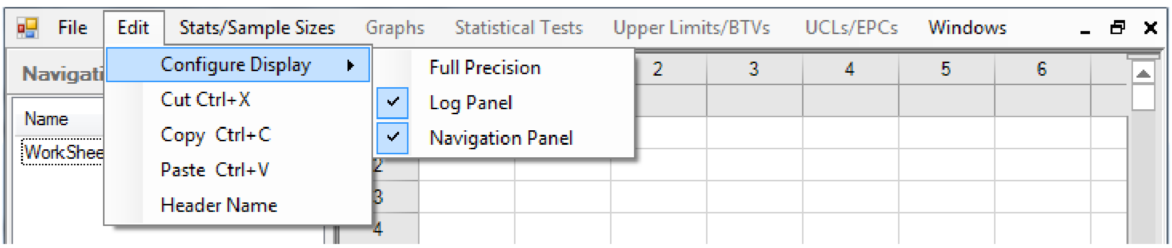

Figure 3. Turning On and Off Panel Displays.

Importing Data in ProUCL

Formatting and importing data for analysis in ProUCL is discussed in detail in Section 1 of this guide.

To import data from Excel spreadsheet, select: **File ► Open Single File Sheet**.

Use Edit module to customize the display and to perform basic editing of imported data.

Statistical Modules

ProUCL 5.2 utilizes the same modules as ProUCL 5.1/5.0 as shown in Figure 3. This document describes how to use each of these modules. The Technical Guide gives some detail about when to use each of them, and the statistical theory behind these methods. For the purpose of a quick introduction statistical functionalities are summarized below.

Stats/Sample Sizes (Section 2): General statistical information in regard to the user's dataset, such as measures of central tendency or variability. It also provides options for regression on order statistics (ROS) imputation of non-detect data as well as estimation of DQO based sample size.

Graphs (Section 3): Provides tools for visual representation of the user's data. These tools include box plots, histograms, and QQ plots.

Statistical Tests (Section 4): Contains all of the different statistical testing methods available within ProUCL, such as outlier tests, goodness of fit tests, single and two sample hypothesis testing methods, as well as ANOVA and trend analysis methods.

Upper Limits/BTVs (Section 5): Methods for upper limit estimates generally used for background threshold value (BTV) analysis. These include options for percentile statistics, upper prediction limits, upper tolerance limits, as well as upper simultaneous limits.

UCLs/EPCs (Section 6): Methods for upper confidence limit (UCL) and exposure point concentration (EPC) estimates based on site data.

EXECUTIVE SUMMARY

ProUCL is software package for commonly used environmental statistics. It was initially developed as a research tool for U.S. EPA scientists and researchers of the Technical Support Center (TSC) and ORD-National Exposure Research Laboratory (NERL), Las Vegas. The intent was to provide a tool for basic statistical calculations that are applicable to site characterization and remediation. As a response to user feedback some additional statistical needs of the environmental projects of the U.S. EPA were addressed in subsequent versions of the ProUCL software from version 1 up to the current 5.2 version. Over the years ProUCL software has been upgraded and enhanced to include more graphical tools and statistical methods described in many EPA guidance documents listed in Reference section of this document.

Methods incorporated in ProUCL cover many common environmental situations and allow environmental practitioners with limited knowledge of statistics to perform calculations to estimate DQO based sample size, establish background levels, compare background and site sample data sets for site evaluation and risk assessment, and perform basic trend analysis. Some methods for analysis of data sets with nondetect values are built in this software. Statistical modules are organized as drop-down menus to allow users easy access to statistical methods and tests.

However, as any software, ProUCL has limitations. The software (version 5.2) does not include advanced statistical methods applicable to very skewed data sets or biased sampling designs and does not include geostatistical methods. ProUCL also lacks capabilities to perform simulations or automation of repeating tasks. Therefore, environmental practitioners are strongly encouraged to seek advice from environmental statisticians on planning of environmental studies and choosing applicable statistical methods for sampling design used in the project.

Several improvements have been made to the decision logic for the recommendation of UCLs for version 5.2. The reliance on goodness of fit tests to select appropriate UCLs is reduced. The Chebyshev UCL is no longer recommended, and the H UCL is only recommended in cases of very large sample sizes when there is high confidence that the assumption of lognormality is met to a good approximation. In some cases, data may be too skewed or not numerous enough to determine an appropriate UCL. Version 5.2 does not provide a recommendation in these cases but encourages the user to verify that the data were collected randomly (rather than through biased sampling, such as hot spot delineation sampling), to consider site knowledge that may explain why the data may be skewed (such as small areas of high concentrations), and to contact a statistician if ProUCL cannot provide a recommendation.

Another improvement of ProUCL 5.2 is that libraries and developer tools (Microsoft .NET, Spread.NET (previously FarPoint), ChartFX, and Visual Studio) were updated to the latest available version. These tools have all had one or more version releases since 2016 when version ProUCL 5.1 was released.

In parallel with ProUCL improvements released as version 5.2, the ProUCL User guide and Technical guide were updated as well. The User Guide was reorganized to be better aligned with the software layout. Sections are now organized in the same order as ProUCL software drop-down menus. The last chapter of User Guide provides some limited guidance on the use of statistical methods incorporated in ProUCL software. The Technical Guide was updated to include the description and justification for decision logic improvements incorporated in version 5.2.

ProUCL has been verified against, and is agreement with, the results obtained by using other software packages including Minitab, SAS®, and CRAN R packages. Statistical methods incorporated in ProUCL have also been tested and verified extensively by the developers, researchers, scientists, and users. Software is continuously improved to address findings and observations of hundreds of users with different levels of statistical background spanning from environmental practitioners to professional statisticians performing analysis on thousands of environmental data sets.

ProUCL is available for free at the U.S. EPA Site Characterization and Monitoring Technical Support Center (SCMTSC) website.

https://www.epa.gov/land-research/proucl-software SCMTSC staff also provide some user support. This may include answering questions related to the use of ProUCL software and technical support to EPA superfund project managers or technical staff.

Table of Contents

ACKNOWLEDGEMENTS

We wish to express our gratitude and thanks to our friends and colleagues who have contributed during the development of past versions of ProUCL and to all of the many people who reviewed, tested, and gave helpful suggestions throughout the development of the ProUCL software package. We wish to especially acknowledge current and former EPA scientists including Deana Crumbling, Nancy Rios-Jafolla, Tim Frederick, Jean Balent, Dr. Maliha Nash, Kira Lynch, and Marc Stiffleman; James Durant of ATSDR, Dr. Steve Roberts of University of Florida, Dr. Elise A. Striz of the National Regulatory Commission (NRC), and Drs. Phillip Goodrum and John Samuelian of Integral Consulting Inc. as well as Dr. D. Beal of Leidos for testing and reviewing ProUCL and its associated guidance documents, and for providing helpful comments and suggestions. Finally, we want to express gratitude to statisticians and computer scientists of Neptune and Company, Inc. for the latest improvements included in ProUCL version 5.2.

Special thanks go to Dr. Anita Singh, Ms. Donna Getty and Mr. Richard Leuser of Lockheed Martin, for significant contribution to the development of ProUCL software and providing a thorough technical and editorial review of ProUCL 5.1 and also ProUCL 5.0 User Guide and Technical Guide. A special note of thanks is due to Ms. Felicia Barnett of EPA ORD Site Characterization and Monitoring Technical Support Center (SCMTSC), without whose assistance the development of the ProUCL 5.1 software and associated guidance documents would not have been possible.

Finally, we wish to dedicate the ProUCL 5.1 (and ProUCL 5.0) software package to our friend and colleague, John M. Nocerino who had contributed significantly in the development of ProUCL and Scout software packages.

ACRONYMS and ABBREVIATIONS

ACL	Alternative compliance or concentration limit
A-D, AD	Anderson-Darling test
AL	Action limit
AOC	Area(s) of concern
ANOVA	Analysis of variance
A_0	Not to exceed compliance limit or specified action level
BC	Box-Cox transformation
BCA	Bias-corrected accelerated bootstrap method
BD	Binomial distribution
BISS	Background Incremental Sample Simulator
BTV	Background threshold value
CC, cc	Confidence coefficient
CERCLA	Comprehensive Environmental Recovery, Compensation, and Liability Act
CL	Compliance limit
CLT	Central Limit Theorem
COPC	Contaminant/constituent of potential concern
C_s	Cleanup standards
CSM	Conceptual site model
CV	Coefficient of variation
Df	Degrees of freedom
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 NDs replaced by 1/2 of the respective detection

	limits
DOE	Department of Energy
DQOs	Data quality objectives
DU	Decision unit
EA	Exposure area
EDF	Empirical distribution function
EM	Expectation maximization
EPA	United States Environmental Protection Agency
EPC	Exposure point concentration
GA	Georgia
GB	Gigabyte
GHz	Gigahertz
GROS	Gamma ROS
GOF, G.O.F.	Goodness-of-fit
GUI	Graphical user interface
GW	Groundwater
H_A	Alternative hypothesis
H_0	Null hypothesis
H-UCL	UCL based upon Land's H-statistic
ISM	Incremental sampling methodology
ITRC	Interstate Technology & Regulatory Council
k, K	Positive integer representing future or next k observations
K	Shape parameter of a gamma distribution
K,k	Number of nondetects in a data set

k hat	MLE of the shape parameter of a gamma distribution
k star	Biased corrected MLE of the shape parameter of a gamma distribution
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 critical value
KM (z)	UCL based upon Kaplan-Meier estimates using critical value of a standard normal distribution
K-M, KM	Kaplan-Meier
K-S, KS	Kolmogorov-Smirnov
K-W	Kruskal Wallis
LCL	Lower confidence limit
LN, *ln*	Lognormal distribution
LCL	Lower confidence limit of mean
LPL	Lower prediction limit
LROS	LogROS; robust ROS
LTL	Lower tolerance limit
LSL	Lower simultaneous limit
M,m	Applied to incremental sampling: number in increments in an ISM sample
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MCL	Maximum concentration limit, maximum compliance limit
MDD	Minimum detectable difference
MDL	Method detection limit
MK, M-K	Mann-Kendall
ML	Maximum likelihood

MLE	Maximum likelihood estimate
n	Number of observations/measurements in a sample
N	Number of observations/measurements in a population
MVUE	Minimum variance unbiased estimate
MW	Monitoring well
NARPM	National Association of Remedial Project Managers
ND, nd, Nd	Nondetect
NERL	National Exposure Research Laboratory
NRC	Nuclear Regulatory Commission
OKG	Orthogonalized Kettenring Gnanadesikan
OLS	Ordinary least squares
ORD	Office of Research and Development
OSRTI	Office of Superfund Remediation and Technology Innovation
OU	Operating unit
PCA	Principal component analysis
PDF, pdf	Probability density function
.pdf	Files in Portable Document Format
PRG	Preliminary remediation goals
PROP	Proposed influence function
p-values	Probability-values
QA	Quality assurance
QC	Quality
Q-Q	Quantile-quantile
R,r	Applied to incremental sampling: number of replicates of ISM samples

RAGS	Risk Assessment Guidance for Superfund
RCRA	Resource Conservation and Recovery Act
RL	Reporting limit
RMLE	Restricted maximum likelihood estimate
ROS	Regression on order statistics
RPM	Remedial Project Manager
RSD	Relative standard deviation
RV	Random variable
S	Substantial difference
SCMTSC	Site Characterization and Monitoring Technical Support Center
SD, Sd, sd	Standard deviation
SE	Standard error
SND	Standard Normal Distribution
SNV	Standard Normal Variate
SSL	Soil screening levels
SQL	Sample quantitation limit
SU	Sampling unit
S-W, SW	Shapiro-Wilk
T-S	Theil-Sen
TSC	Technical Support Center
TW, T-W	Tarone-Ware
UCL	Upper confidence limit
UCL95	95% upper confidence limit
UPL	Upper prediction limit

U.S. EPA	United States Environmental Protection Agency
UTL	Upper tolerance limit
UTL95-95	95% upper tolerance limit with 95% coverage
USGS	U.S. Geological Survey
USL	Upper simultaneous limit
vs.	Versus
WMW	Wilcoxon-Mann-Whitney
WRS	Wilcoxon Rank Sum
WSR	Wilcoxon Signed Rank
x_p	p^{th} percentile of a distribution
<	Less than
>	Greater than
≥	Greater than or equal to
≤	Less than or equal to
Δ	Greek letter denoting the width of the gray region associated with hypothesis testing
Σ	Greek letter representing the summation of several mathematical quantities, numbers
%	Percent
α	Type I error rate
β	Type II error rate
ϴ	Scale parameter of the gamma distribution
Σ	Standard deviation of the log-transformed data
^	carat sign over a parameter, indicates that it represents a statistic/estimate computed using the sampled data

1 Preparing and Entering Data

The majority of the information provided in Chapter 1 is also available in the first of the ProUCL 2020 presentations available online here: ProUCL Utilization 2020: Part 1: ProUCL A to Z.

1.1 Entering and Manipulating Data

1.1.1 Creating a New Data Set

By executing ProUCL, the following options in Figure 1-1will appear (the title will show ProUCL version installed).

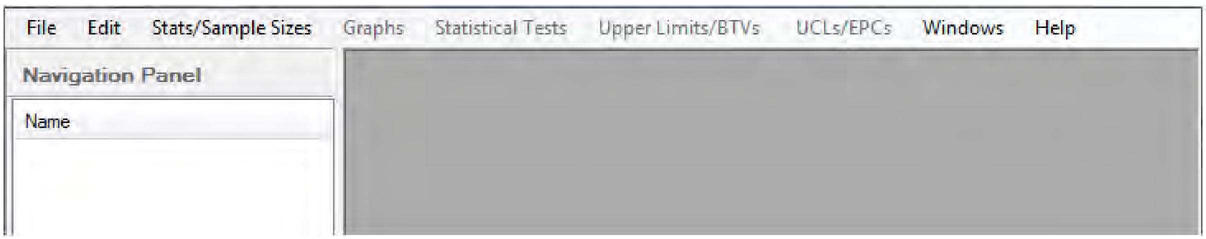

Figure 1-1. Toolbar Upon Execution of the Program.

By choosing the **File ► New** option, a new worksheet shown below will appear (Figure 1-2). The user enters variable names and data following the ProUCL input file format requirements described in Section 1.3.

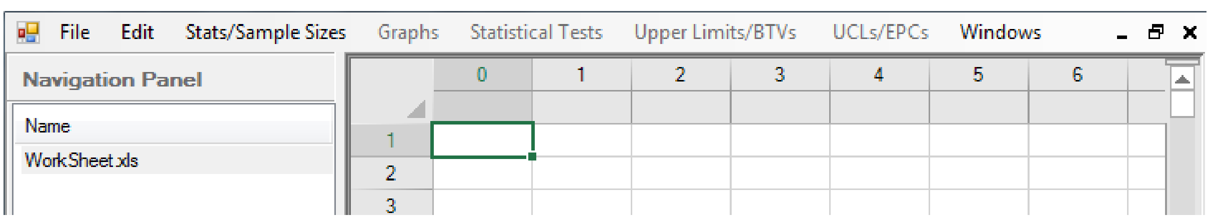

Figure 1-2. Creating a New Worksheet.

Note: When entering data or loading data from existing source ProUCL will only read data that is presented in long format. That is to mean, each column represents exactly one variable with each row being one observation of all available variables. Additionally data types within a column should be consistent as ProUCL will read text strings within a numeric column as a missing value, see Section 1.2.4.

1.2 Opening an Existing Data Set

The user can open an existing worksheet (*.xls, *.xlsx, *.wst, and *.ost) by choosing the **File ► Open Single File Sheet** option. The drop-down menu in Figure 1-3 will appear:

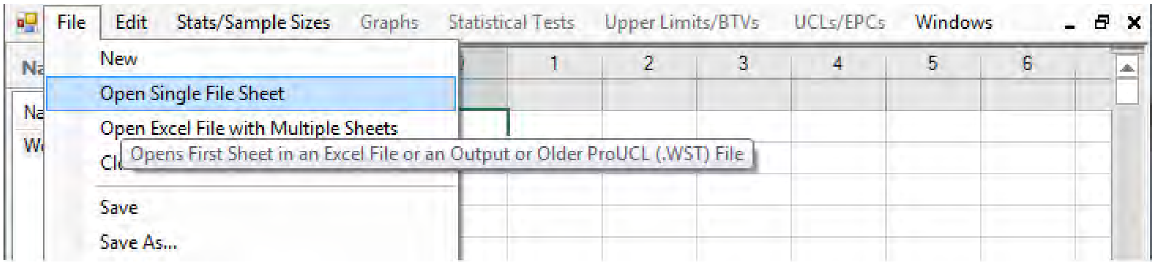

Figure 1-3. Opening an Existing Worksheet – Part One.

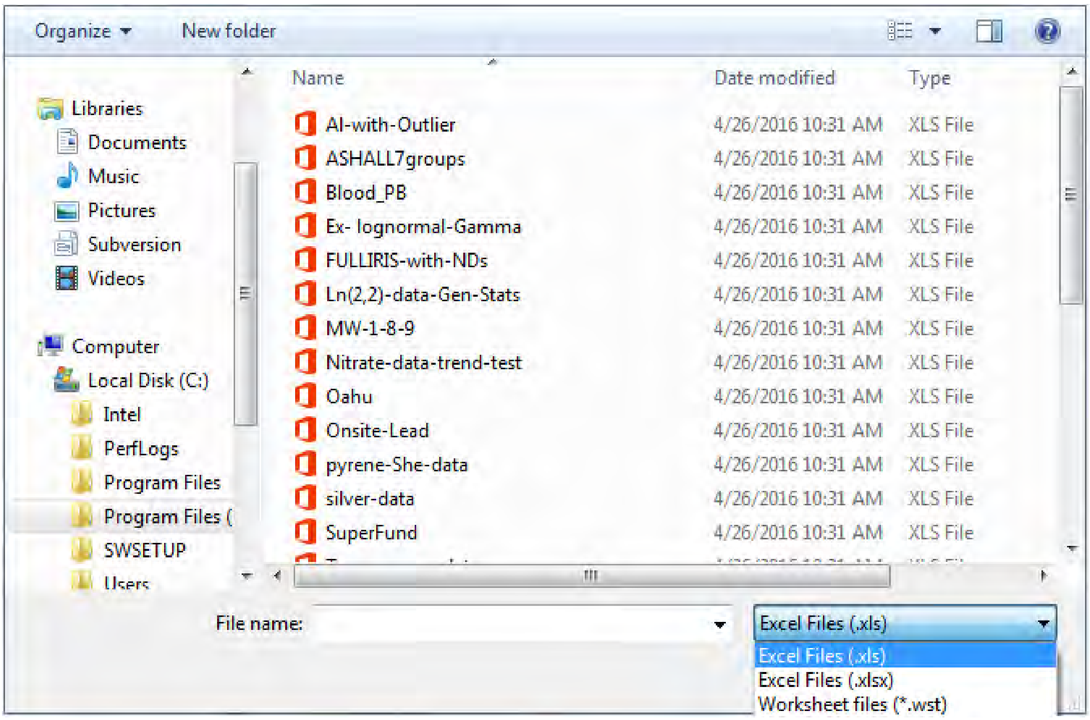

Figure 1-4. Opening an Existing Worksheet – Part Two.

Choose a file by highlighting the type of file such as **.xls** as shown in Figure 1-4. This option can also be used to read in a *.wst worksheet and *.ost output sheet generated by earlier versions (e.g., ProUCL 4.1 and older) of ProUCL.

By choosing the **File ► Excel Multiple Sheets** option, the user can open an Excel file consisting of multiple sheets. Each sheet will be opened as a separate file to be processed individually by ProUCL.

Caution: If you are editing a file (e.g., an excel file using Excel), make sure to close the file before importing the file into ProUCL using the file open option.

Caution: ProUCL 5.2 will often successfully read **.xlsx** files however this is inconsistent, and it is advised to save any excel data a user wishes to import as **.xls** before import.

1.2.1 Input File Format

The program can read Excel files. The user can perform typical Cut, Paste, and Copy operations available under the Edit Menu Option as shown below.

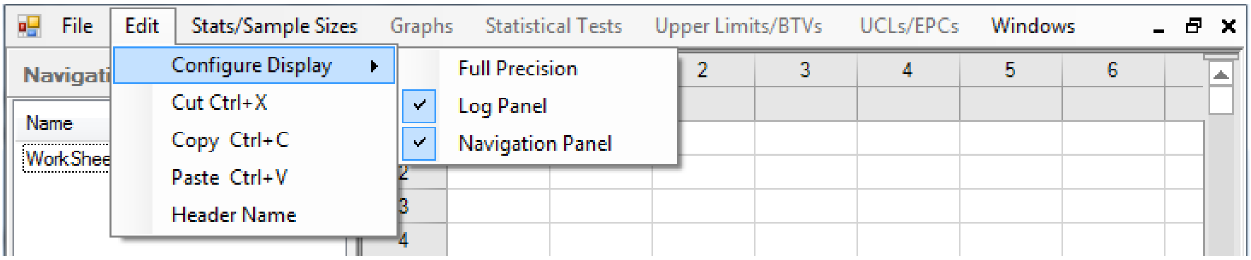

Figure 1-5. Turning On and Off Panel Displays.

The first row in all input data files must 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.

An example Group-ID column could hold the labels for the groups (e.g., Background, AOC1, AOC2, 1, 2, 3, a, b, c, Site 1, Site 2) that might be present in the data set. Alphanumeric strings (e.g., Surface, Sub-surface) can be used to label the various groups. Most of the modules of ProUCL can process data by a group variable.

The data file can have multiple variables (columns) with unequal numbers of observations. Most of the modules of ProUCL can process data by a group variable.

1.2.2 Handling Non-detect Observations and Generating Files with Non-detects

Several modules of ProUCL (e.g., Statistical Tests, Upper limits/BTVs, UCLs/EPCs) handle data sets containing ND observations with single and multiple DLs.

The user informs the program about the status of a variable consisting of NDs. For a variable with ND 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. No qualifiers or flags (e.g., J, B, U, UJ, X) should be entered in data files with ND observations.

Data for variables with ND values are provided in two columns. One column consists of numerical values of detected observations and numerical values of detection limits (or reporting limits) associated with non-detect observations. The second column represents their detection status consisting of only 0 (ND) and 1 (detected) values. The name of this second column, representing the detection status should start with d_, or D_ (not case sensitive) and the column name associated with this detection status. The detection status column with variable name starting with a D_ (or a d_) should have only two values: 0 for ND values, and 1 for detected observations.

For example, if an observation column has the header name, Arsenic, then the associated detection status column would be named D_Arsenic. If this format is not followed, the program will not recognize that the data set has NDs.

An example data set illustrating these points is given as follows. ProUCL does not distinguish between lowercase and uppercase letters

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			

Figure 1-6. Example Data Set with Non-Detects.

1.2.3 Caution Regarding Non-detects

Care should be taken to avoid any misrepresentation of detected and non-detected values. Specifically, do not include 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 ND, 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 (such as qualifiers) 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 an ND value.**

When computing statistics for full uncensored data sets without any ND values, it is important to note that ProUCL will treat all observations in the selected variable column as detected values regardless of an associated d_variable column. Therefore, the user should use only columns with no NDs if they wish to compute statistics without ND values.

1.2.4 Handling Missing Values

Within ProUCL there are three types of cell entry that are treated as missing values. Those missing values are omitted from all future statistical evaluations.

These types are

a. Alphanumeric Strings- Any value entered that consists of non-numerical values will be discarded ie: “three” will be treated as a missing value not counted as 3. The one exception to this is that E can be used for scientific notation such as 1E5.
b. Blank Cells- Any cell that is left blank will be treated as a missing value.
c. Note: If a missing value is located in a non-detect column, for example D_Arsenic, while the associated value in the Arsenic column is not missing, the associated value will be treated as a non-detect.
d. A specific large value cutoff- The value 1E31 (= 1×10^{31}) or any number greater than that value is counted as a special character that will be discarded from future analysis and treated as a missing value.

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 5.0 and newer will treat that blank value (or 1e31 value) as a new group. All variables and values that correspond to this missing value will be treated as part of a new group and not with any existing groups. It is therefore important to check the consistency and validity of all data sets before performing statistical evaluations.

ProUCL prints out the number of missing values (if any) and the number of reported values (excluding the missing values) associated with each variable in the data sheet. This information is provided in several output sheets (e.g., General statistics, BTVs, UCLs, Outliers, OLS, Trend Tests) generated by ProUCL.

Example 1-1: The following example illustrates the notion of Valid Samples, Distinct Samples, and Missing Values with a toy 17 sample dataset. The data set also has ND values.

Table 1-1. Example 1-1 Data.

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

2.3	1	Used
2.3	1	Used
34	1	Used
73	1	Used

Valid Samples: Represents the total number of observations (censored and uncensored) excluding the missing values. In this case the number of valid samples = 13 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 treated as missing values. Specifically, all alphanumeric values including blanks are considered to be missing values. Big numbers such as 1.0e31 are also treated as missing values and are considered as not valid observations. In the example above the number of missing values = 4.

Distinct Samples: The number of unique samples or number of distinct samples represents all unique (or distinct) detected and non-detected values. This is computed separately for detects and NDs. 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 small. In the example above total number of unique or distinct samples = 8, number of distinct detects = 6, and number of distinct NDs (with different detection limits) = 2.

Table 1-2. Summary Statistics for Example 5-1.

General Statistics			
Total Number of Observations	7	Number of Distinct Observations	6
Number of Detects	2	Number of Non-Detects	5
Number of Distinct Detects	2	Number of Distinct Non-Detects	4
Minimum Detect	10	Minimum Non-Detect	1
Maximum Detect	13	Maximum Non-Detect	5
Variance Detects	4.5	Percent Non-Detects	71.43%
Mean Detects	11.5	SD Detects	2.121
Median Detects	11.5	CV Detects	0.184
Skewness Detects	N/A	Kurtosis Detects	N/A
Mean of Logged Detects	2.434	SD of Logged Detects	0.186
Warning: Data set has only 2 Detected Values.			
This is not enough to compute meaningful or reliable statistics and estimates.			
Note: Sample size is small (e.g., <10), if data are collected using ISM approach, you should use			
guidance provided in ITRC Tech Reg Guide on ISM (ITRC, 2012) to compute statistics of interest.			
For example, you may want to use Chebyshev UCL to estimate EPC (ITRC, 2012).			
Chebyshev UCL can be computed using the Nonparametric and All UCL Options of ProUCL 5.1			

1.2.5 *Number Precision*

The user may turn “Full Precision” on or off by choosing **Edit ► Configure Display ► Full Precision On/OFF**

By leaving “Full Precision” turned **off**, ProUCL will display numerical values using an appropriate (default) decimal digit option; and by turning “Full Precision” **on**, numbers will be carried out to 7 decimal places.

The “Full Precision” **on** option is specifically useful when 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 very small with several leading zeros (e.g., 0.00007332) after the decimal. In such situations, one may want to use the "Full Precision” **on** option to see nonzero values after the decimal.

<u>Note:</u> For the purpose of this User Guide, unless noted otherwise, all examples have used the “Full Precision” **OFF** option. This option prints out results up to 3 significant digits after the decimal.

1.2.6 *Entering and Changing a Header Name*

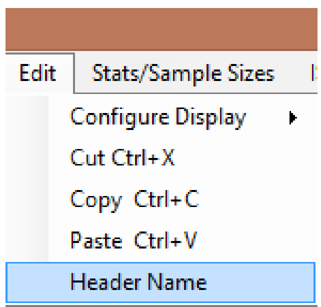

Figure 1-7. Editing a Header Name – Part One.

The user can change variable names (Header Name) using the following process. Highlight the column whose header name (variable name) you want to change by clicking either the column number or the header as shown below.

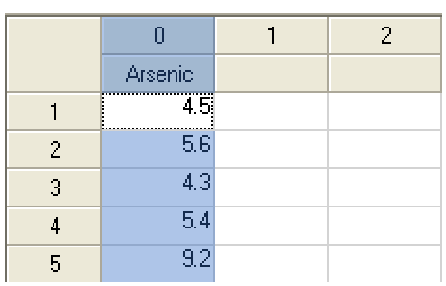

Figure 1-8. Editing a Header Name – Part Two.

Right-click and then click **Header Name**.

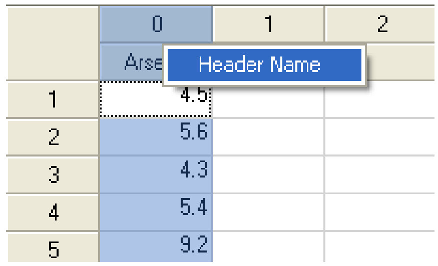

Figure 1-9. Editing a Header Name – Part Three.

Change the Header Name.

Figure 1-10. Editing a Header Name – Part Four.

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

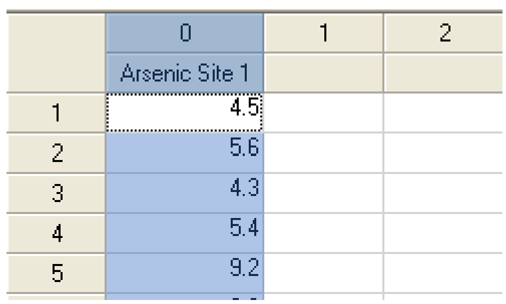

Figure 1-11. Changed Header Name.

1.2.7 Saving Files

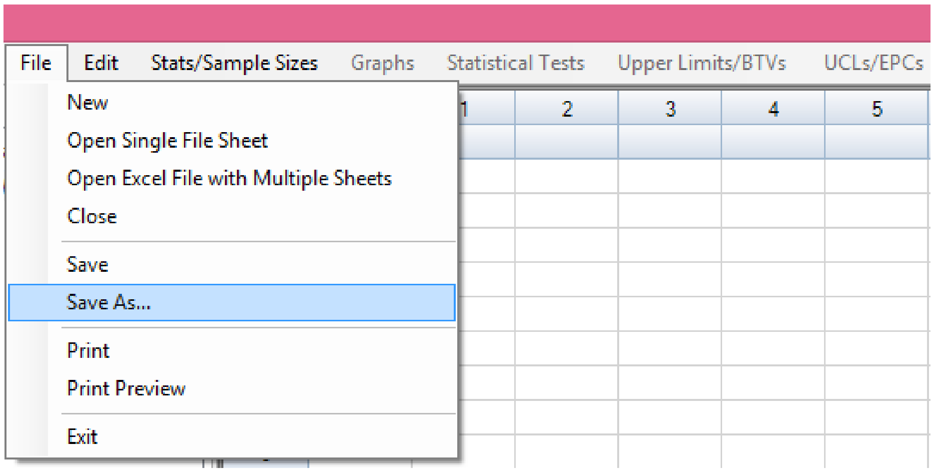

Figure 1-12. Saving as Excel File.

The **Save** option allows the user to save the active window in .xls or .xlsx formats.

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

1.2.8 Editing

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

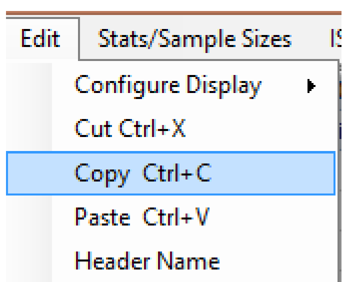

Figure 1-13. Edit Options.

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.

1.3 Common Options and Functionalities

1.3.1 Warning Messages and Recommendations

ProUCL 5.2 provides warning messages to alert the user when there might be a problem with the data or computations. In addition to the warnings given by ProUCL 5.1, version 5.2 encourages the user to 1) verify that the data were collected randomly (rather than through biased sampling, such as hot spot delineation sampling or best professional judgment sampling); 2) consider site knowledge that may explain why the data may be skewed (such as small areas of high concentrations), 3) and to contact a statistician if ProUCL cannot provide a recommendation.

1.3.1.1 Insufficient Amount of Data

ProUCL provides warning messages and recommendations for data sets with an insufficient amount of data for calculating meaningful estimates and statistics of interest. For example, it is not desirable to compute an estimate of the EPC term based upon a discrete (as opposed to composite or ISM) data set of size less than 5, especially when NDs are also present in the data set.

However, to accommodate the computation of UCLs and other limits based upon ISM data sets, ProUCL allows users to compute UCLs, UPLs, and UTLs based upon data sets of sizes as small as 3. The user is advised to follow the guidance provided in the ITRC ISM Technical Regulatory Guidance Document (2012) to select an appropriate UCL95 to estimate the EPC term. Due to lower variability in ISM data, the minimum sample size requirements for statistical methods used on ISM data are lower than the minimum sample size requirements for statistical methods used on discrete data sets.

It is suggested that for data sets composed of observations resulting from discrete sampling, at least 10 observations should be collected to compute UCLs and various other limits.

Some examples of data sets with insufficient amount of data include data sets with less than 3 distinct observations, data sets with only two detected observations, and data sets consisting of all non-detects.

Some of the warning messages generated by ProUCL are shown as follows.

Table 1-2. Warning Messages.

UCL Statistics for Uncensored Full Data Sets			
User Selected Options			
Date/Time of Computation	3/13/2013 9:26:43 PM		
From File	Not-enough-data-set.xls		
Full Precision	OFF		
Confidence Coefficient	95%		
Number of Bootstrap Operations	2000		
x			
General Statistics			
Total Number of Observations	2	Number of Distinct Observations	2
		Number of Missing Observations	0
Minimum	2	Mean	4.5
Maximum	7	Median	4.5
Warning: This data set only has 2 observations!			
Data set is too small to compute reliable and meaningful statistics and estimates!			
The data set for variable x was not processed!			
It is suggested to collect at least 8 to 10 observations before using these statistical methods!			
If possible, compute and collect Data Quality Objectives (DQO) based sample size and analytical results.			

UCL Statistics for Data Sets with Non-Detects			
User Selected Options			
Date/Time of Computation	3/13/2013 9:27:39 PM		
From File	Not-enough-data-set.xls		
Full Precision	OFF		
Confidence Coefficient	95%		
Number of Bootstrap Operations	2000		
y			
General Statistics			
Total Number of Observations	7	Number of Distinct Observations	6
Number of Detects	2	Number of Non-Detects	5
Number of Distinct Detects	2	Number of Distinct Non-Detects	4
Minimum Detect	10	Minimum Non-Detect	1
Maximum Detect	13	Maximum Non-Detect	5
Variance Detects	4.5	Percent Non-Detects	71.43%
Mean Detects	11.5	SD Detects	2.121
Median Detects	11.5	CV Detects	0.184
Skewness Detects	N/A	Kurtosis Detects	N/A
Mean of Logged Detects	2.434	SD of Logged Detects	0.186
Warning: Data set has only 2 Detected Values.			
This is not enough to compute meaningful or reliable statistics and estimates.			
Normal GOF Test on Detects Only			
Not Enough Data to Perform GOF Test			

Table 1-2 (continued). Warning Messages.

	Background Statistics for Data Sets with Non-Detects
User Selected Options	
From File	Not-enough-data-set_a.xls
Full Precision	OFF
Confidence Coefficient	95%
Coverage	95%
Different or Future K Observations	1
Number of Bootstrap Operations	2000

yy

	General Statistics		
Total Number of Observations	7	Number of Missing Observations	0
Number of Distinct Observations	6		
Number of Detects	0	Number of Non-Detects	7
Number of Distinct Detects	0	Number of Distinct Non-Detects	6
Minimum Detect	N/A	Minimum Non-Detect	1
Maximum Detect	N/A	Maximum Non-Detect	13
Variance Detected	N/A	Percent Non-Detects	100%
Mean Detected	N/A	SD Detected	N/A
Mean of Detected Logged Data	N/A	SD of Detected Logged Data	N/A

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 yy was not processed!

1.3.1.2 Biased Sampling

Due to the nature of environmental contamination, sampling based on professional judgement, rather than random sampling, is quite common. Especially if some data are historical, the methodology for selecting locations may be unknown. Typically, moderate to high skew in the data is an indication that the data may include a small number of locations that were specifically selected to characterize areas of particularly high concentrations (i.e., judgmental sampling). ProUCL currently supports calculation of UCLs from randomly collected locations only. However, statistical methods exist that can account for the bias in sample collection. Users should contact a statistician for assistance with such calculations. Therefore, ProUCL 5.2 includes a warning if the coefficient of variation (CV) of the data is greater than 1, alerting the user to confirm that all the data were collected from randomly selected locations.

1.3.1.3 Recommendation Not Available

ProUCL is intended to provide guidance for the most common environmental data sets and situations, and to allow practitioners with limited knowledge of statistics to perform calculations to estimate UCLs as well as perform other basic statistical analyses. However, it cannot replace analysis performed by a trained statistician. There are certain situations where all choices of UCL methods have serious drawbacks (for example, if the sample size is small and the data are highly skewed). Section 2.5.1 of the Technical Guide) provides further details. Rather than recommending a UCL that may seriously overestimate or underestimate the mean, ProUCL 5.2 encourages the user to contact a trained statistician in such situations.

1.3.2 Select Variables Screen and the Grouping Variable

- The **Select Variable** screen is associated with all modules of ProUCL.
- Variables need to be selected to perform statistical analyses.
- When the user clicks on a drop-down menu for a statistical procedure (e.g., UCLs/EPCs), the following window will appear.

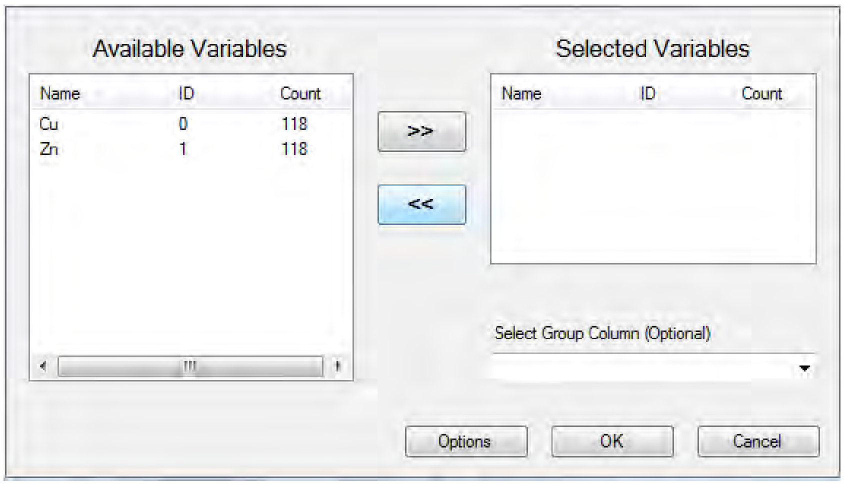

Figure 1-14. Selecting Variables.

- The **Options** button is available in certain menus. The use of this option leads to another pop-up window such as shown below. This window provides the options associated with the selected statistical method (e.g., BTVs, OLS Regression).

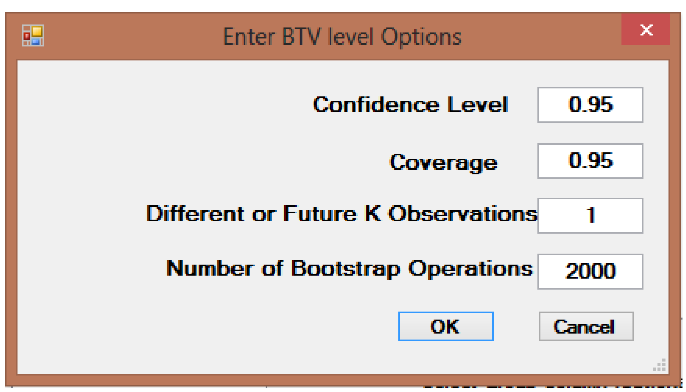

Figure 1-15. Options Associated with BTVs.

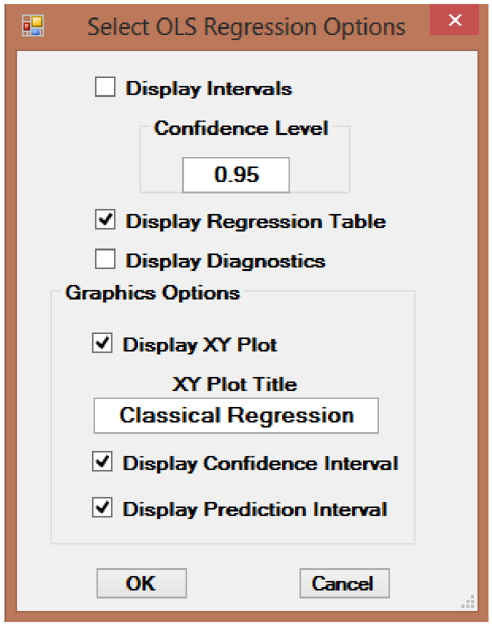

Figure 1-16. Options Associated with OLS Regression.

- ProUCL can process multiple variables simultaneously. ProUCL software can generate graphs, and compute UCLs, and background statistics simultaneously for all selected variables shown in the right panel of the screen shot displayed on the previous page.

- If the user wants to perform statistical analysis on a variable (e.g., manganese) by a Group variable, click the arrow below the **Select Group Column (Optional)** to get a drop-down list of available variables from which to select an appropriate group variable. For example, a group variable (e.g., Zone) can have alphanumeric values such as MW8, 27, or in this case two options of Alluvial Fan, and Basin Trough. Thus, in this example, the group variable name, Zone, takes 2 values: Alluvial Fan, and Basin Trough. The selected statistical method (e.g., GOF test) performs computations on data sets for all the groups associated with the selected group variable (e.g., Zone).

	0	1	2	3	4
	Cu	Zn	Zone	D_Cu	D_Zn
49	5	10	Alluvial Fan	0	1
50	2	20	Alluvial Fan	1	1
51	10	20	Alluvial Fan	0	1
52	5	20	Alluvial Fan	0	1
53	5	10	Alluvial Fan	0	0
54	2	20	Alluvial Fan	1	1
55	10	23	Alluvial Fan	1	1
56	2	17	Alluvial Fan	1	1
57	4	10	Alluvial Fan	1	1
58	5	10	Alluvial Fan	0	0
59	2	10	Alluvial Fan	1	1
60	3	20	Alluvial Fan	1	1
61	9	29	Alluvial Fan	1	1
62	5	20	Alluvial Fan	0	1
63	2	10	Alluvial Fan	1	0
64	2	10	Alluvial Fan	1	1
65	2	10	Alluvial Fan	1	0
66	2	10	Alluvial Fan	1	1
67	1	7	Alluvial Fan	1	1
68	1	10	Alluvial Fan	1	0

Figure 1-17. Grouping Variables – Part One. .

- The Group variable is 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 with meaningful categories or value ranges 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 "Zone" has 118 observations. If it is selected as the grouping variable, then only variables with the same row index of 118 observations can be used for statistical analysis.

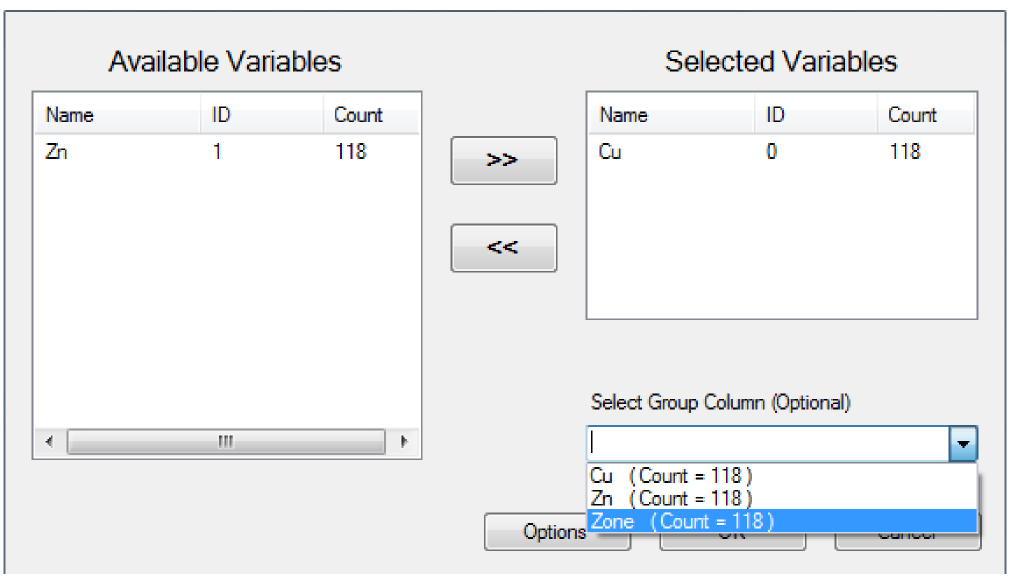

Figure 1-18. Grouping Variables – Part Two.

- As mentioned earlier, one should not assign any columns with missing values (such as a blank data value) for the group variable. If there is a missing value (represented by blanks, strings or dummy values for a group variable, ProUCL 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. For example, if missing values of the grouping variable were assigned the word "blank", all missing values assigned as such would be grouped together.

The **Group Option** is a useful tool for performing 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 example, the same data set may consist of samples from multiple 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.

<u>Notes:</u> Once again, care should be taken to avoid misrepresentation and improper use of group variables. Do not assign any form of a missing value for the group variable.

2 Stats / Sample Sizes

The Stats/Sample Sizes module of ProUCL Contains the General Statistics, Imputed NDs and ROS Methods, as well as the DQO Based Sample Sizes drop down options. This chapter will walk the user through the operation of those three options and give a basic level of understanding to their output. Additionally, most of the information provided in this chapter is also available online in the first of the three ProUCL 2020 webinars, available at:

ProUCL Utilization 2020: Part 1: ProUCL A to Z
https://clu-in.org/conf/tio/ProUCLAtoZ1/

2.1 General Statistics

The **General Statistics** option is available under the **Stats/Sample Sizes** module of ProUCL. This option is used to compute general statistics including simple summary statistics (e.g., mean, standard deviation) for all selected variables. In addition to simple summary statistics, several other statistics such as skewness or %NDs among others can help users to determine which later tests are appropriate, should they wish to run more statistical tests or produce potential estimates such as a UTL or UCL. These can be computed for both full uncensored data sets (**Full w/o NDs**), and for data sets with non-detect (**with NDs**) observations (e.g., estimates based upon the KM method).

Two Menu options: **Full w/o NDs** and **With NDs** are available.

- **Full (w/o NDs)**: This option computes general statistics for all selected variables.
- **With NDs**: This option computes general statistics including the KM method based mean and standard deviations for all selected variables with ND observations.

Each menu option (**Full (w/o NDs)** and **With NDs**) has two sub-menu options:

- Raw Statistics
- Log-Transformed

When computing general statistics for raw data, a message will be displayed for each variable that contains non-numeric values. The **General Statistics** option computes log-transformed (natural log) statistics 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 a selected variable.

2.1.1 General Statistics for Data Sets with or without NDs

Click General Statistics

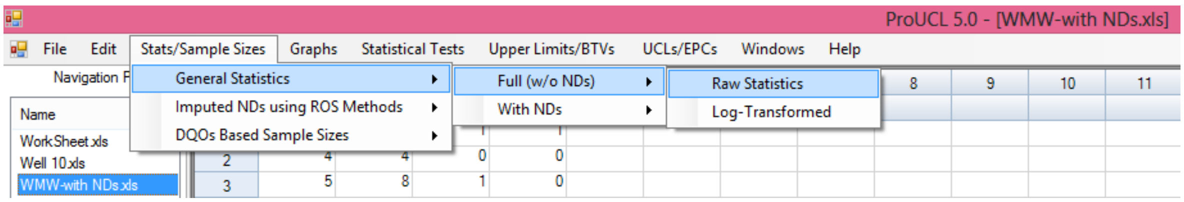

Figure 2-1. Computing General Statistics

- Select either Full (w/o NDs) or With NDs
- Select either Log-Transformed or Raw Statistics option.
- The **Select Variables** screen (see Chapter 1) will appear.
- Select one or more variables from the **Select Variables** screen.

If statistics are to be computed by a Group variable, then select a group variable by clicking the arrow below the **Select Group Column (Optional)** button. This will result in drop-down list of available variables and select a proper group variable.

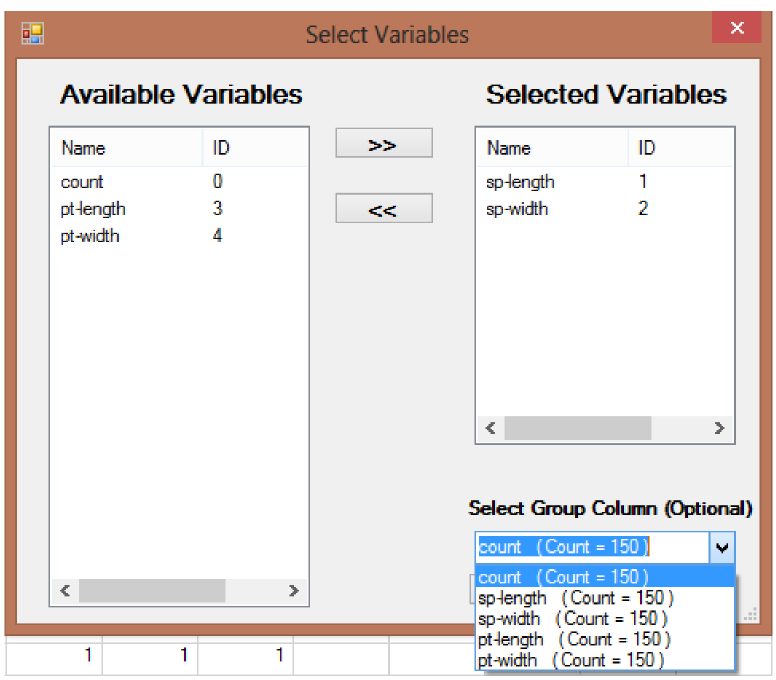

Figure 2-2. Selecting a Grouping Variable

Click on the **OK** button to continue or on the **Cancel** button to cancel the **General Statistics** option.

The Raw or log statistics results will appear similar to the images below. The first two show examples for **Full Datasets (w/o NDs)** while the final shows an example **With NDs**

Table 2-1. Raw Statistics- w/o NDs

User Selected Options	
From File	FULLIRIS-nds.xls
Full Precision	OFF

From File: FULLIRIS-nds.xls

Summary Statistics for Uncensored Data Sets

Variable	NumObs	# Missing	Minimum	Maximum	Mean	SD	SEM	MAD/0.675	Skewness	Kurtosis	CV
sp-length (1)	50	0	4.3	5.8	5.006	0.352	0.0498	0.297	0.12	-0.253	0.0704
sp-length (2)	50	0	4.9	7	5.936	0.516	0.073	0.519	0.105	-0.533	0.087
sp-length (3)	50	0	4.9	7.9	6.588	0.636	0.0899	0.593	0.118	0.0329	0.0965

Percentiles for Uncensored Data Sets

Variable	NumObs	# Missing	10%ile	20%ile	25%ile(Q1)	50%ile(Q2)	75%ile(Q3)	80%ile	90%ile	95%ile	99%ile
sp-length (1)	50	0	4.59	4.7	4.8	5	5.2	5.32	5.41	5.61	5.751
sp-length (2)	50	0	5.38	5.5	5.6	5.9	6.3	6.4	6.7	6.755	6.951
sp-length (3)	50	0	5.8	6.1	6.225	6.5	6.9	7.2	7.61	7.7	7.802

Table 2-2. Log-Transformed Statistics- w/o NDs

User Selected Options	
From File	FULLIRIS-nds.xls
Full Precision	OFF

From File: FULLIRIS-nds.xls

Summary Statistics for Uncensored Log-Transformed Data Sets

Variable	NumObs	# Missing	Minimum	Maximum	Mean	Variance	SD	MAD/0.675	Skewness	Kurtosis	CV
sp-length (1)	50	0	1.459	1.758	1.608	0.00497	0.0705	0.0605	-0.0553	-0.291	0.0438
sp-length (2)	50	0	1.589	1.946	1.777	0.00761	0.0872	0.0873	-0.0852	-0.463	0.0491
sp-length (3)	50	0	1.589	2.067	1.881	0.00943	0.0971	0.0885	-0.196	0.492	0.0516

Percentiles for Uncensored Log-Transformed Data Sets

Variable	NumObs	# Missing	10%ile	20%ile	25%ile(Q1)	50%ile(Q2)	75%ile(Q3)	80%ile	90%ile	95%ile	99%ile
sp-length (1)	50	0	1.524	1.548	1.569	1.609	1.649	1.671	1.688	1.724	1.749
sp-length (2)	50	0	1.683	1.705	1.723	1.775	1.841	1.856	1.902	1.91	1.939
sp-length (3)	50	0	1.758	1.808	1.829	1.872	1.932	1.974	2.029	2.041	2.054

Table 2-3. Raw Statistics – Data Set with NDs

User Selected Options											
From File	Zn-alluvial-fan-data.xls										
Full Precision	OFF										
From File: Zn-alluvial-fan-data.xls											
Summary Statistics for Censored Data Set (with NDs) using Kaplan Meier Method											
Variable	**NumObs**	**# Missing**	**Num Ds**	**NumNDs**	**% NDs**	**Min ND**	**Max ND**	**KM Mean**	**KM Var**	**KM SD**	**KM CV**
Cu (alluvial fan)	65	3	48	17	26.15%	1	20	3.608	13.08	3.616	1.002
Cu (basin trough)	49	1	35	14	28.57%	1	15	4.362	21.64	4.651	1.066
Summary Statistics for Raw Data Sets using Detected Data Only											
Variable	**NumObs**	**# Missing**	**Minimum**	**Maximum**	**Mean**	**Median**	**Var**	**SD**	**MAD/0.675**	**Skewness**	**CV**
Cu (alluvial fan)	48	3	1	20	4.146	2	16.04	4.005	1.483	2.256	0.966
Cu (basin trough)	35	1	1	23	5.229	3	27.18	5.214	2.965	1.878	0.997
Percentiles using all Detects (Ds) and Non-Detects (NDs)											
Variable	**NumObs**	**# Missing**	**10%ile**	**20%ile**	**25%ile(Q1)**	**50%ile(Q2)**	**75%ile(Q3)**	**80%ile**	**90%ile**	**95%ile**	**99%ile**
Cu (alluvial fan)	65	3	1	2	2	3	5	7	10	15.2	20
Cu (basin trough)	49	1	1	2	2	4	8	9.4	12.4	15	20.12

Note:
MAD = Median absolute deviation
MAD/0.675 = Robust and resistant (to outliers) estimate of variability, population standard deviation, σ.

The **General Statistics** screen (and all other output screens generated by other modules) shown above can be saved as an Excel 2003 (.xls) or 2007 (.xlsx) file. Click **Save** from the file menu.

On the output screens shown above, most of the statistics are self-explanatory and described in the ProUCL Technical Guide (EPA 2013, 2015).

2.2 Imputing Non-Detects Using ROS Methods

ROS methods can be used to impute ND observations using a normal, lognormal, or gamma model. The use of this option generates additional columns consisting of all imputed NDs and detected observations. These columns are appended to the existing open spreadsheet file. The user should save the updated file if they want to use the imputed data for their other application(s) such as PCA or discriminant analysis. It is not easy to perform multivariate statistical methods on data sets with NDs. The availability of imputed NDs in a data file helps the advanced users who want to use exploratory methods on data sets with ND observations. Like other statistical methods in ProUCL, NDs can also be imputed by a group variable. An example using lognormal ROS for ND imputation is presented below, however utilizing this tool for Normal or Gamma ROS depending on the distributional form of the user's data have effectively the same workflow.

Note: ROS methods should not be used when the data is highly skewed, contains outliers, or consists of a high percentage of NDs (>50%). For more detailed information on this subject see Section 4.5 of the ProUCL Technical Guide.

Click Imputed NDs using ROS Methods ► Lognormal ROS

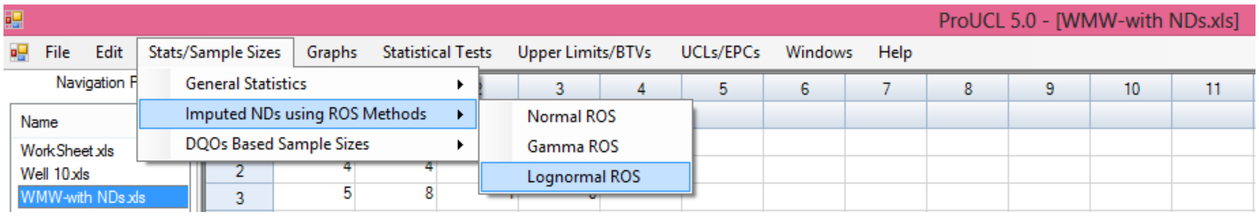

Figure 2-3. Using Lognormal ROS Method for NDs.

The **Select Variables** screen (Section 1.3.1.2) will appear.

Select one or more variable(s) from the **Select Variables** screen; NDs can be imputed using a group variable as shown in the following screen shot.

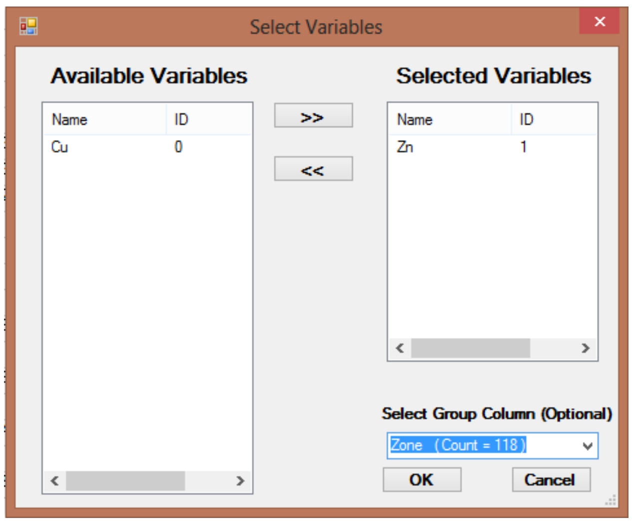

Figure 2-4. Using Grouping Variables to Impute NDs.

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

Table 2-4. Output Screen for ROS Est. NDs (Lognormal ROS) Option

0	1	2	3	4	5	6
Cu	Zn	Zone	D_Cu	D_Zn	LnROS_Zn (alluvial fan)	LnROS_Zn (basin trough)
1	10	Alluvial Fan	0	0	2.12437794466611	20
1	9	Alluvial Fan	0	1	9	10
3		Alluvial Fan	1		1.000000E+031	60
3	5	Alluvial Fan	1	1	5	20
5	18	Alluvial Fan	1	1	18	12
1	10	Alluvial Fan	1	0	2.7045642735474	8
4	12	Alluvial Fan	1	1	12	3.48713118440742
4	10	Alluvial Fan	1	1	10	14
2	11	Alluvial Fan	1	1	11	4.98477186220711
2	11	Alluvial Fan	1	1	11	17
1	19	Alluvial Fan	1	1	19	1.87132713438924
2	8	Alluvial Fan	1	1	8	11
5	3	Alluvial Fan	0	0	2.49463676896719	5
11	10	Alluvial Fan	1	0	3.1603475071042	12
1	10	Alluvial Fan	0	0	3.55892730586941	4
2	10	Alluvial Fan	1	1	10	3
2	10	Alluvial Fan	1	1	10	6
2	10	Alluvial Fan	1	1	10	3
2	10	Alluvial Fan	1	1	10	15
20	10	Alluvial Fan	0	0	3.92469067412296	13
2	10	Alluvial Fan	1	1	10	4
2	10	Alluvial Fan	1	0	4.26969100939485	20
3	10	Alluvial Fan	1	1	10	20
3	10	Alluvial Fan	1	0	4.60094330444612	70
	10	Alluvial Fan		1	10	60
20	10	Alluvial Fan	0	0	4.92298559179133	40
10	10	Alluvial Fan	0	1	10	30
7	10	Alluvial Fan	1	1	10	40
5	20	Alluvial Fan	1	1	20	17

Notes: For grouped data, ProUCL generates a separate column for each group in the data set as shown in the above table. Columns with a similar naming convention are generated for each selected variable and distribution using the ROS option.

2.3 DQO Based Sample Sizes

2.3.1 Sample Sizes Based Upon User Specified Data Quality Objectives (DQOs) and Power Assessment

One of the most frequent problems in the application of statistical theory to practical applications, including environmental projects, is to determine the minimum number of samples needed for sampling of reference/background areas and survey units (e.g., potentially impacted site areas, areas of concern, decision units) to make cost-effective and defensible decisions about the population parameters based upon the sampled discrete data. The sample size determination formulae for estimation of the population mean (or some other parameters) depends upon certain decision parameters including the confidence coefficient, $(1-\alpha)$ and the specified error margin (difference), Δ from the unknown true population mean, μ. Similarly, for hypotheses testing approaches, sample size determination formulae depends upon pre-specified values of the decision parameters selected while describing the data quality objectives (DQOs) associated with an environmental project. The decision parameters 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 acceptable from both human health and cost-effectiveness point of view.

Refer to Figure 2-5 for the relationship between Type I and Type II error. Note that H0 represents the null hypothesis while H1 represents the alternative. Type I error represents the risk of rejecting the null when the null is true, while Type II error represents the risk of not rejecting the null when the null is false. By moving the cut-off value (black vertical line) to the left, the rate of false negative error β (depicted with blue shaded area) can be decreased at the cost of increasing the rate of false positive error α (depicted with red shaded area) or vice versa.

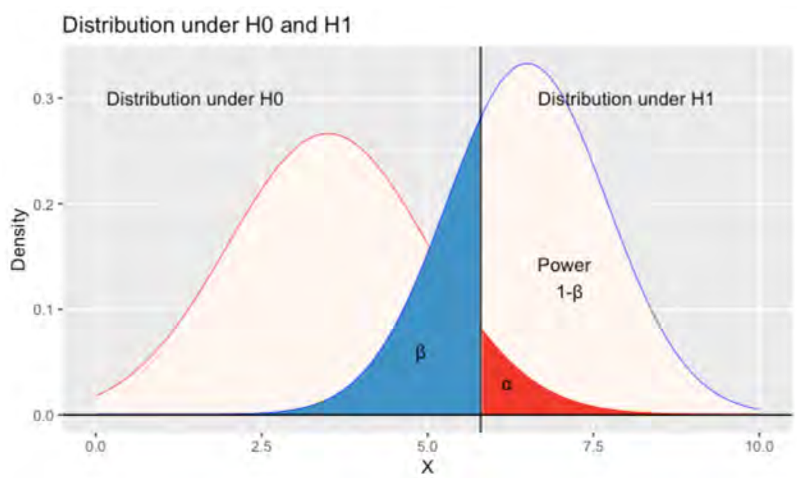

Figure 2-5. The Relationship Between Type I and Type II Error.

Note: Initially, the **Sample Sizes** module was incorporated in ProUCL 4.0/ProUCL 4.1. Not many changes have been made since then except those described below. Therefore, many screenshots generated using an earlier 2010 version of ProUCL have been used in the examples described in this chapter.

Both parametric (assuming normality) and nonparametric (distribution free) sample size determination formulae as described in guidance documents (MARSSIM 2000, EPA 2002c and 2006a) have been incorporated in the ProUCL software. Specifically, the **DQOs Based Sample Sizes** module of ProUCL can be used to determine sample sizes to estimate the mean, perform parametric and nonparametric single-sample and two-sample hypothesis tests, and apply acceptance sampling approaches to address project needs of the various CERCLA and RCRA site projects. The details can be found in Chapter 8 of the ProUCL Technical Guide and in EPA guidance documents (EPA 2006a, 2006b).

The Sample size module in ProUCL can be used at two different stages of a project. Most of the sample size formulae require some estimate of the population standard deviation (variability). Depending upon the project stage, a standard deviation: 1) represents a preliminary estimate of the population (e.g., study area) variability needed to compute the minimum sample size during the planning and design stage; or 2) represents the sample standard deviation computed using the data collected without considering DQOs process which is used to assess the power of the test based upon the collected data. During the power assessment stage, if the computed sample size is larger than the size of already collected data set, it can be inferred that the size of the collected data set is not large enough to achieve the desired power. The formulae to compute the sample sizes during the planning stage and after performing a statistical test are the same except that the estimates of standard deviations are computed/estimated differently.

Planning stage before collecting data: Sample size formulae are commonly used during the planning stage of a project to determine the minimum sample sizes needed to address project objectives (estimation, hypothesis testing) with specified values of the decision parameters (e.g., Type I and II errors, width of gray region). During the planning stage, since the data are not collected *a priori*, a preliminary rough estimate of the population standard deviation (to be expected in sampled data) is obtained from other similar sites, pilot studies, or expert opinions. An estimate of the expected standard deviation along with the specified values of the other decision parameters are used to compute the minimum sample sizes needed to address the project objectives during the sampling planning stage; the project team is expected to collect the number of samples thus obtained. The detailed discussion of the sample size determination approaches during the planning stage can be found in MARSSIM 2000 and U.S. EPA 2006a.

Power assessment stage after performing a statistical method: Often, in practice, environmental samples/data sets are collected without taking the DQOs process into consideration or the observed standard deviation is different than anticipated. Under this scenario, the project team performs statistical tests on the available already collected data set. However, once a statistical test (e.g., WMW test) has been performed, the project team attempts to assess the power associated with the test in retrospect. The user should refer to EPA (2006b) for guidance on a second-stage power analysis. During this process, it will be necessary to re-evaluate assumptions as well as the project objectives to determine if the previous goal for power is still adequate. Once this is done, the practitioner can use the sample size module in ProUCL and the observed sample standard deviation computed based upon the already collected data, to estimate the minimum sample size needed to perform the test and achieve adequate power. The module asks the user to estimate the allowable margin of error as well as the variation. Although it may be tempting to use the observed difference between two sample means, or the observed difference between the sample mean and the screening level, this is not an appropriate second-stage power analysis. It is important that this margin of error is based on what is actually meaningful to the project. This will likely be the same as the margin of error used for the initial sample size calculation, but it may be different if the understanding of the site has fundamentally changed.

- If the computed sample size obtained using the sample variance is less than the size of the already collected data set used to perform the test, it may be determined that the power of the test has been achieved. However, if the sample size of the collected data is less than the minimum sample size computed in retrospect, the user may want to collect additional samples to assure that the test achieves the desired power.

- Frequently, differences in the sample sizes computed in two different stages due to the differences in the values of the estimated variability. Specifically, the preliminary estimate of the variance computed using information from similar sites could be significantly different from the variance computed using the available data already collected from the study area under investigation which will yield different values of the sample size. If during the preliminary sample size estimation, the variation was underestimated compared to what was actually observed in the data, and exactly the recommended number of samples were taken from this preliminary estimate, the second-stage power analysis will indicate additional samples are needed, if no other parameters have changed.

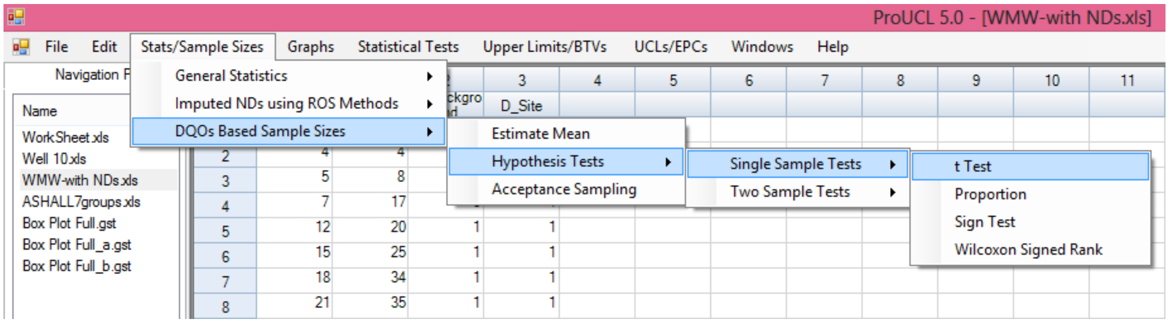

Figure 2-6. Computing Sufficient Sample Size

2.3.2 *Sample Size for Estimation of Mean*

Click **Stats/Sample Sizes► DQOs Based Sample Sizes ► Estimate Mean**

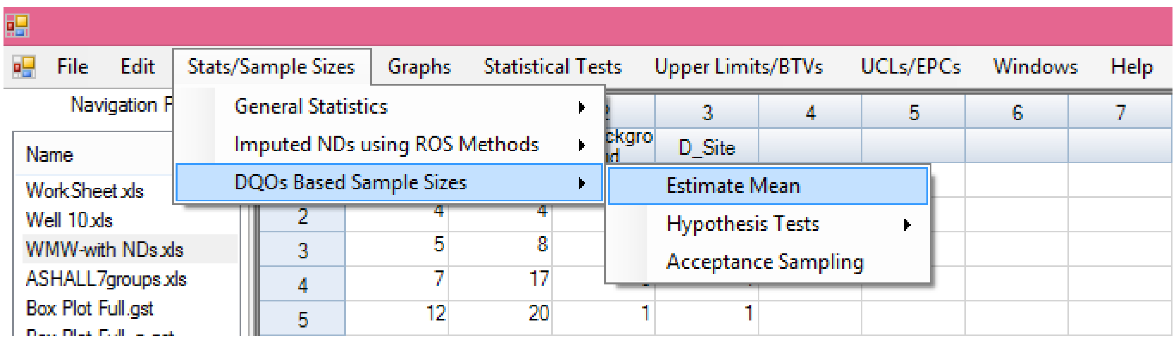

Figure 2-7. Computing Sufficient Sample Size for Estimating the Mean.

The following options window is shown.

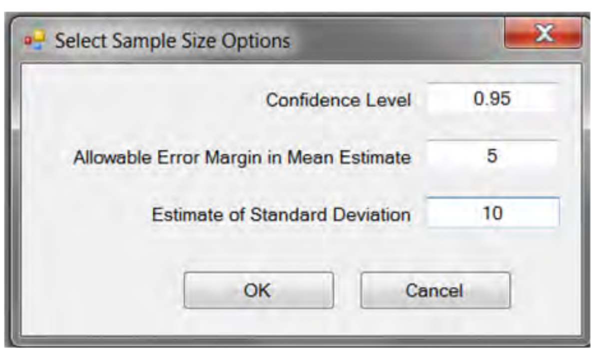

Figure 2-8. Options Related to Computing Sufficient Sample Size for Estimating the Mean.

- Specify the **Confidence Level**. Default is **0.95**.
- Specify the Estimate of standard deviation.
- Specify the Allowable Error Margin in Mean Estimate.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-5. 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	

2.3.3 Sample Sizes for Single-Sample Hypothesis Tests

2.3.3.1 Sample Size for Single-Sample t-Test

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► t Test

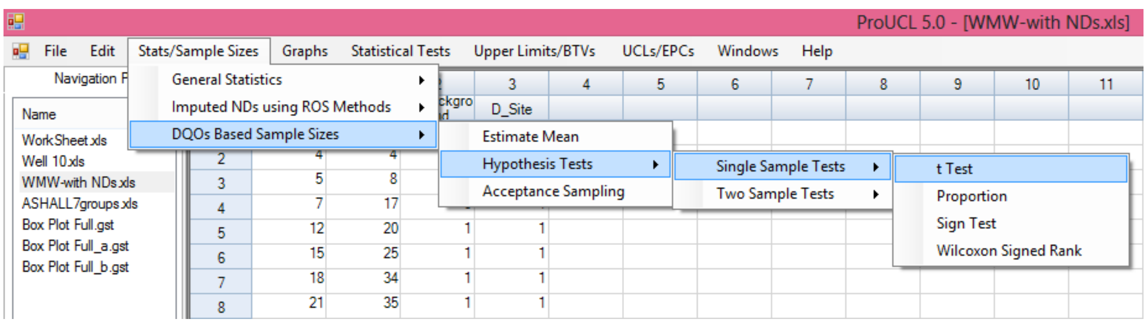

Figure 2-9. Computing Sufficient Sample Size for a Single-Sample t-Test

The following options window is shown.

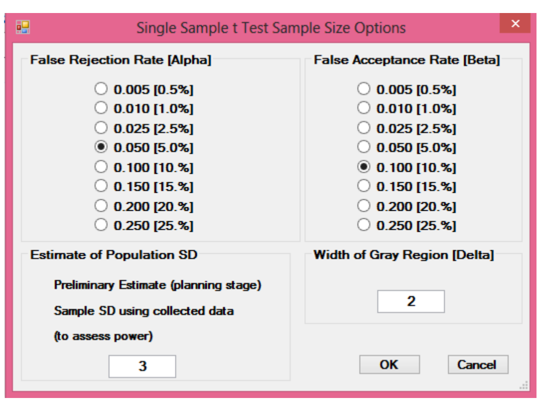

Figure 2-10. Options Related to Computing Sufficient Sample Size for a Single-Sample t-Test.

- Specify the False Rejection Rate (Alpha, α). Default is 0.05.

- Specify the False Acceptance Rate (Beta, β). Default is 0.1.
- Specify the Estimate of population standard deviation (SD). Default is 3.
- Specify the Width of the Gray Region (Delta, Δ). Default is 2.

Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-6. Output Screen for Sample Sizes for Single-Sample t-Test (α = 0.05, β = 0.2, sd = 10.41, Δ = 10) Example from EPA 2006a (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						

2.3.3.2 Sample Size for Single-Sample Proportion Test

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► Proportion

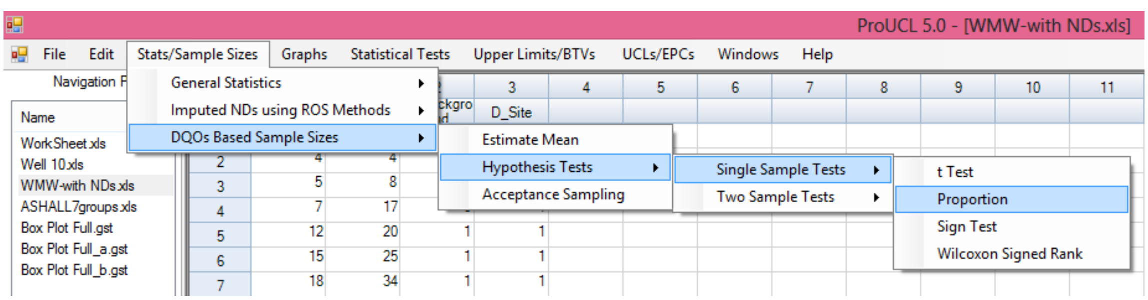

Figure 2-11. Computing Sufficient Sample Size for Single-Sample Proportion Test

The following options window is shown.

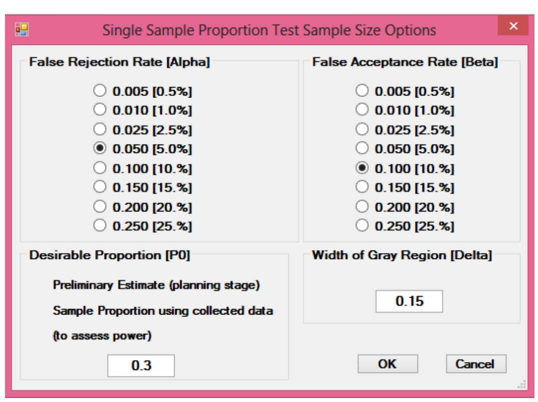

Figure 2-12. Options Related to Computing Sufficient Sample Size for Single-Sample Proportion Test.

- Specify the False Rejection Rate (Alpha, α). Default is 0.05.
- Specify the False Acceptance Rate (Beta, β). Default is 0.1.
- Specify the Desirable Proportion (P0). Default is 0.3.
- Specify the Width of the Gray Region (Delta, Δ). Default is 0.15.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-7. Output Screen for Sample Size for Single-Sample Proportion Test (α = 0.05, β = 0.2, P0 = 0.2, Δ = 0.05) Example from EPA 2006a (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)

2.3.3.3 Sample Size for Single-Sample Sign Test

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► Sign Test

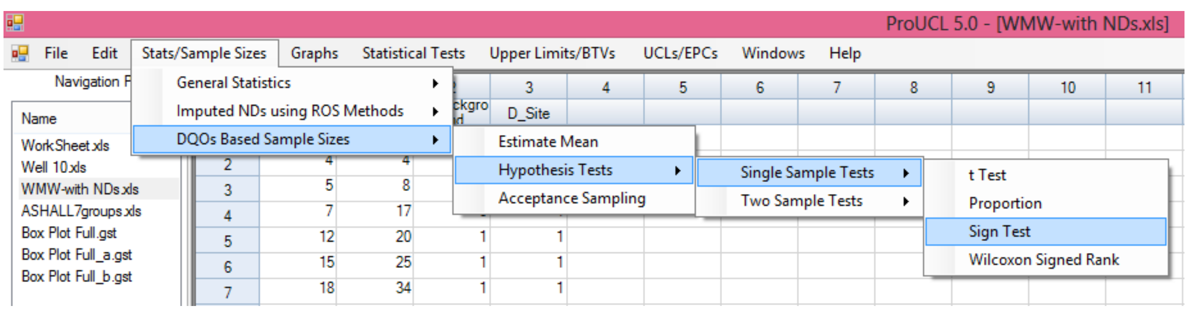

Figure 2-13. Computing Sufficient Sample Size for Single-Sample Sign Test.

The following options window is shown.

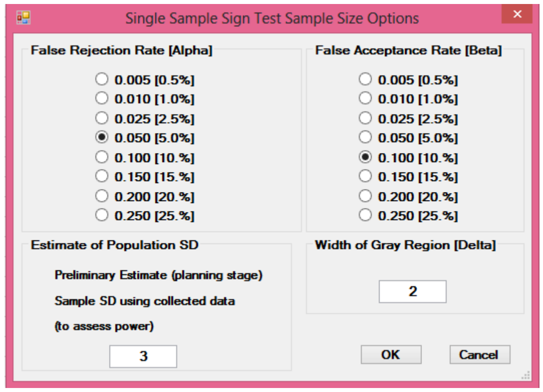

Figure 2-14. Options Related to Computing Sample Size for Single-Sample Sign Test.

- Specify the False Rejection Rate (Alpha, α). Default is 0.05.
- Specify the False Acceptance Rate (Beta, β). Default is 0.1.
- Specify the Width of the Gray Region (Delta, Δ). Default is 2.
- Specify the Estimate of standard deviation. Default is 3.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-8. 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					

2.3.3.4 Sample Size for Single-Sample Wilcoxon Signed Rank Test

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► Wilcoxon Signed Rank

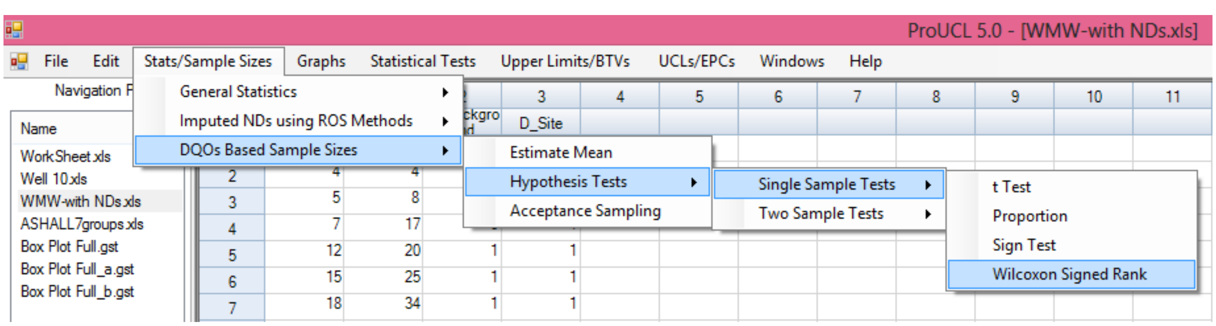

Figure 2-15. Computing Sufficient Sample Size for Single-Sample Wilcoxon Signed Rank Test.

The following options window is shown.

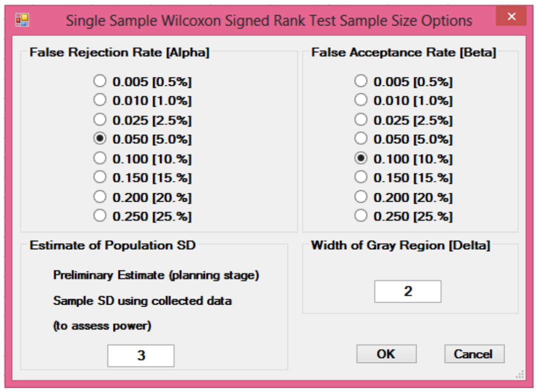

Figure 2-16. Options Related to Computing Sufficient Sample Size for Single-Sample Wilcoxon Signed Rank Test.

- Specify the False Rejection Rate (Alpha, α). Default is 0.05.
- Specify the False Acceptance Rate (Beta, β). 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, Δ). Default is 2.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-9. Output Screen for Sample Sizes for Single-Sample WSR Test (α = 0.1, β = 0.2, sd = 130, Δ = 100) Example from EPA 2006a (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

2.3.4 Sample Sizes for Two-Sample Hypothesis Tests

2.3.4.1 Sample Size for Two-Sample t-Test

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Hypothesis Tests► Two Sample Tests► t Test

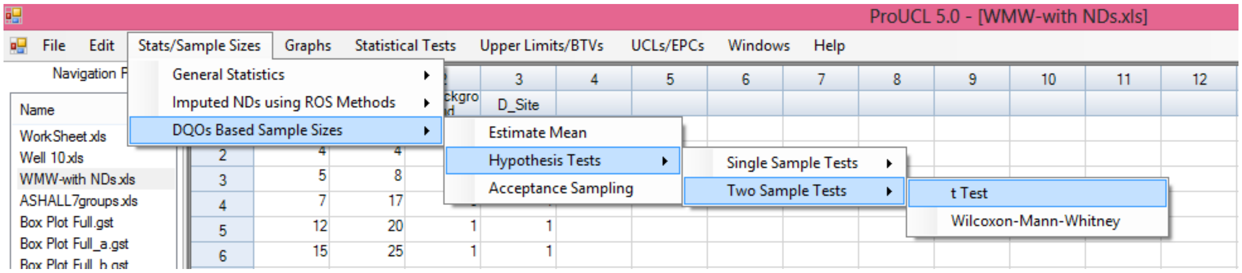

Figure 2-17. Computing Sufficient Sample Size for Two-Sample t-Test.

The following options window is shown.

Figure 2-18. Options Related to Computing Sufficient Sample Size for Two-Sample t-Test.

- Specify the False Rejection Rate (Alpha, α). Default is 0.05.
- Specify the False Acceptance Rate (Beta, β). Default is 0.1.
- Specify the Estimate of standard deviation. Default is 3
- Specify the Width of the Gray Region (Delta, Δ). Default is 2.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-10. Output Screen for Sample Sizes for Two-Sample t-Test (α = 0.05, β = 0.2, sd = 1.467, Δ = 2.5) example from EPA 2006a (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

The sample sizes shown apply to each of the two samples from the two populations used in the hypothesis test.

2.3.4.2 Sample Size for Two-Sample Wilcoxon Mann-Whitney Test

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Hypothesis Tests► Two Sample Tests► Wilcoxon-Mann-Whitney

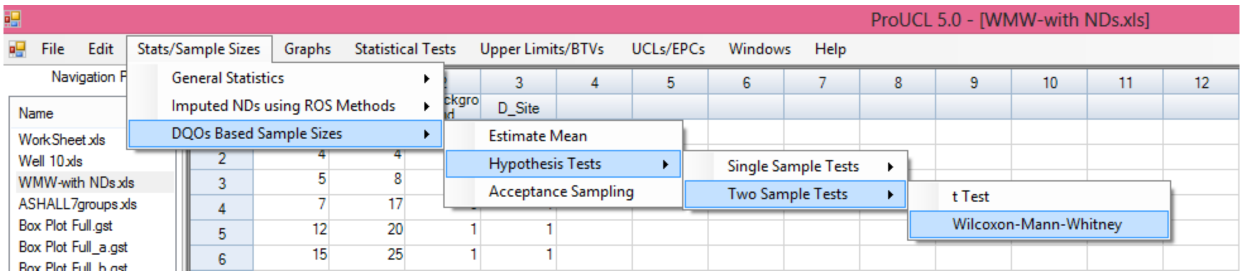

Figure 2-19. Computing Sufficient Sample Size for Two-Sample Wilcoxon Mann-Whitney Test.

The following options window is shown.

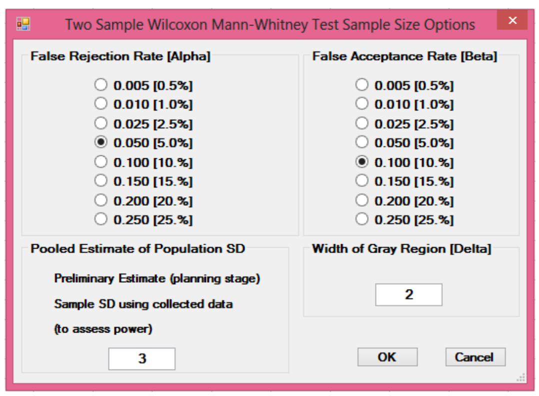

Figure 2-20. Options Related to Computing Sufficient Sample Size for Two-Sample Wilcoxon Mann-Whitney Test.

- Specify the False Rejection Rate (Alpha, α). Default is 0.05.
- Specify the False Acceptance Rate (Beta, β). 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, Δ). Default is 2.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-11. 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					

The sample sizes shown apply to each of the two samples from the two populations used in the hypothesis test.

2.3.4.3 Sample Sizes for Acceptance Sampling

Stats/Sample Sizes ► DQOs Based Sample Sizes ► Acceptance Sampling

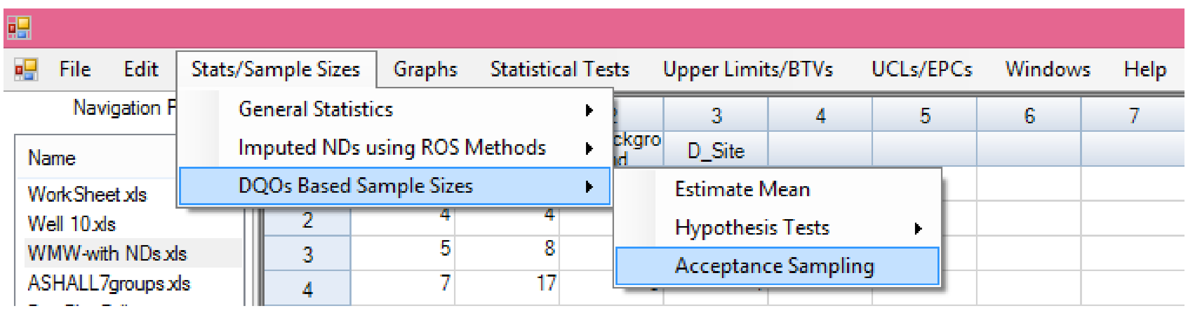

Figure 2-21. Computing Sufficient Sample Size for Acceptance Sampling.

The following options window is shown.

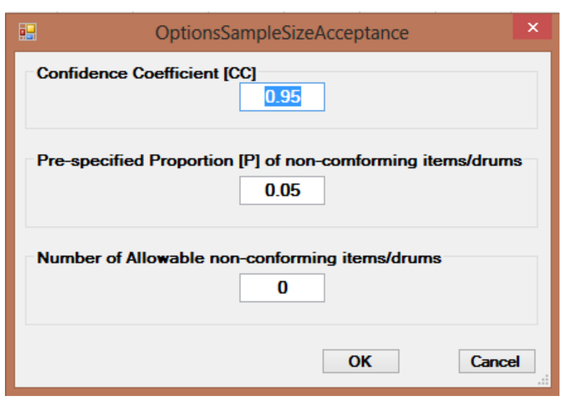

Figure 2-22. Options Related to Computing Sufficient Sample Size for Acceptance Sampling.

- Specify the Confidence Coefficient. Default is 0.95.

- Specify the Proportion [P] of non-conforming items/drums. Default is 0.05.
- Specify the Number of Allowable non-conforming items/drums. Default is 0.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Table 2-12. 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

3 Graphical Methods (Graphs)

The graphical methods described here are used as exploratory tools to get some idea about data distributions (e.g., skewed, symmetric), potential outliers and/or multiple populations present in a data set. The following graphical methods are available under the **Graphs** option of ProUCL 5.2. Additionally, these graphical methods are described in detail in the first ProUCL 2020 webinar located here ProUCL Utilization 2020: Part 1: ProUCL A to Z.

3.1 Handling Non-detects

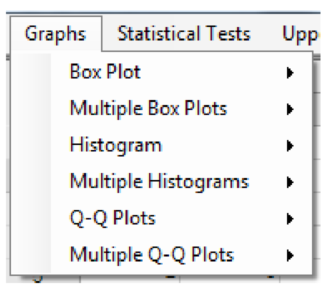

Figure 3-1. Graphical Options.

All graphical displays listed above can be generated using uncensored full data sets (**Full w/o NDs**) as well as left-censored data sets with non-detect (**With NDs**) observations. For the histogram and QQ plot options these three choices of how to display those non-detects are available

- **Use Reported Detection Limit**: Selection of this option treats DLs/RLs as detected values associated with the ND values. The graphs are generated using the numerical values of detection limits and statistics displayed on Q-Q plots are computed accordingly.
- **Use Detection Limit Divided by 2.0**: Selection of this option replaces the DLs with their half values. All Q-Q plots and histograms are generated using the half detection limits and detected values. The statistics displayed on Q-Q plots are computed accordingly.

- **Do not Display Non-detects**: Selection of this option excludes all NDs from a graphical method (Q-Q plots and histograms) and plots only detected values. The statistics shown on Q-Q plots are computed only using the detected data.

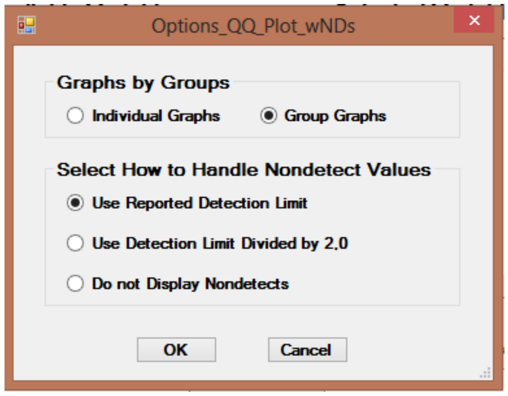

Figure 3-2. Options Related to Q-Q Plots with NDs.

3.2 Making Changes in Graphs using the Toolbar

One can use the toolbar to make changes in a graph generated by ProUCL. The toolbar can be activated by right clicking the mouse on your graph and selecting "Toolbar". The context menu on the box plot shown below appears. By using the context menu, one can change color, title, font size, legend box and label points. For example, one can add the title by clicking title in the context menu. These are typical windows operations which can also be used in ProUCL. The menu applicable to each graph element is activated by right-clicking to the element (e.g. to box plot, title). These operations are illustrated by several screen captures displayed as follows.

Note: Options that affect the computation of statistics displayed on a graph do not adjust the data displayed and as such can yield incorrect results. For example, changing scales along the x-axis or y-axis (e.g., to log scale) will not automatically display statistics in the changed log- scale.

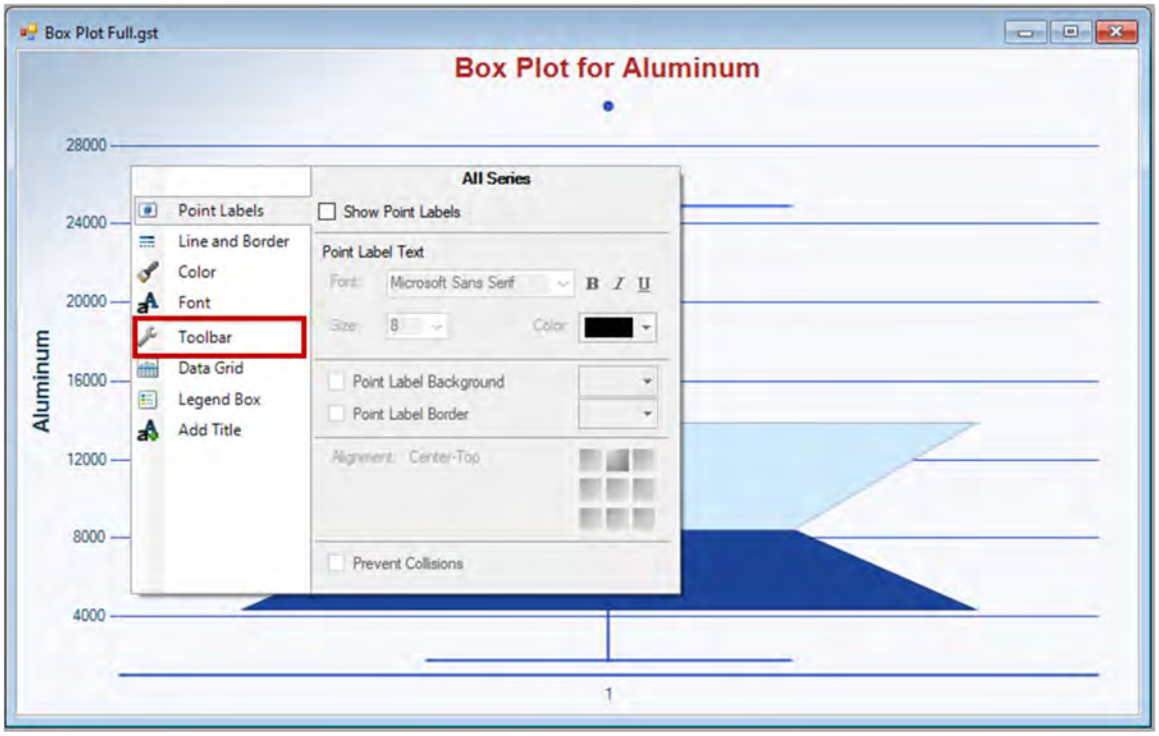

Figure 3-3. Activating the Graphs Toolbar.

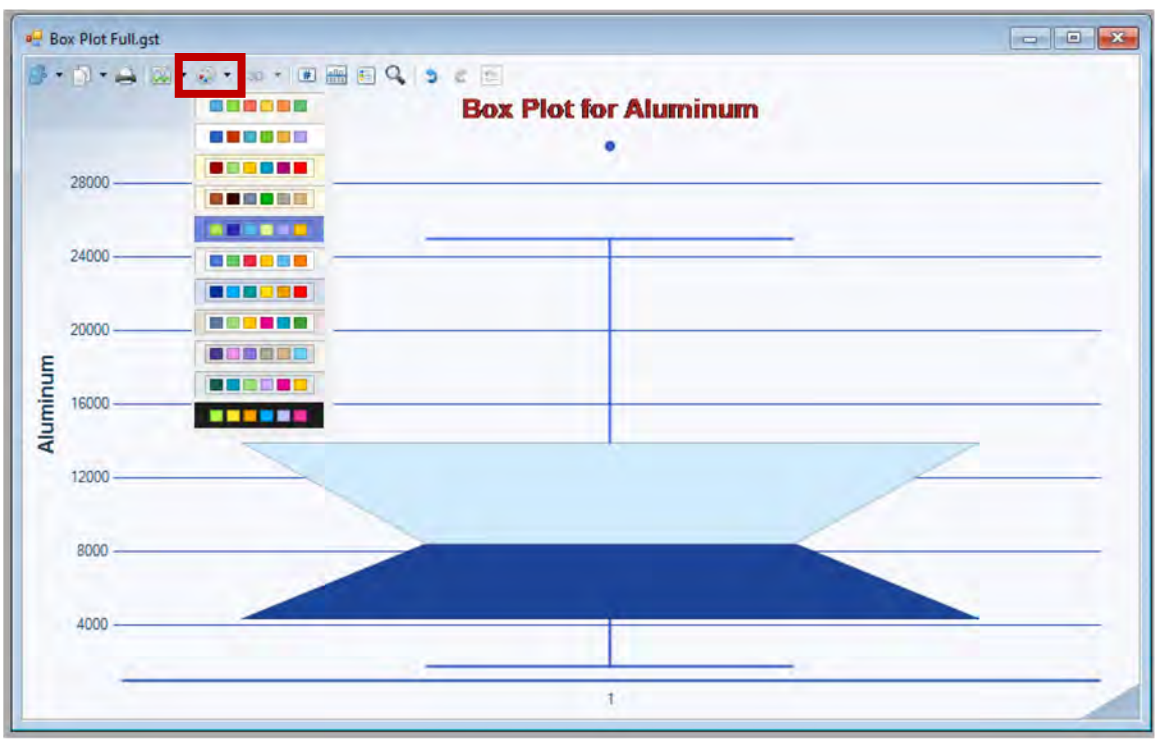

Figure 3-4. Changing the Color of the Graph.

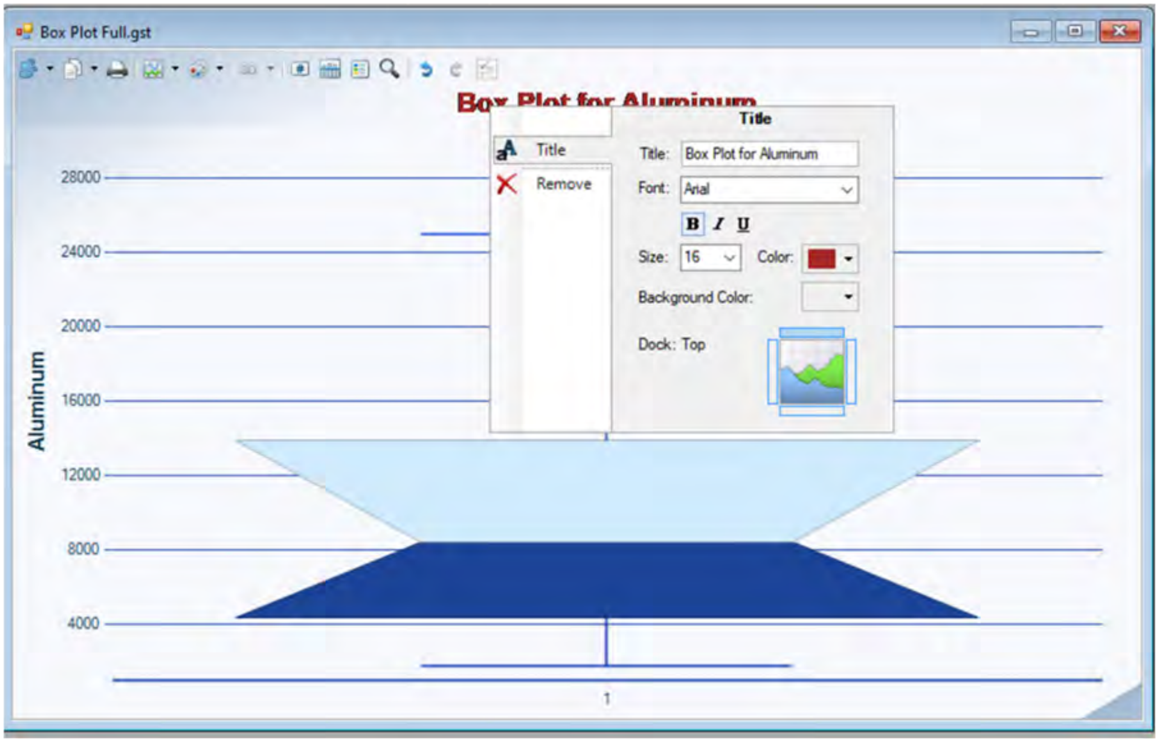

Figure 3-5. Changing the Title of the Graph.

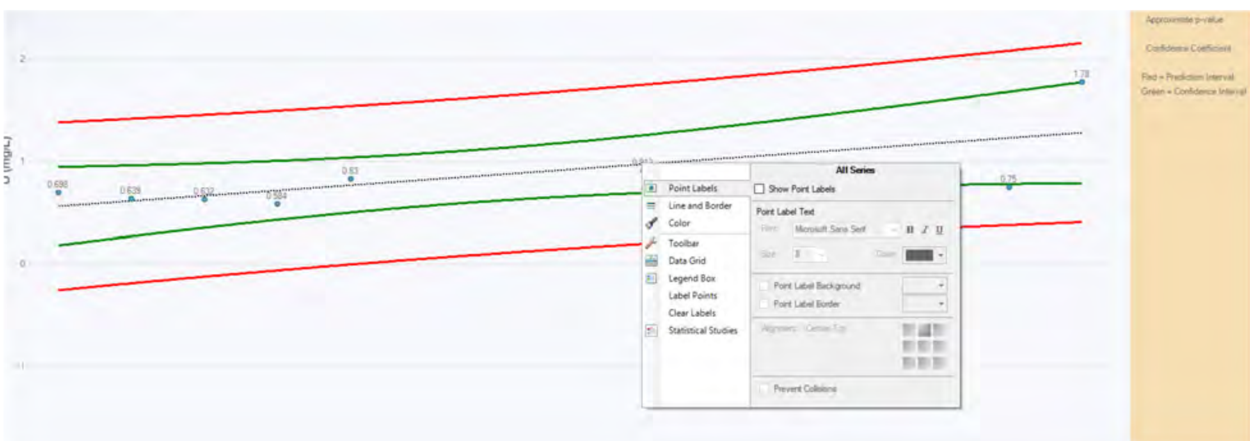

Figure 3-6. Label Points / Clear Labels to show or hide data labels.
Right click just above the point

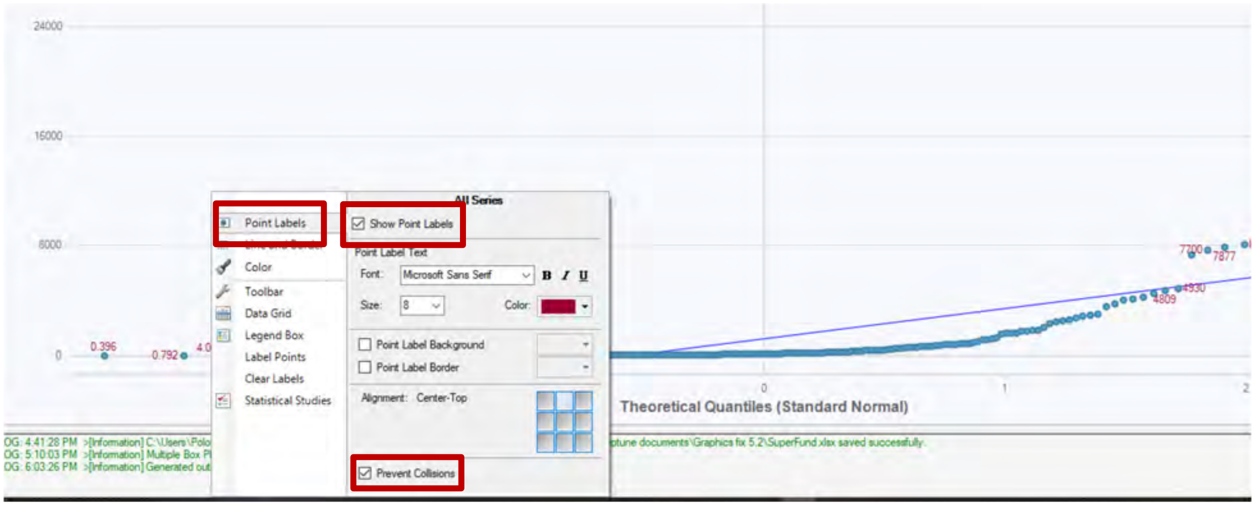

Figure 3-7. If labels overlap click on Point Labels
checkmark Show Point Labels and Prevent Collision

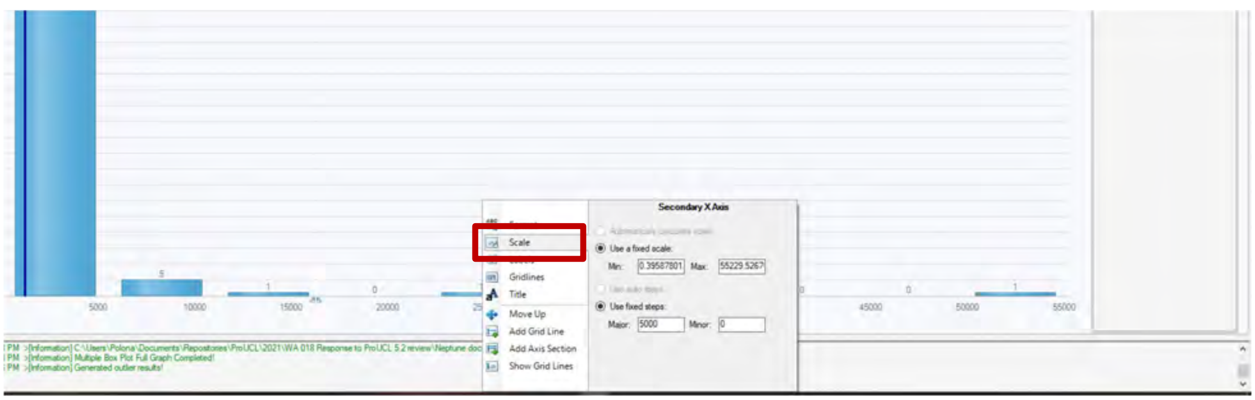

Figure 3-8. Right click the desired axis to modify scale and display more or fewer number labels

3.3 Box Plots

<u>Box Plot (Box and Whiskers Plot):</u> A box plot (box and whiskers plot) represents a convenient exploratory tool and provides a quick five-point summary of a data set. In statistical literature, one can find several ways to generate box plots. The practitioners may have their own preferences to use one method over the other. Box plots are well documented in the statistical literature and a description of differing methodology for box plots can be easily obtained online. Therefore, only the description of the methodology employed in ProUCL is provided below.

All box plot methods including the one in ProUCL represent five-point summary graphs including: the lowest and the highest data values, median (50^{th} percentile=second quartile, Q2), 25^{th} percentile (lower quartile, Q1), and 75^{th} percentile (upper quartile, Q3). A box and whisker plot also provides information about the degree of dispersion (interquartile range (IQR) = Q3-Q1=length/height of the box in a box plot), the degree of skewness (suggested by the length of the whiskers) and unusual data values known as outliers. Specifically, ProUCL (and various other software packages) use the following to generate a box and whisker plot.

- Q1= 25th percentile, Q2= 50th (median), and Q3 = 75th percentile
- Interquartile range= IQR = Q3-Q1 (the height of the box in a box plot)
- Lower whisker starts at Q1 and the upper whisker starts at Q3.
- Lower whisker extends up to the lowest observation or (Q1 - 1.5 * IQR) whichever is higher
- Upper whisker extends up to the highest observation or (Q3 + 1.5 * IQR) whichever is lower
- Horizontal bars (also known as fences) are drawn at the end of whiskers
- Observations lying outside the fences (above the upper bar and below the lower bar) represent potential outliers

ProUCL uses a couple of development tools such as FarPoint spread (for Excel type input and output operations) and ChartFx (for graphical displays). ProUCL generates box plots using the built-in box plot feature in ChartFx. The programmer has no control over computing the statistics (e.g., Q1, Q2, Q3, IQR) using ChartFx. Box plots generated by ProUCL can slightly differ from box plots generated by other programs (e.g., Excel). However, for all practical and exploratory purposes, box plots in ProUCL are equally good compared to those available in the various commercial software packages for exploring data distribution (skewed or symmetric), identifying outliers, and comparing multiple groups.

Note: When producing a box plot using non-detect data a red horizontal line will be added to the graph at the highest non-detect.

Click Graphs ► Box Plot

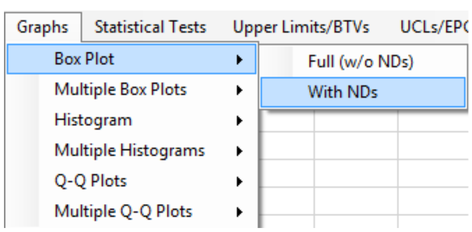

Figure 3-6. Producing Box Plots.

The **Select Variables** screen (Section 1.3.1.2) will appear.

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

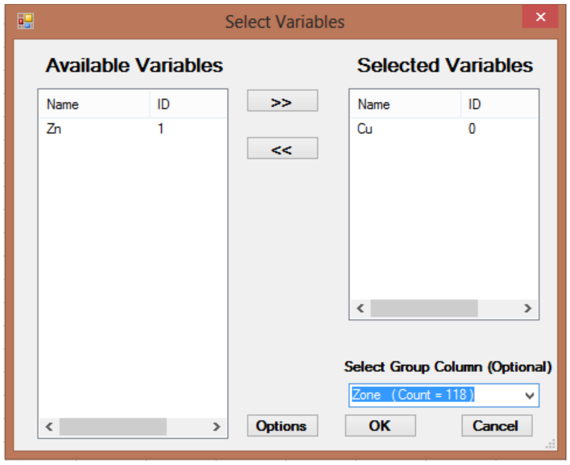

Figure 3-7. Selecting Variables.

The default option for **Graph by Groups** is **Group Graphs**. This option produces side-by- side box plots for all groups included in the selected Group ID Column (e.g., Zone here). The **Group Graphs** option is used when multiple graphs categorized by a group variable need to be produced on the same graph. The **Individual Graphs** option generates individual graphs for each selected variable or one box plot for each group for the variable categorized by a Group ID column (variable).

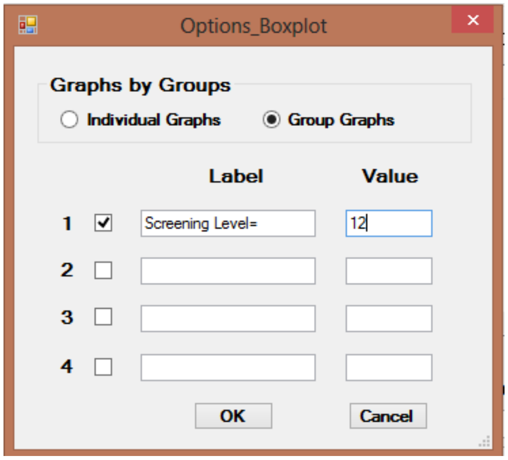

Figure 3-8. Options Related to Producing Box plots.

- While generating box plots, one can display horizontal lines at specified screening levels or a BTV estimate (e.g., UTL95-95) computed using a background data set. A line is added by checking the numbered box, the label for the line is entered in the “label” space, and the value where the line is placed is entered in the “value” space. For data sets with NDs, a horizontal line is also displayed at the largest reported DL associated with a ND value. The use of this option may provide information about the analytical methods used to analyze field samples.
- Click on the **OK** button to continue or on the **Cancel** button to cancel the Box Plot (or other selected graphical) option.

- By clicking anywhere on the graph, a text box will appear that includes the first quartile, median, and third quartile values.

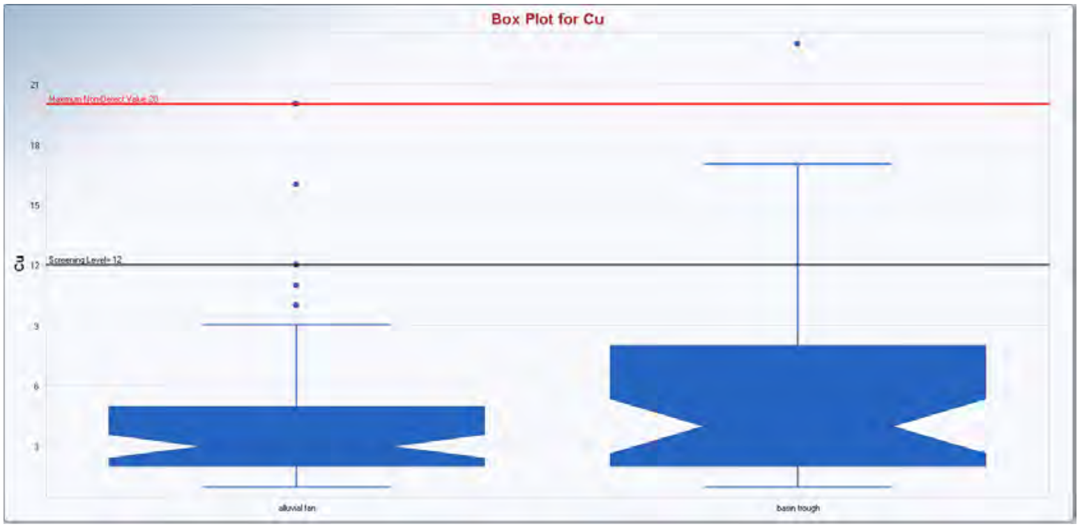

Figure 3-9. Box Plot Output Screen (Group Graph) Selected options: Label (Screening Level), Value (12)

3.4 Multiple Box Plots

Within ProUCL, box plots can also be produced as multiple box plots. To do so simply select the multiple box plots option from the Graphs drop down menu. Then select your variables and groups in the same manner described for single box plots.

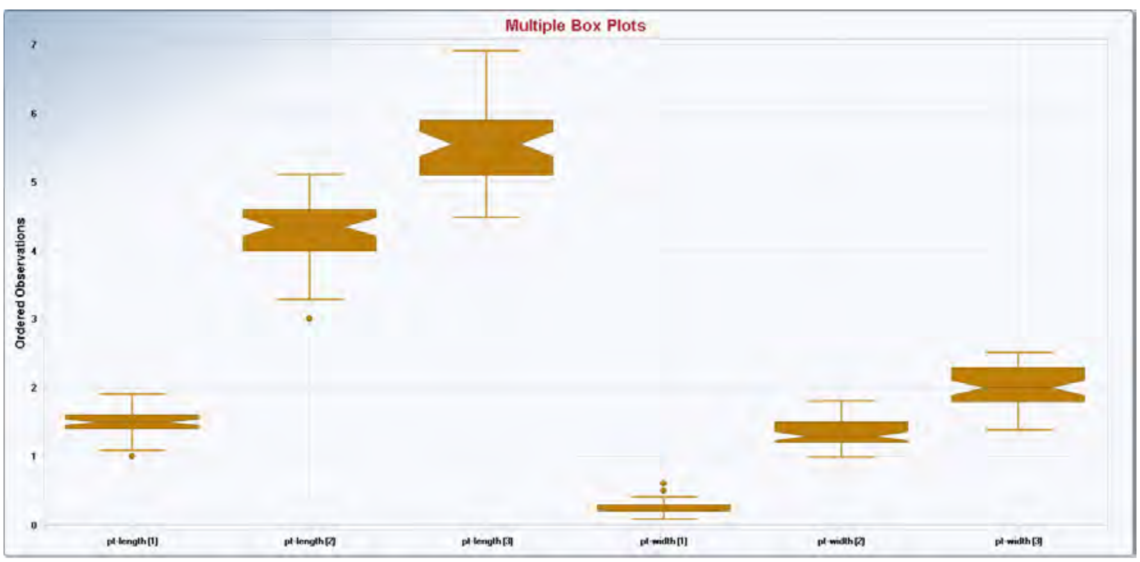

Figure 3-10. Output Screen for Multiple Box Plots (Full w/o NDs) Selected options: Group Graph

3.5 Histograms

Click Graphs ► Histogram

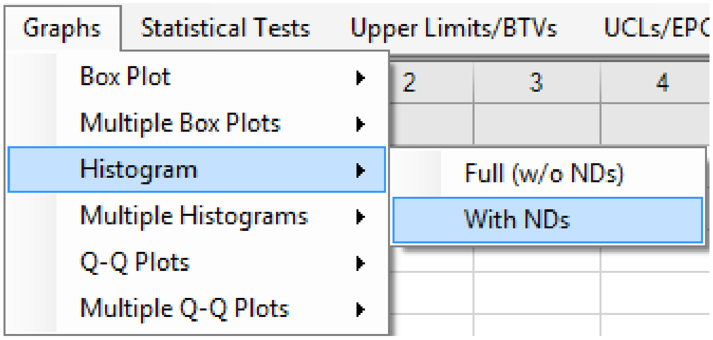

Figure 3-11. Producing Histograms.

- The **Select** Variables screen (Section 1.3.1.2) 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 **Select Group Column (Optional)** button. This will result in a drop-down list of available variables. The user should select an appropriate variable representing a group variable as shown below.
- When the option button is clicked for data sets with NDs, the following window will be shown. By default, histograms are generating using the RLs for NDs.

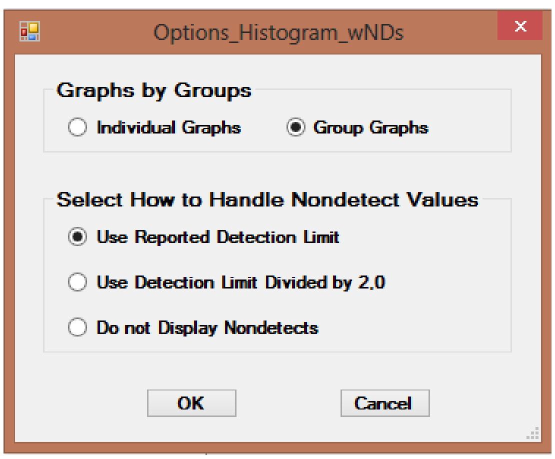

Figure 3-12. Options Related to Producing Histograms with NDs.

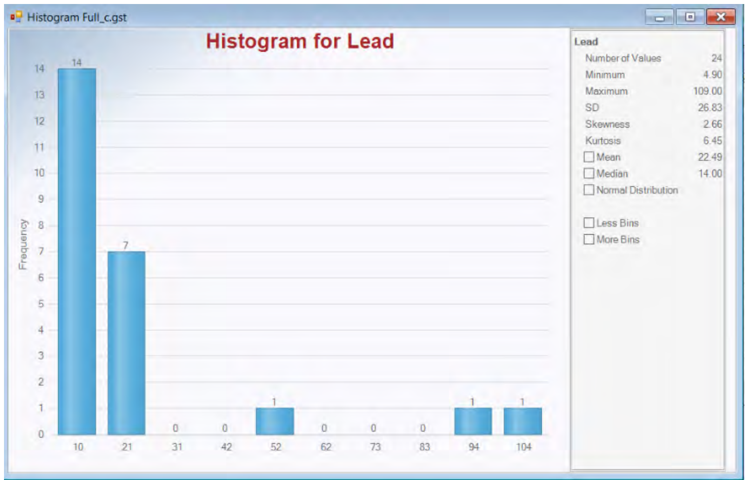

Figure 3-13. Histogram Output.

- After producing a histogram, the user can adjust the number of bins, display a normal distribution curve, and show the mean and median on the histogram using the check boxes on the right side of the graph.

Figure 3-14. Histogram Output with Additional Options.

- The default selection for histograms (and for all other graphs) by a group variable is **Group Graphs**. This option produces multiple histograms on the same graph. If histograms are needed to be displayed individually, the user should check the radio button next to **Individual Graphs**.
- Click on the **OK** button to continue or on the **Cancel** button to cancel the histogram (or other selected graphical) option.

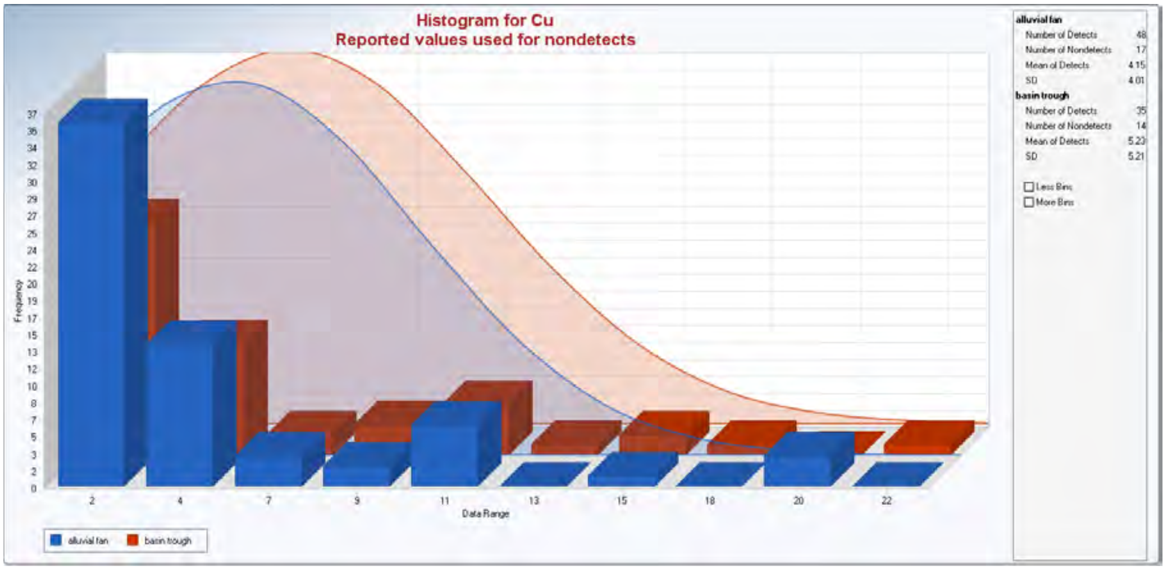

Figure 3-15. Histogram Output Screen Selected options: Group Graphs

Note: ProUCL does not perform any GOF tests when generating histograms. Histograms are generated using the development software ChartFx and not many options are available to alter the histograms. The labeling along the x-axis is done by the development software and it is less than perfect. However, if one hovers the mouse on a bar, relevant statistics (e.g., begin point, midpoint, and end point) about the bar will appear on the screen. The **Histogram** option automatically generates a normal probability density function (pdf) curve irrespective of the data distribution. At this time, ProUCL does not display a pdf curve for any other distribution (e.g., gamma) on a histogram. The user can increase or decrease the number of bins to be used in a histogram.

3.6 Q-Q Plots

Click Graphs ► Q-Q Plots.

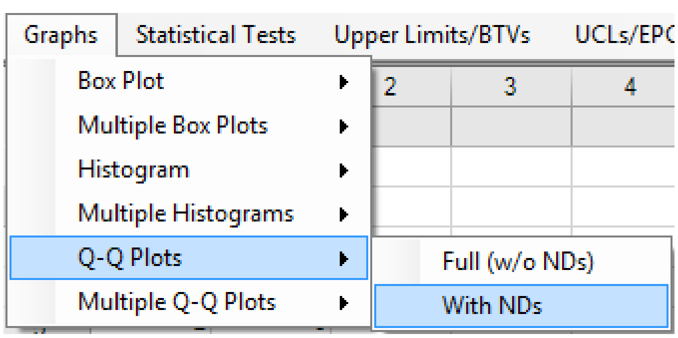

Figure 3-16. Producing Q-Q Plots.

- Select either **Full (w/o NDs)** or **With NDs** option.
- The **Select Variables** screen (Section 1.3.1.2) 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 **Select Group Column (Optional)** 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 as shown below.
- Click on the **OK** button to continue or on the **Cancel** button to cancel the selected Q-Q plots option. The following options screen appears providing choices on how to treat NDs. The default option is to use the reported values for all NDs.

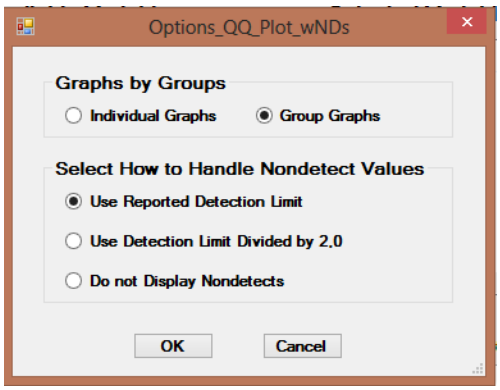

Figure 3-17. Options Related to Producing Q-Q Plots.

- Click on the OK button to continue or on the Cancel button to cancel the selected Q-Q plots option. The following Q-Q plot appears when used on the copper concentrations of two zones: Alluvial Fan and Basin Trough.

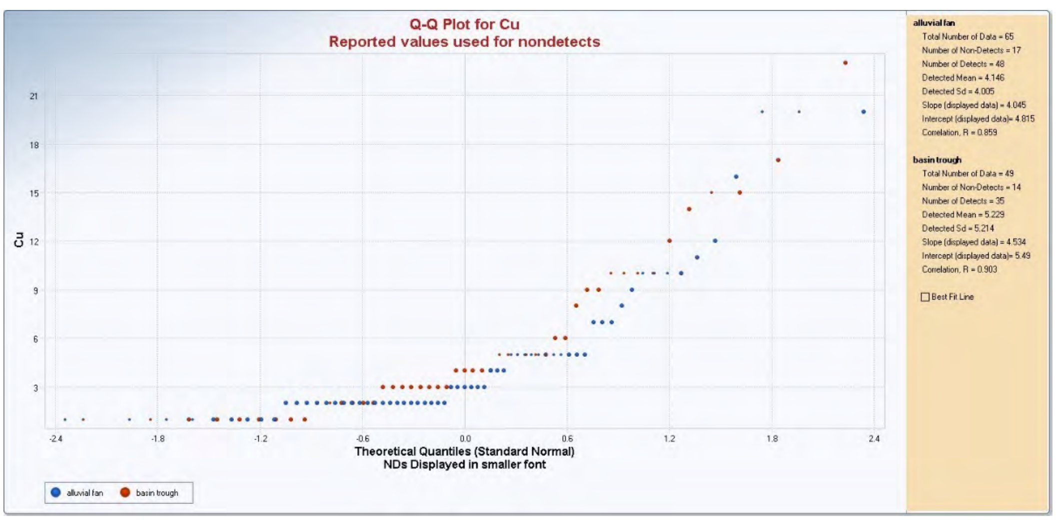

Figure 3-18. Output Screen for Q-Q plots (With NDs) Selected options: Group Graph, No Best Fit Line

<u>Note:</u> The font size of dots representing ND values is smaller than those of the detected values.

3.7 Multiple Q-Q Plots

Similar to box plots, multiple Q-Q plots can be produced in ProUCL. Simply select multiple Q-Q plot from the Graphs dropdown and select repeat the steps from the single Q-Q plot process with the desired variables.

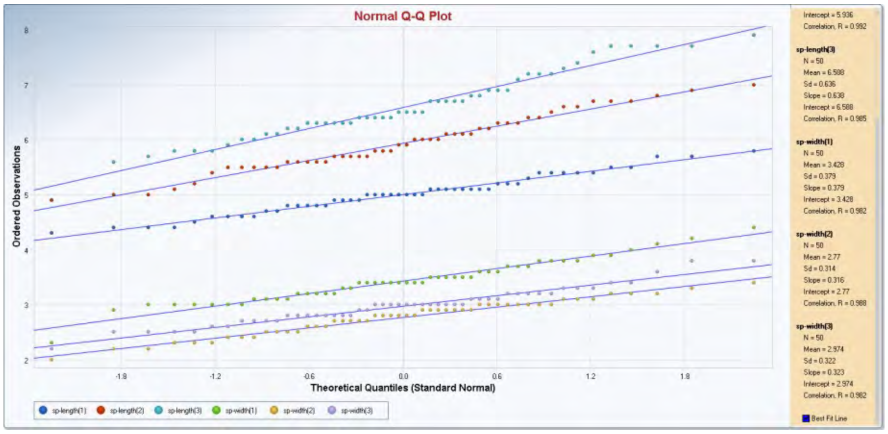

Figure 3-19. Output Screen for Multiple Q-Q Plots (Full w/o NDs) Selected Options: Group Graph, Best Fit Line

If the user does not want the regression lines shown above, click the toggle for the **Best Fit Line** and all regression lines will disappear as shown below.

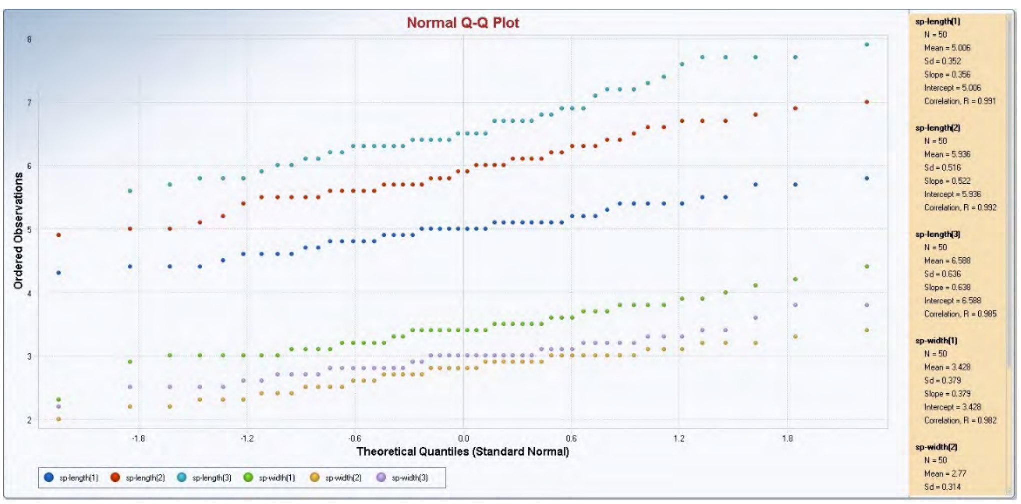

Figure 3-20. Output Screen for Multiple Q-Q Plots (Full w/o NDs) Selected Options: Group Graph

Notes: For Q-Q plots and Multiple Q-Q plots 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 the **Statistical Tests ► Goodness-of-Fit Tests** option.

3.8 Gallery

On any graph, the user can access the gallery by right-clicking on the graph and selecting Toolbar. A Toolbar will appear; the gallery is accessed by selecting the button between the print icon and the color palette icon. The gallery includes several options that can be performed on the current data selection. For example, if the user has produced a histogram with the current data set, they may produce a box plot from the same data by using the gallery and selecting Box Whiskers, or they can do so from the Graphs menu and re-selecting the desired data.

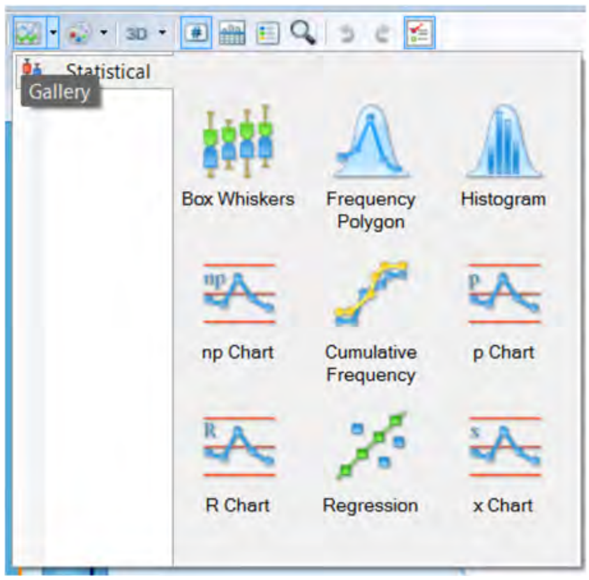

Figure 3-21. The Gallery.

4 Statistical Tests

This section is thoroughly covered in two of the ProUCL 2020 webinar presentations. All of the outlier tests, as well as hypothesis testing, and goodness of fit are discussed in training recording ProUCL Utilization 2020: Part 1: ProUCL A to Z. Trend analysis in ProUCL is presented in training recording ProUCL Utilization 2020: Part 2: Trend Analysis.

4.1 Outlier Tests

Since environmental data tend to be right-skewed, extreme values often occur in data sets originating from environmental applications. A datapoint is not necessarily an outlier just because it is greatly larger or smaller in magnitude than anticipated. When an outlier is identified using statistical test, the best practice is to first scientifically investigate extreme values in the context of site processes, geology and historical use, and based on this information decide whether there is a reason to discard the data. One may also conduct the planned analysis with and without the datapoint in question, as this can lead to better understanding of sub-populations that may be present within a site, such as hot spots. Another important step is to carefully document the reasoning and statistical methods used for treatment of outliers

Two classical outlier tests, Dixon's and Rosner's tests (EPA 2006a; Gilbert 1987), are available in ProUCL 4.0 and later. These tests can be used on data sets with and without ND observations. These tests require the assumption of normality of the data set without the outliers. However, this is very often not the case since environmental data tend to be right-skewed, either naturally or due to subsampling error. 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 requiring further investigations. Therefore, for data sets with NDs, two options are available in ProUCL to deal with data sets with outliers. These options are: 1) exclude NDs 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 processes. To compute the various statistics of interest, ProUCL uses statistical methods suited for left-censored data sets with multiple DLs.

It is suggested that the outlier identification procedures be supplemented with graphical displays such as normal Q-Q plots and box plots. Also, significant and obvious jumps and breaks in a normal Q-Q plot can be indications of the presence of more than one population and/or data gaps due to lack of enough data points (data sets of smaller sizes). Data sets of large sizes (e.g., >100) exhibiting such behavior on Q-Q plots may need to be partitioned out into component sub-populations before estimating EPCs or BTVs.

Outlier tests in ProUCL are available under the **Statistical Tests** module.

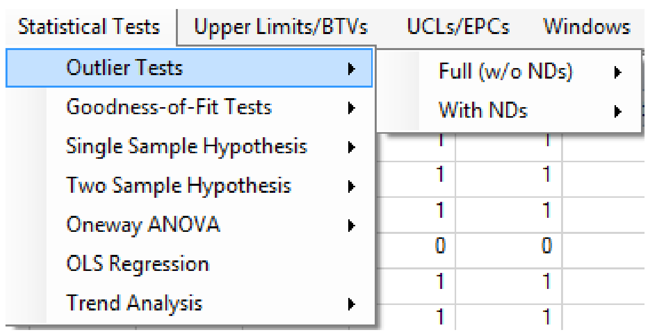

Figure 4-1. Performing Outlier Tests.

Dixon's Outlier Test (Extreme Value Test): Dixon's test is used to identify statistical outliers when the sample size is ≤ 25. This test identifies outliers or extreme values in the left tail (Case 2) and also in the right tail (Case 1) of a data distribution. In environmental data sets, outliers found in the right tail, potentially representing impacted locations, are of interest. The Dixon test assumes that the data without the suspected outlier (s) are normally distributed. This test tends to suffer from masking in the presence of multiple outliers. This means that if more than one outlier (in either tail) is suspected, this test may fail to identify all of the outliers.

Rosner Outlier Test: This test can be used to identify up to 10 outliers in data sets of sizes 25 and higher. This test also assumes that the data set without the suspected outliers is normally distributed. The detailed discussion of these two tests is given in the associated ProUCL Technical Guide. A couple of examples illustrating the identification of outliers in data sets with NDs are described in the following sections.

4.1.1 Outlier Test Example

For this example, we use a dataset with NDs and chose to exclude them from the outlier test. If your dataset does not include NDs simply select the **Full (w/o NDs)** option, or if the dataset has NDs and you wish to impute ½ the detection limit select the **DL/2 Estimates** option.

Click Statistical Tests ► Outlier Tests ► With NDs ► Exclude NDs

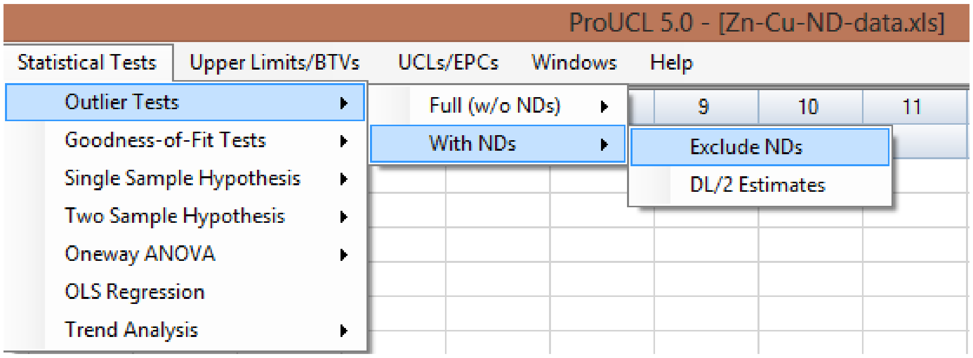

Figure 4-2. Performing Outlier Tests with NDs Excluded.

The **Select Variables** screen (Section 1.3.1.2) will appear.

- Select one or more variable(s) from the **Select Variables** screen.
- If outlier test needs to be performed by using a Group variable, then select a group variable by clicking the arrow below the **Select Group Column (Optional)** 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 (or group) has 25 or more observations, then click the option button for the Rosner Test. ProUCL automatically performs the Dixon test for data sets of sizes ≤ 25.

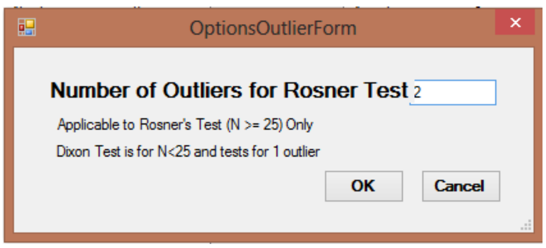

Figure 4-3. Options Related to Performing Outlier Tests.

- The default option for the number of suspected outliers is 1. To use the Rosner test, the user has to obtain an initial guess about the number of suspected outliers that may be present in the data set. This can be done by using graphical displays such as a Q-Q plot. On a 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 Test.

Table 4-1. Output Screen for Dixon's Outlier Test

Dixon's Outlier Test for TCE-1
Total N = 12
Number NDs = 4
Number Detects = 8
10% critical value: 0.479
5% critical value: 0.554
1% critical value: 0.683
Note: NDs excluded from Outlier Test
1. Data Value 9.29 is a Potential Outlier (Upper Tail)?
Test Statistic: 0.392
For 10% significance level, 9.29 is not an outlier.
For 5% significance level, 9.29 is not an outlier.
For 1% significance level, 9.29 is not an outlier.

2. Data Value 0.75 is a Potential Outlier (Lower Tail)?
Test Statistic: 0.011
For 10% significance level, 0.75 is not an outlier.
For 5% significance level, 0.75 is not an outlier.
For 1% significance level, 0.75 is not an outlier.

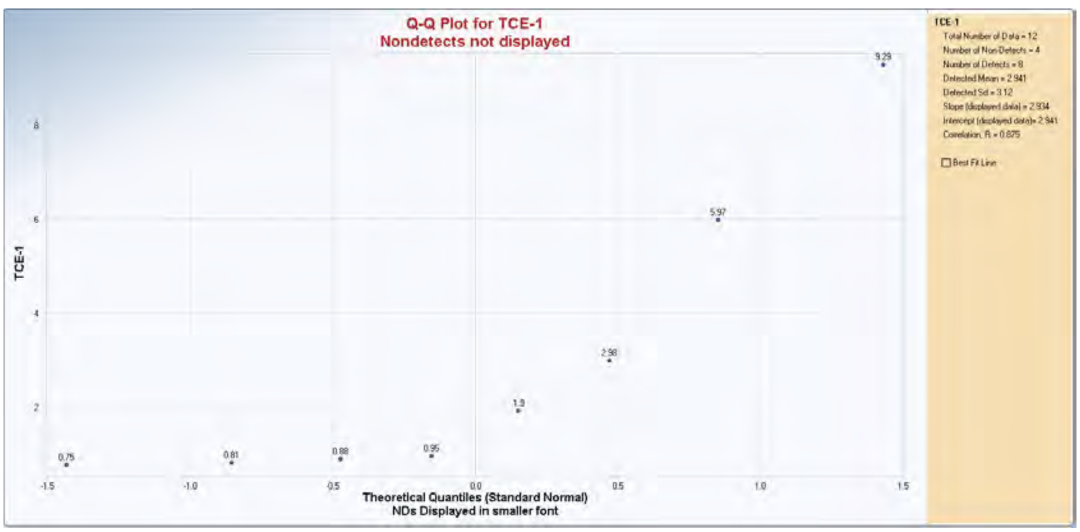

Figure 4-4. Q-Q plot without Four Non-detect Observations

Example 4-1: Rosner's Outlier Test by a Group Variable, Zone

Selected Options: Number of Suspected Outliers = 4

- NDs excluded from the Rosner Test
- Outlier test performed using the **Select Group Column (Optional)**

Table 4-2. Output Screen for Rosner's Outlier Test for Zinc in Zone: Alluvial Fan

Rosner's Outlier Test for 4 Outliers in Zn (alluvial fan)

Total N	**67**
Number NDs	**16**
Number Detects	**51**
Mean of Detects	**27.88**
SD of Detects	**85.02**
Number of data	**51**
Number of suspected outliers	**4**
s not included in the following:	

#	Mean	sd	Potential outlier	Obs. Number	Test value	Critical value (5%)	Critical value (1%)
1	27.88	84.18	620	26	7.034	3.137	3.488
2	16.04	8.776	50	28	3.87	3.127	3.478
3	15.35	7.356	40	27	3.352	3.118	3.469
4	14.83	6.485	33	29	2.801	3.108	3.468

For 5% significance level, there are 3 Potential Outliers

620, 50, 40

For 1% Significance Level, there are 2 Potential Outliers

620, 50

Figure 4-5. Q-Q plot for Zinc Based upon Detected Data (Alluvial Fan)

Table 4-3. Output Screen for Rosner's Outlier Test for Zinc in Zone: Basin Trough

Rosner's Outlier Test for 4 Outliers in Zn (basin trough)	
Total N	50
Number NDs	4
Number Detects	46
Mean of Detects	23.13
SD of Detects	19.03
Number of data	46
Number of suspected outliers	4
s not included in the following:	

#	Mean	sd	Potential outlier	Obs. Number	Test value	Critical value (5%)	Critical value (1%)
1	23.13	18.82	90	45	3.553	3.09	3.45
2	21.64	16.32	70	21	2.963	3.09	3.44
3	20.55	14.73	60	3	2.679	3.08	3.43
4	19.63	13.57	60	22	2.975	2.07	3.41

For 5% significance level, there are 4 Potential Outliers

90, 70, 60, 60

For 1% Significance Level, there is 1 Potential Outlier

4.2 Goodness-of-Fit (GOF) Tests

GOF tests are available under the **Statistical Test** module of ProUCL. The details and usage of the various GOF tests are described in Chapter 2 of the associated ProUCL Technical Guide. Several GOF tests for uncensored full (**Full (w/o NDs)**) and left-censored (**With NDs**) data sets are available in the ProUCL software.

Note that GOF test may fail to detect the actual non-normality of the population distribution for small sample sizes (n<20). For large sample sizes (n>50), a small deviation from normality may lead to rejecting the normality hypothesis.

4.2.1 Full (w/o NDs)

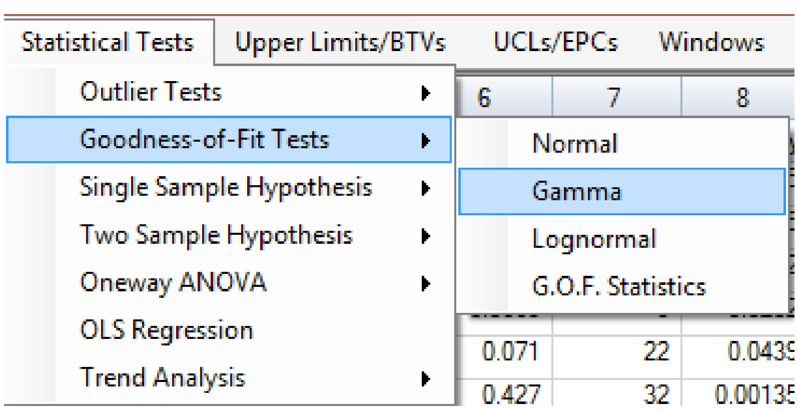

Figure 4-6. Performing GOF Tests with no NDs.

- This option is used on uncensored full data sets without any ND observations. This option can be used to determine GOF for normal, gamma, or lognormal distribution of the variable(s) selected using the **Select Variables** option.
- Like all other methods in ProUCL, GOF tests can also be performed on variables categorized by a Group ID variable.
- Based upon the hypothesized distribution (normal, gamma, lognormal), a Q-Q plot displaying all statistics of interest including the derived conclusion is also generated.
- The **G.O.F. Statistics** option generates a detailed output log (Excel type spreadsheet) showing all GOF test statistics (with derived conclusions) available in ProUCL. This option helps a user to determine the distribution of a data set before generating a GOF Q-Q plot for the hypothesized distribution. This option was included at the request of some users in earlier versions of ProUCL.

4.2.2 With NDs

- This option performs GOF tests on data sets consisting of both non-detected and detected data values.
- Several sub-menu items shown below are available for this option.

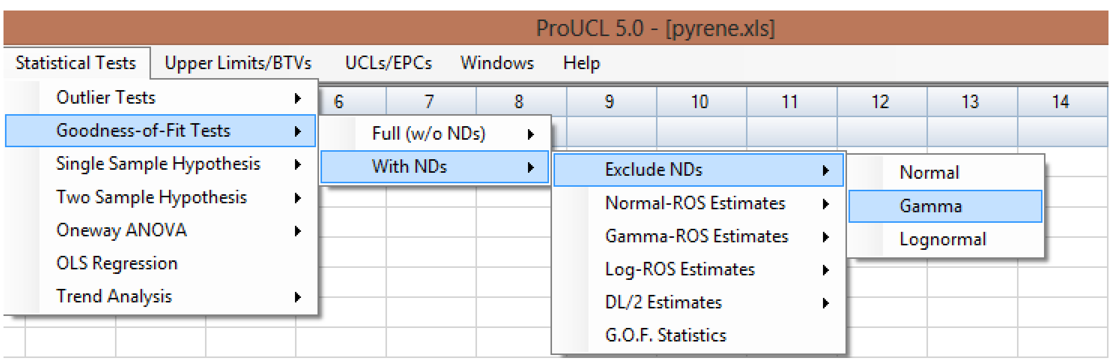

Figure 4-7. Performing GOF Tests with NDs.

Exclude NDs: tests for normal, gamma, or lognormal distribution of the selected variable(s) using only the detected values. This option is the most important option for a GOF test applied to data sets with ND observations. Based upon the skewness and distribution of detected data, ProUCL computes the appropriate decision statistics (UCLs, UPLs, UTLs, and USLs) which accommodate data skewness. Specifically, depending upon the distribution of detected data, ProUCL uses KM estimates in parametric or nonparametric upper limits computation formulae (UCLs, UTLs) to estimate EPC and BTV estimates.

ROS Estimates: Three ROS methods for normal, lognormal (Log), and gamma distributions are available. This option imputes the NDs based upon the specified distribution and performs the specified GOF test on the data set consisting of detects and imputed non-detects. However, it is not recommended to use ROS estimates in the presence of a large amount of non-detects. Please see the ProUCL Technical Guide section 4.5 for more information.

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. ProUCL does not make any recommendations based upon this option.

G.O.F. Statistics: Like full uncensored data sets, this option generates an output log of all GOF test statistics available in ProUCL for data sets with non-detects. The conclusions about the data distributions for all selected variables are also displayed on the generated output file (Excel-type spreadsheet).

Multiple variables: When multiple variables are selected from the **Select Variables** screen, one can use one of the following two options:

Use the **Group Graph**s option to produce multiple GOF Q-Q plots for all selected variables in a single graph. This option may be used when a selected variable has data coming from two or more groups or populations. The relevant statistics (e.g., slope, intercept, correlation, test statistic and critical value) associated with the selected variables are shown on the right panel of the GOF Q-Q plot. 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.

The **Individual Graphs** option is used to generate individual GOF Q-Q plots for each of the selected variables, one variable at a time (or for each group individually of the selected variable categorized by a Group ID). This is the most commonly used option to perform GOF tests for the selected variables.

GOF Q-Q plots for hypothesized distributions: ProUCL computes the relevant test statistic and the associated critical value and prints them on the associated Q-Q plot (called GOF Q-Q plot). On a GOF Q-Q plot, the program informs the user if the data are gamma, normally, or lognormally distributed.

For all options described above, ProUCL generates GOF Q-Q plots based upon the hypothesized distribution (normal, gamma, lognormal). All GOF Q-Q plots display several statistics of interest including the derived conclusion.

The linear pattern displayed by a GOF Q-Q plot suggests an approximate GOF 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) may be an indication of a good fit for that distribution; however, the high correlation should exhibit a definite linear pattern in the Q-Q plot without breaks and discontinuities.

On a GOF Q-Q plot, observations that are well separated from the majority of the data typically represent potential outliers needing further investigation.

Significant and obvious jumps and breaks and curves in a Q-Q plot are indications of the presence of more than one population. Data sets exhibiting such behavior of Q-Q plots may require partitioning of the data set into component subsets (representing sub-populations present in a mixture data set) before computing upper limits to estimate EPCs or BTVs. It is recommended that both graphical and formal goodness-of-fit tests 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 GOF test is also available to test the normality or lognormality of the data set.

<u>Lilliefors Test:</u> a test typically used for samples of size larger than 50 (> 50). However, the Lilliefors test (generalized Kolmogorov Smirnov [KS] test) is available for samples of all sizes. There is no applicable upper limit for sample size for the Lilliefors test.

<u>Shapiro and Wilk</u> (SW, S-W) Test: a test used for samples of size smaller than or equal to 2000 (<= 2000). In ProUCL 5.2, the SW test uses the exact SW critical values for samples of size 50 or less. The SW test statistic is displayed along with the *p*-value of the test (Royston 1982a, 1982b).

<u>Notes:</u> As with other statistical tests, sometimes these two GOF tests might lead to different conclusions. The user is advised to exercise caution when interpreting these test results. When one the GOF tests passes the hypothesized distribution, ProUCL determines that the data set follows an approximate hypothesized distribution. It should be pointed out that for data sets of smaller sizes (e.g., <50), when Lilliefors tests determines that the data set follows a normal (lognormal) distribution the Shapiro-Wilk's test may determine that the data set does not follow a normal (lognormal) distribution. Users should use caution when interpreting GOF tests when the sample size is small.

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 AD test and the KS 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 4-2,500. Also, for these two tests, the value (known or estimated) of the shape parameter, *k* (*k* hat) should lie in the interval [0.01, 100.0]. Consult the associated ProUCL Technical Guide for a detailed description of the gamma distribution and its parameters, including *k*. Extrapolation of critical values beyond these sample sizes and values of *k* is not recommended.

<u>Notes:</u> Even though, the **GOF 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.

4.2.3 GOF Tests for Normal and Lognormal Distributions

Click Goodness-of-Fit Tests ► Chose your handling of NDs if applicable ► Normal or Lognormal

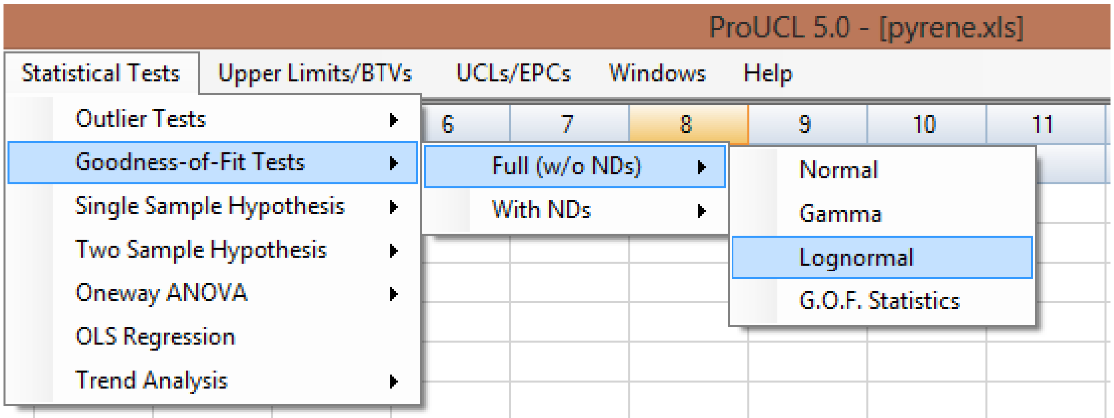

Figure 4-8. Performing GOF Tests for Lognormal Distributions.

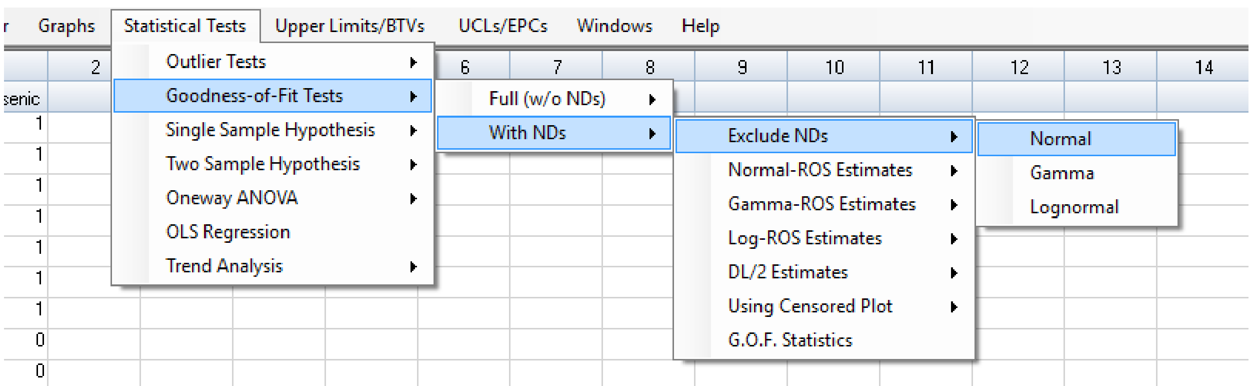

Figure 4-9. Performing GOF Tests for Normal Distributions.

Note: The images above are simply shown as options when using datasets without and with non-detects. The choice of ND imputation method as well as Normal vs Lognormal are available regardless of those choices.

The **Select Variables** screen (Section 1.3.1.2) 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 **Select Group Column (Optional)** 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.

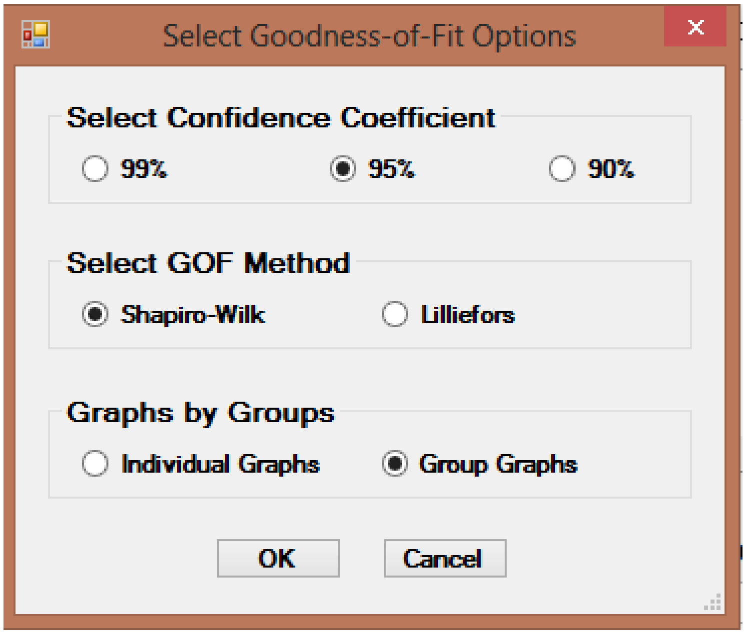

Figure 4-10. Options Related to Performing GOF Tests for Normal and Lognormal Distributions.

- The default option for the **Confidence Level** is **95%**.
- The default GOF Method is **Shapiro-Wilk**.
- The default option for **Graphs by Group** is **Group Graphs**. If you want to see the plots for all selected variables individually, and then check the button next to **Individual Graphs**.
- Click **OK** button to continue or **Cancel** button to cancel the GOF tests.

<u>Notes</u>: This option for **Graphs by Group** is specifically provided for 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).

Example 4-2: Consider the chromium concentrations data set included in your ProUCL download file superfund.xls. The lognormal and normal GOF test results on chromium concentrations are shown in the following figures.

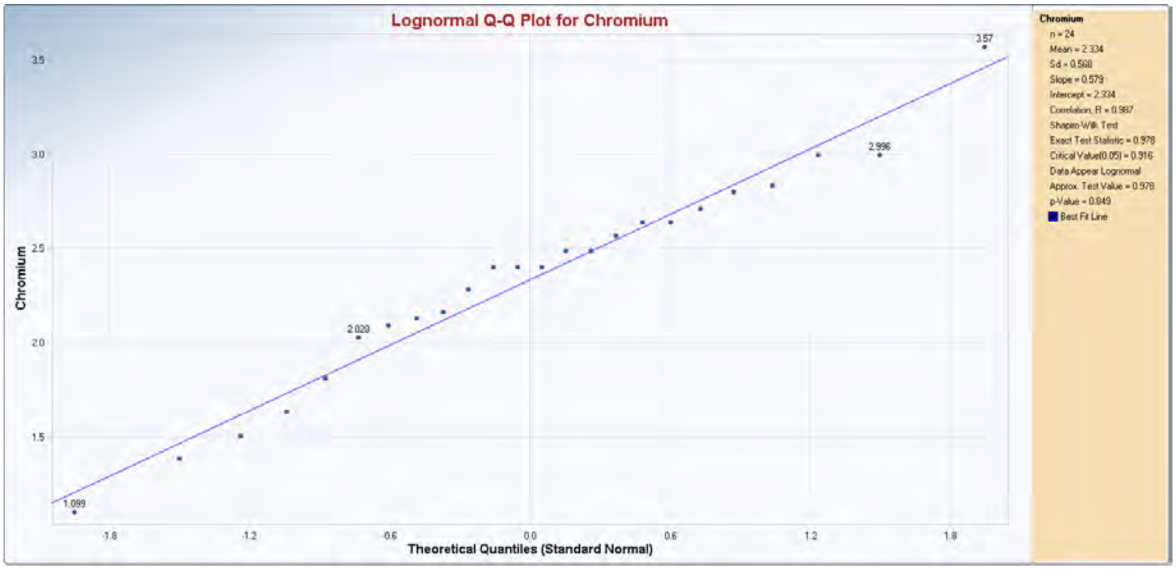

Figure 4-11. Output Screen for Lognormal Distribution (Full (w/o NDs)) Selected Options: Shapiro-Wilk

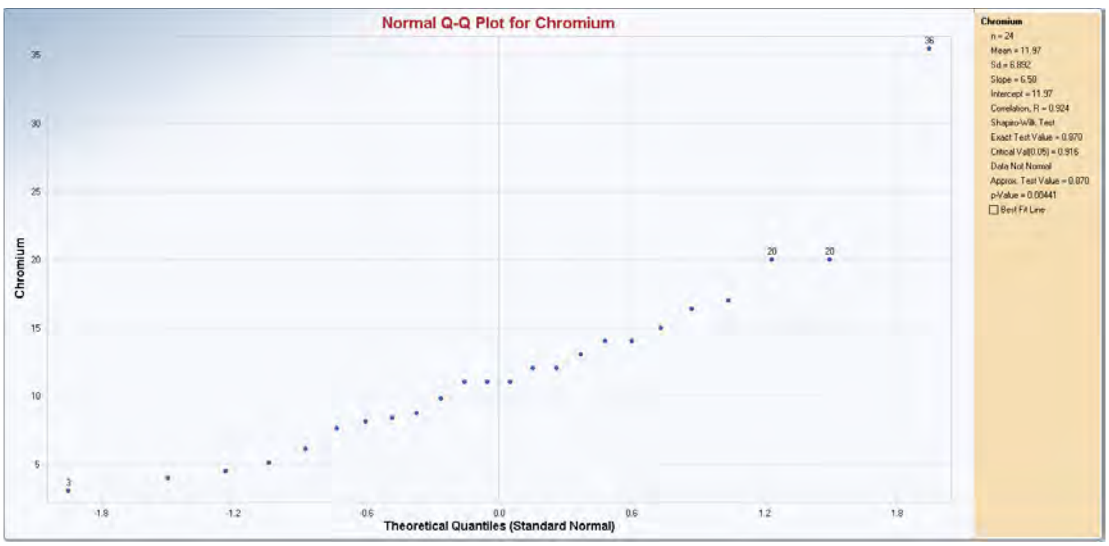

Figure 4-12. Output Screen for Normal Distribution (Full (w/o NDs)) Selected Options: Shapiro-Wilk, Best Fit Line Not Displayed

4.2.4 GOF Tests for Gamma Distribution

Click **Goodness-of-Fit Tests** ► Chose your handling of NDs if applicable ► **Gamma**

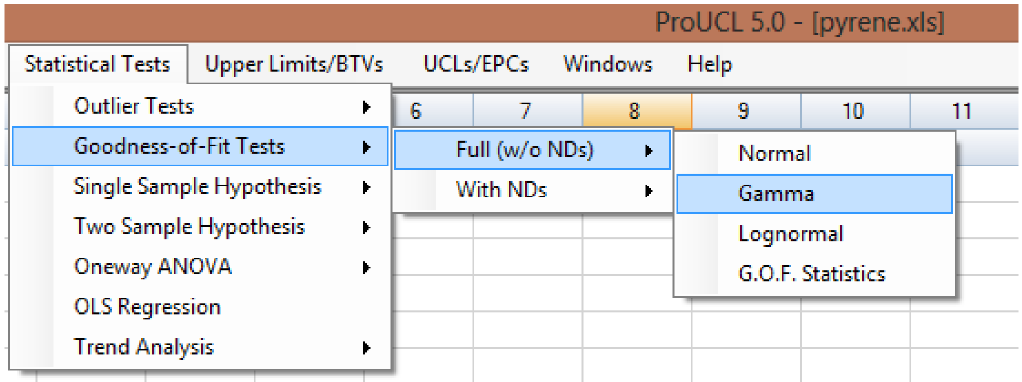

Figure 4-13. Performing GOF Tests for Gamma Distributions with no ND data.

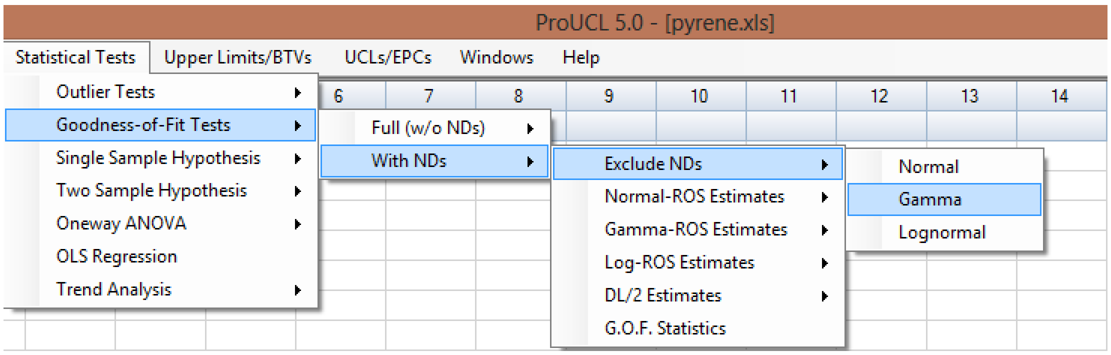

Figure 4-14. Performing GOF Tests for Gamma Distributions with ND data.

The **Select Variables** screen (Section 1.3.2) 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 **Select Group Column (Optional)** 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.

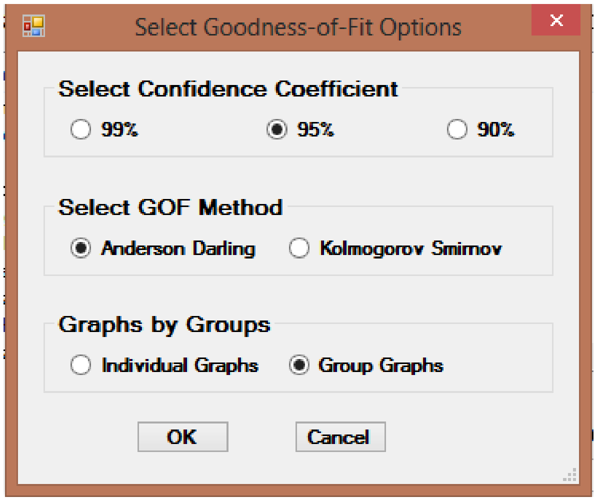

Figure 4-15. Options Related to Performing GOF Tests for Gamma Distributions.

- The default option for the **Confidence Coefficient** is **95%**.
- The default GOF method is **Anderson Darling**.
- The default option for **Graph by Groups** is **Group Graphs**. If you want to see individual graphs, then check the radio button next to **Individual Graphs**.
- Click the **OK** button to continue or the **Cancel** button to cancel the option.
- Click **OK** button to continue or **Cancel** button to cancel the GOF tests.

Example 4-3: Consider arsenic concentrations data set provided in the ProUCL download as superfund.xls. The Gamma GOF test results for arsenic concentrations, are shown in the following G.O.F. Q-Q plot.

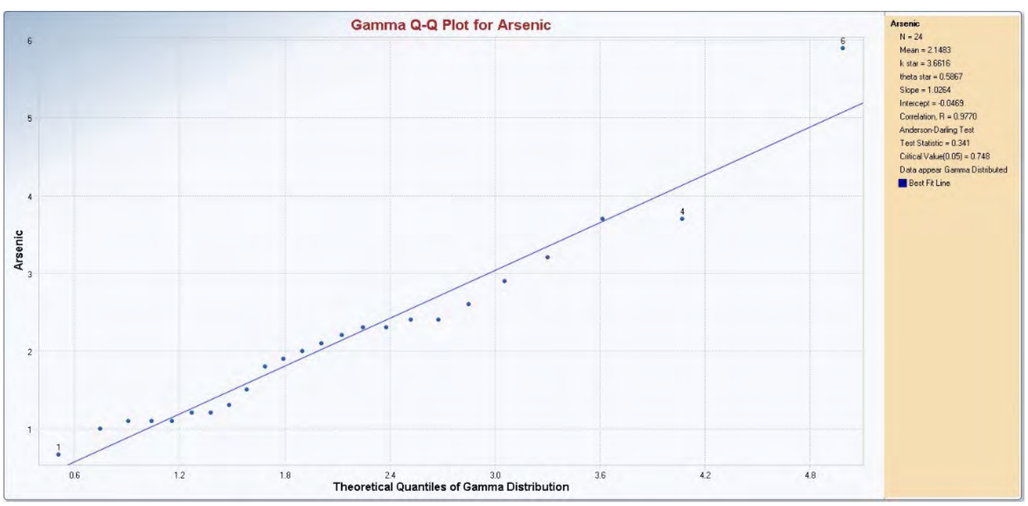

Figure 4-16. Output Screen for Gamma Distribution (Full (w/o NDs)) Selected Options: Anderson Darling with Best Line Fit

4.2.5 Goodness-of-Fit Test Statistics

The **G.O.F.** option displays all GOF test statistics available in ProUCL. This option is used when the user does not know which GOF test to use to determine the data distribution. Based upon the information provided by the GOF test results, the user can perform an appropriate GOF test to generate GOF Q-Q plot based upon the hypothesized distribution. This option is available for uncensored as well as left censored data sets. Input and output screens associated with the G.O.F statistics option for data sets with NDs are summarized as follows.

Click Goodness-of-Fit ► Chose your handling of NDs if applicable ► G.O.F. Statistics

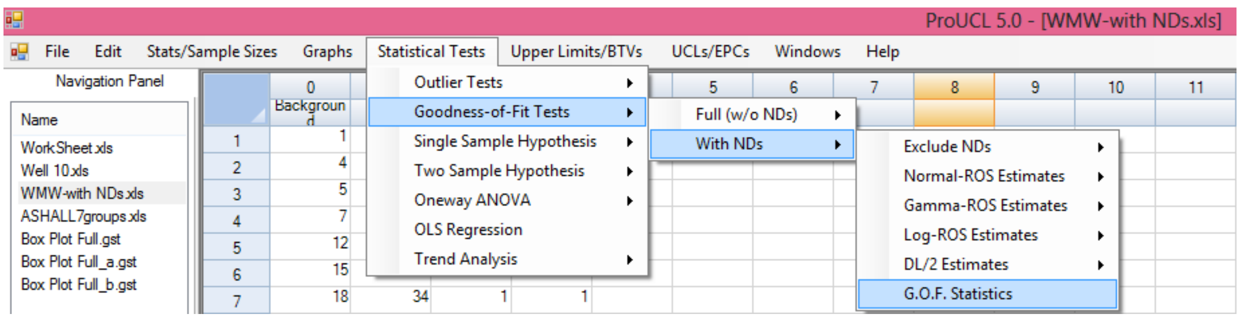

Figure 4-17. Computing GOF Statistics.

The **Select Variables** screen (Section 1.3.2) 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.

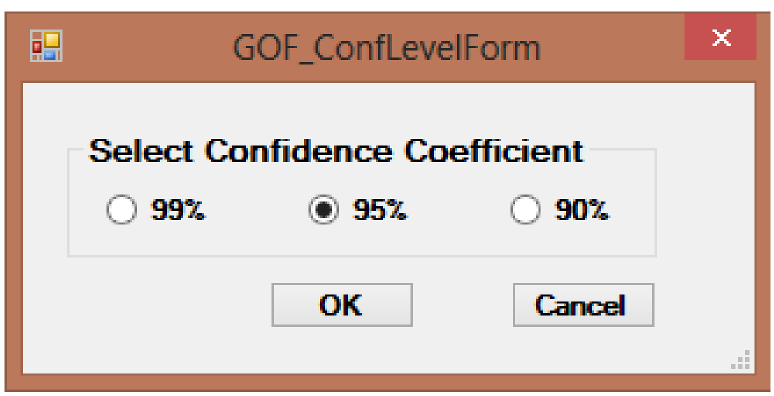

Figure 4-18. Options Related to Computing GOF Statistics.

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

Example 4-3: (continued). Consider the arsenic Oahu data set with NDs. Partial GOF test results, obtained using the **G.O.F. Statistics** option, are summarized in the following table. Note that "K hat", "K star", and "Theta hat" refer to parameter estimates of a gamma distribution, while "Log Mean" and "Log Stdv" refer to parameter estimates of a lognormal distribution (i.e., the mean and SD off the log-transformed dataset).

Table 4-4. Sample Output Screen for G.O.F. Test Statistics on Data Sets with Non-detect Observations

Arsenic

	Num Obs	Num Miss	Num Valid	Detects	NDs	% NDs
Raw Statistics	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 Imputed Data)	24	-0.0995	3.2	0.997	0.737	0.776
Statistics (Gamma ROS Imputed Data)	24	0.119	3.2	0.956	0.7	0.758
Statistics (Lognormal ROS Imputed Data)	24	0.349	3.2	0.972	0.7	0.718

	K hat	K Star	Theta hat	Log Mean	Log Stdv	Log CV
Statistics (Detects Only)	2.257	1.702	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
Statistics (Gamma ROS Estimates)	2.071	1.84	0.461	--	--	--
Statistics (Lognormal ROS Estimates)	--	--	--	-0.209	0.571	-2.727

Normal GOF Test Results

	No NDs	NDs = DL	NDs = DL/2	Normal ROS
Correlation Coefficient R	0.887	0.948	0.833	0.928

	Test value	Crit. (0.05)	Conclusion with Alpha(0.05)
Shapiro-Wilk (Detects Only)	0.777	0.85	Data Not Normal
Shapiro-Wilk (NDs = DL)	0.89	0.916	Data Not Normal
Shapiro-Wilk (NDs = DL/2)	0.701	0.916	Data Not Normal
Shapiro-Wilk (Normal ROS Estimates)	0.868	0.916	Data Not Normal
Lilliefors (Detects Only)	0.273	0.251	Data Not Normal
Lilliefors (NDs = DL)	0.217	0.177	Data Not Normal
Lilliefors (NDs = DL/2)	0.335	0.177	Data Not Normal
Lilliefors (Normal ROS Estimates)	0.17	0.177	Data Appear Normal

Table 4-4 (continued). Sample Output Screen for G.O.F. Test Statistics on Data Sets with Non-detect Observations

Gamma GOF Test Results				
	No NDs	NDs = DL	NDs = DL/2	Gamma ROS
Correlation Coefficient R	0.964	0.956	0.924	0.975

	Test value	Crit. (0.05)	Conclusion with Alpha(0.05)
Anderson-Darling (Detects Only)	0.787	0.738	
Kolmogorov-Smirnov (Detects Only)	0.254	0.258	Detected Data appear Approximate Gamma Disti
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)	0.48	0.755	
Kolmogorov-Smirnov (Gamma ROS Est.)	0.126	0.18	Data Appear Gamma Distributed

Lognormal GOF Test Results				
	No NDs	NDs = DL	NDs = DL/2	Log ROS
Correlation Coefficient R	0.939	0.959	0.933	0.963

	Test value	Crit. (0.05)	Conclusion with Alpha(0.05)
Shapiro-Wilk (Detects Only)	0.86	0.85	Data Appear Lognormal
Shapiro-Wilk (NDs = DL)	0.906	0.916	Data Not Lognormal
Shapiro-Wilk (NDs = DL/2)	0.865	0.916	Data Not Lognormal
Shapiro-Wilk (Lognormal ROS Estimates)	0.924	0.916	Data Appear Lognormal
Lilliefors (Detects Only)	0.229	0.251	Data Appear Lognormal
Lilliefors (NDs = DL)	0.214	0.177	Data Not Lognormal
Lilliefors (NDs = DL/2)	0.217	0.177	Data Not Lognormal
Lilliefors (Lognormal ROS Estimates)	0.143	0.177	Data Appear Lognormal

Note: Substitution methods such as DL or DL/2 are not recommended.

4.3 Hypothesis Testing

This chapter illustrates single-sample and two-sample parametric and nonparametric hypotheses testing approaches as incorporated in the ProUCL software. All hypothesis tests are available under the **Statistical Tests** module of ProUCL. ProUCL software can perform these hypotheses tests on data sets with and without ND observations. It should be pointed out that when one wants to use two-sample hypotheses tests on data sets with NDs, ProUCL assumes that samples from both of the samples/groups have ND observations. All this means is that a ND column (with 0 or 1 entries only) needs to be provided for the variable in each of the two samples. This has to be done even if one of the samples (e.g., Site) has all detected entries; in this case the associated ND column will have '1' for all entries. 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.

4.3.1 Single-Sample Hypothesis Tests

In many environmental applications, single-sample hypotheses tests are used to compare site data with pre-specified Cs or CLs. The single-sample hypotheses tests are useful when the environmental parameters such as the Cs, action level, or CLs are known, and the objective is to compare site concentrations with those known pre-established 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 an AOC (or a MW) exceeds a certain specified proportion (e.g., 1%, 5%, 10%).

ProUCL can perform these hypotheses tests on data sets with and without ND observations. However, a single-sample t-test will not account for NDs; the user must select Single Sample Hypothesis > Full (w/o NDs) > t Test. ND observations will be taken at face-value as if they were detected. 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 Cs or a CL (e.g., proportion test), all NDs (if any) should lie below the cleanup standard, Cs. For proper use of these hypotheses testing approaches, the differences between these tests should be noted and understood. Specifically, a t-test or a Wilcoxon Signed Rank (WSR) test is 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, Cs, 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 CL exceeds a specified proportion, *P0* (e.g., 5%, 10%). ProUCL has graphical methods that may be used to visually compare the concentrations of a site AOC with an action level. This can be done using a box plot of site data with horizontal lines displayed at action levels on the same graph. The details of the various single-sample hypotheses testing approaches are provided in the associated ProUCL Technical Guide.

Statistical Tests ►Single Sample Hypothesis ► Chose whether or not your dataset has NDs ► Select appropriate test

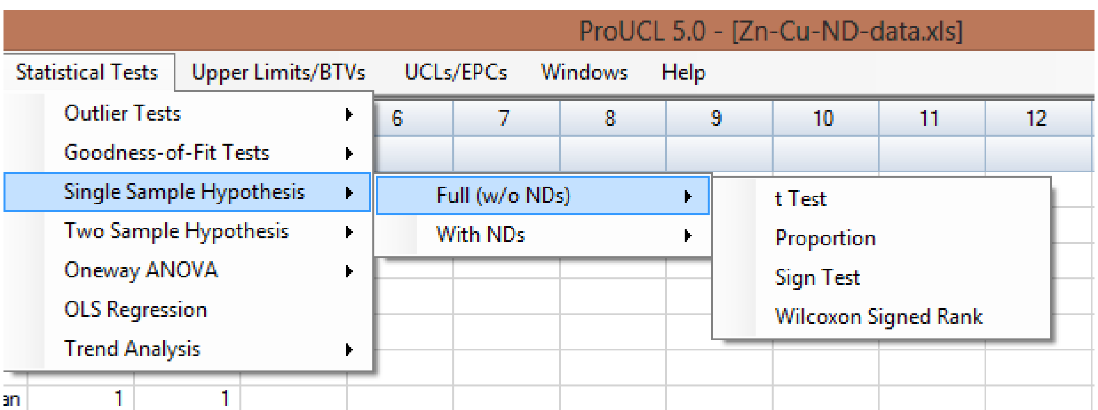

Figure 4-19. Performing Single-Sample Hypothesis Tests.

- To perform a t-test, click on **t-Test** from the drop-down menu as shown above. Note: This test is only available for full datasets without non-detects
- 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 (WSR) test, click on **Wilcoxon Signed Rank** from the drop-down menu.

All single-sample hypothesis tests for uncensored and left-censored data sets can be performed by a group variable. The user selects a group variable by clicking the arrow below the **Select Group Column (Optional)** 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.

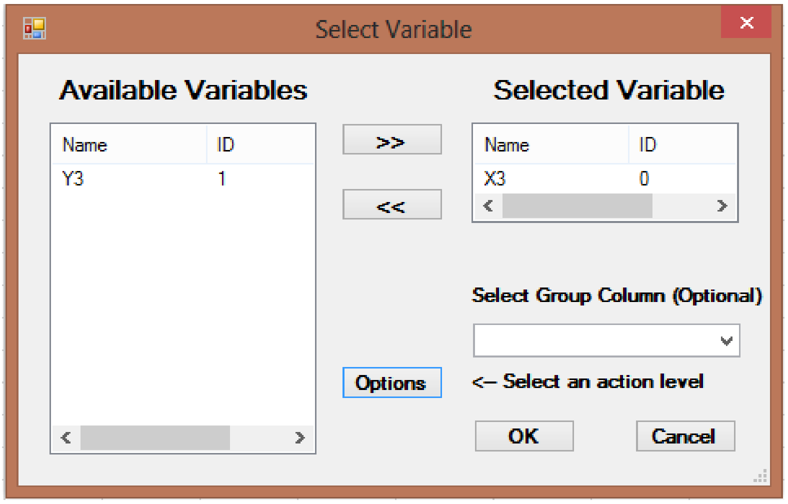

Figure 4-20. Selecting Variables for Single-Sample Hypothesis Tests.

4.3.1.1 Single-Sample t-Test

Note: The single-sample t-Test can only be run on full datasets without non-detects.

Click Statistical Tests ► Single Sample Hypothesis ► Full (w/o NDs) ► t-Test

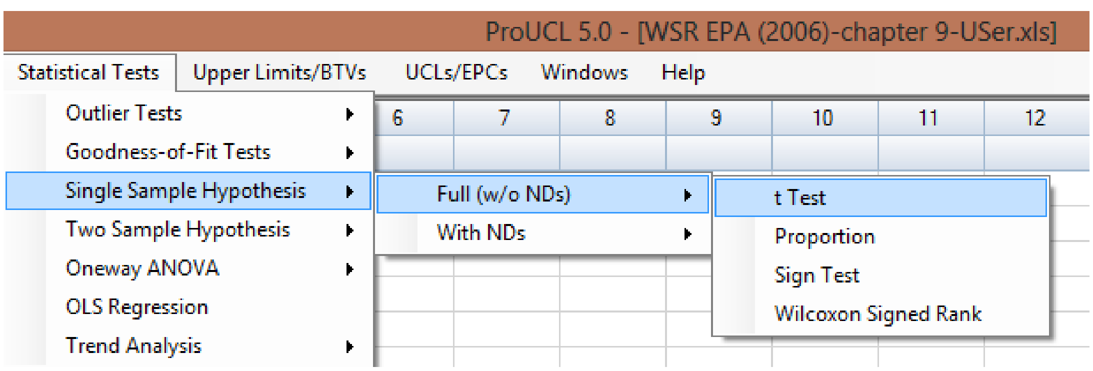

Figure 4-21. Performing a Single-Sample t-Test with no ND data.

The **Select Variables** screen will appear.

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

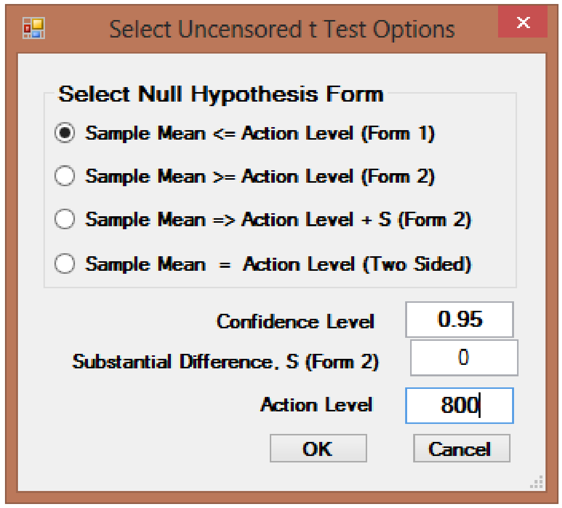

Figure 4-22. Options Related to Performing a Single-Sample t-Test.

- 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 form of Null Hypothesis; default is **Sample Mean <= Action Level (Form 1)**.
- Click on **OK** button to continue or on **Cancel** button to cancel the test.

Example 4-4: Consider the WSR data set described in EPA (2006a). One Sample t-test results are summarized as follows.

Table 4-5. Output for Single-Sample t-Test (Full Data w/o NDs)

From File	WSR EPA (2006)-chapter 9-USer.xls
Full Precision	OFF
Confidence Coefficient	95%
Substantial Difference	0.000
Action Level	800.000
Selected Null Hypothesis	Mean <= Action Level (Form 1)
Alternative Hypothesis	Mean > the Action Level

WSR1	
One Sample t-Test	
Raw Statistics	
Number of Valid Observations	10
Number of Distinct Observations	10
Minimum	750
Maximum	1161
Mean	925.7
Median	888
SD	136.7
SE of Mean	43.24
H0: Sample Mean <= 800 (Form 1)	
Test Value	2.907
Degrees of Freedom	9
Critical Value (0.05)	1.833
P-Value	0.00869
Conclusion with Alpha = 0.05	
Reject H0, Conclude Mean > 800	
P-Value < Alpha (0.05)	

4.3.1.2 Single Sample Proportion Test

Note: When NDs are present, the Proportion test assumes that all ND observations lie below the specified action level, A_0. These single-sample tests are not performed if ND observations exceed the action levels.

Statistical Tests ► Single Sample Hypothesis ► Chose whether or not your dataset has NDs ► Proportion

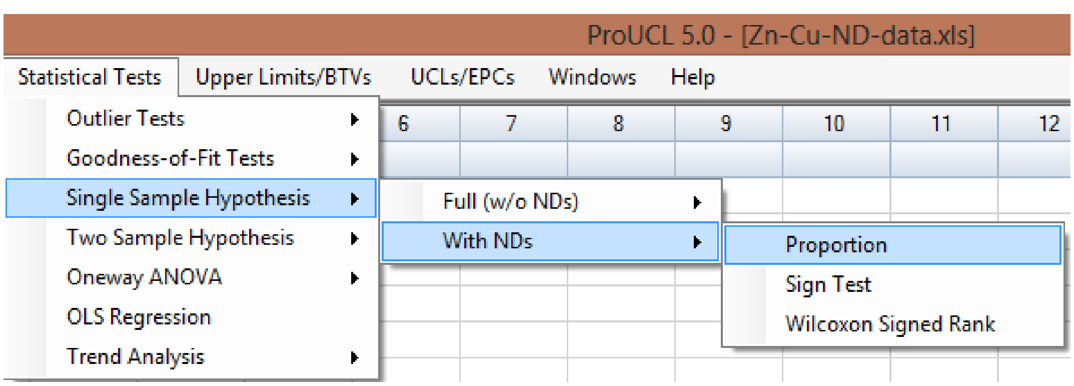

Figure 4-23. Performing a Single-Sample Proportion Test with ND Data.

- Select variable(s) from the **Select Variables** screen.
- If hypothesis test has to be performed by using a Group variable, then select a group variable by clicking the arrow below the **Select Group Column (Optional)** 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. This option has been used in the following screen shot for the single-sample proportion test.

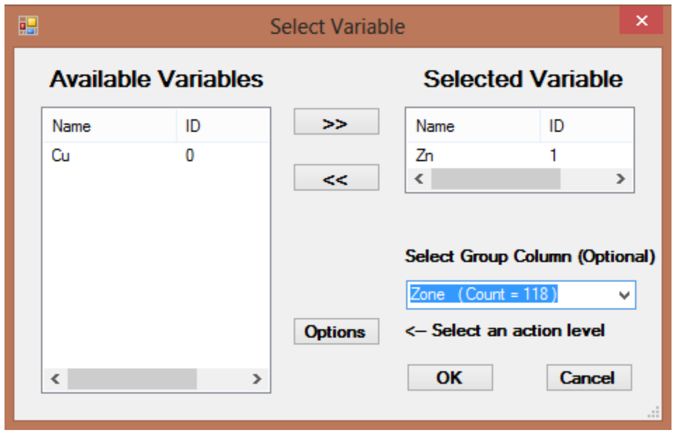

Figure 4-24. Selecting Variables for a Proportion Test.

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

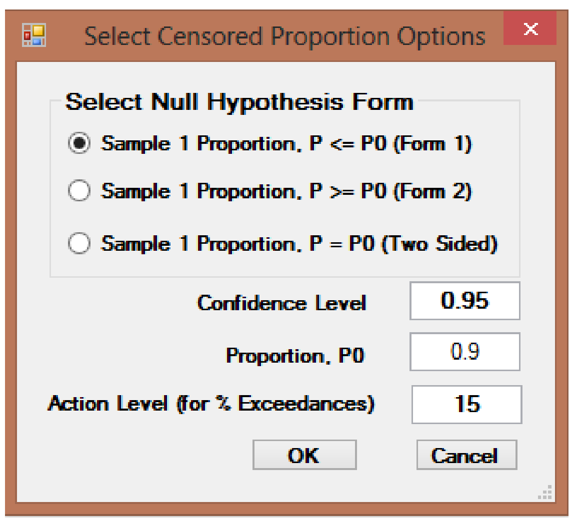

Figure 4-25. Options Related to Performing a Single-Sample Proportion Test.

- Specify the **Confidence Level**; default is **0.95**.
- Specify meaningful values for **Proportion** and the **Action Level (=15 here).**
- Select form of Null Hypothesis; default is **Sample 1 Proportion, P <= P0 (Form 1)**.
- Click on **OK** button to continue or on **Cancel** button to cancel the test.

Example 4-5: Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed in the literature (Helsel 2012b, NADA in R [Helsel 2013]). This data set is used here to illustrate the one sample proportion test on a data set with NDs and is available with your ProUCL 5.2 download as Zn-CU-two-zones-NDs.xls. The output sheet generated by ProUCL is presented below.

Table 4-6. Output for Single-Sample Proportion Test (with NDs) by Groups: Alluvial Fan and Basin Trough

User Selected Options	
Date/Time of Computation	3/18/2013 9:55:58 AM
From File	Zn-Cu-ND-data.xls
Full Precision	OFF
Confidence Coefficient	95%
User Specified Proportion	0.900 (P0 of Exceedances of Action Level)
Action Level	15.000
Select Null Hypothesis	Sample Proportion, P of Exceedances of Action Level <= User Specified Proportion (Form 1)
Alternative Hypothesis	Sample Proportion, P of Exceedances of Action Level > User Specified Proportion

Zn (alluvial fan)

One Sample Proportion Test

Note: All nondetects are treated as detects at values (e.g., DLs) included in Data File

Raw Statistics	
Number of Valid Data	67
Number of Missing Observations	1
Number of Distinct Data	19
Number of Non-Detects	16
Number of Detects	51
Percent Non-Detects	23.88%
Minimum Non-detect	3
Maximum Non-detect	10
Minimum Detect	5
Maximum Detect	620
Mean of Detects	27.88
Median of Detects	11
SD of Detects	85.02
Number of Exceedances	24
Sample Proportion of Exceedances	0.358

H0: Sample Proportion <= 0.9 (Form 1)

Large Sample z-Test Statistic	-14.58
Critical Value (0.05)	1.645
P-Value	1

Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Sample Proportion <= 0.9

P-Value > Alpha (0.05)

Zn (basin trough)

One Sample Proportion Test

Note: All nondetects are treated as detects at values (e.g., DLs) included in Data File

Raw Statistics	
Number of Valid Data	50
Number of Distinct Data	20
Number of Non-Detects	4
Number of Detects	46
Percent Non-Detects	8.00%
Minimum Non-detect	3
Maximum Non-detect	10
Minimum Detect	3
Maximum Detect	90
Mean of Detects	23.13
Median of Detects	20
SD of Detects	19.03
Number of Exceedances	27
Sample Proportion of Exceedances	0.54

H0: Sample Proportion <= 0.9 (Form 1)

Exact P-Value	1

Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Sample Proportion <= 0.9

P-Value > Alpha (0.05)

4.3.1.3 Single-Sample Sign Test

Note: When NDs are present, the Sign test assumes that all ND observations <u>lie below</u> the specified action level, A_0. These single-sample tests are not performed if ND observations exceed the action levels.

Statistical Tests ► Single Sample Hypothesis ► Chose whether or not your dataset has NDs ► Sign test

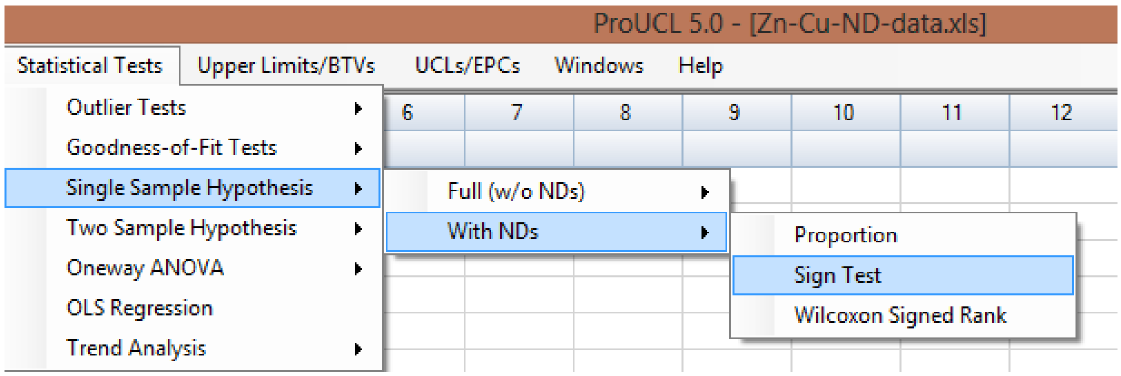

Figure 4-26. Performing a Single-Sample Sign Test with ND Data.

The **Select Variables** screen will appear.

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

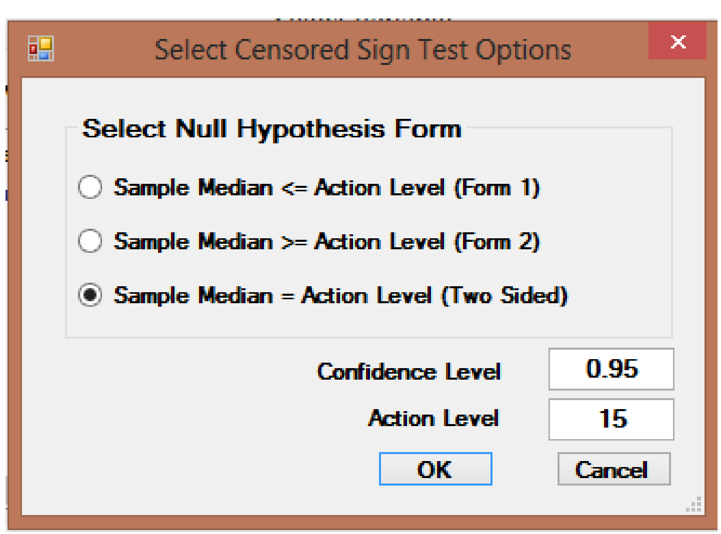

Figure 4-27. Options Related to Performing a Single-Sample Sign Test.

- Specify the **Confidence Level**; default is **0.95**.
- Select an Action Level.
- Select the form of Null Hypothesis; default is **Sample Median <= Action Level (Form 1)**.
- Click on **OK** button to continue or on **Cancel** button to cancel the test.

Example 4-5: (continued). Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed above. This data set is used here to illustrate the Single-Sample Sign test on a data set with NDs. The output sheet generated by ProUCL follows.

Table 4-7. Output for Single-Sample Sign Test (Data with Non-detects)

Selected Null Hypothesis	Median = Action/compliance Limit (Two Sided Alternative)
Alternative Hypothesis	Median <> Action/compliance Limit
Zn (alluvial fan)	
One Sample Sign Test	
Note: All nondetects are treated as detects at values (e.g., DLs) included in Data File	
Raw Statistics	
Number of Valid Data	67
Number of Missing Observations	1
Number of Distinct Data	19
Number of Non-Detects	16
Number of Detects	51
Percent Non-Detects	23.88%
Minimum Non-detect	3
Maximum Non-detect	10
Minimum Detect	5
Maximum Detect	620
Mean of Detects	27.88
Median of Detects	11
SD of Detects	85.02
Number Above Action Level	24
Number Equal Action Level	0
Number Below Action Level	43
H0: Sample Median = 15	
Standardized Test Value using Normal Appx.	-2.321
P-Value	0.0203
Conclusion with Alpha = 0.05	
Reject H0 at the specified level of significance (0.05). Conclude Median <> 15	
P-Value < Alpha (0.05)	

4.3.1.4 Single-Sample Wilcoxon Signed Rank Test

Click Statistical Tests ► Single Sample Hypothesis ► Chose whether or not your dataset has NDs ► Wilcoxon Signed Rank

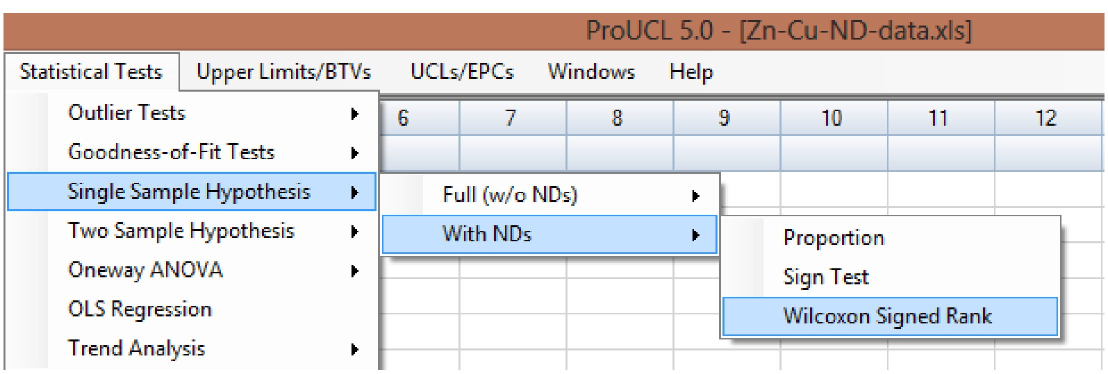

Figure 4-28. Performing a Single-Sample Wilcoxon Signed Rank Test with ND Data.

The **Select Variables** screen will appear.

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

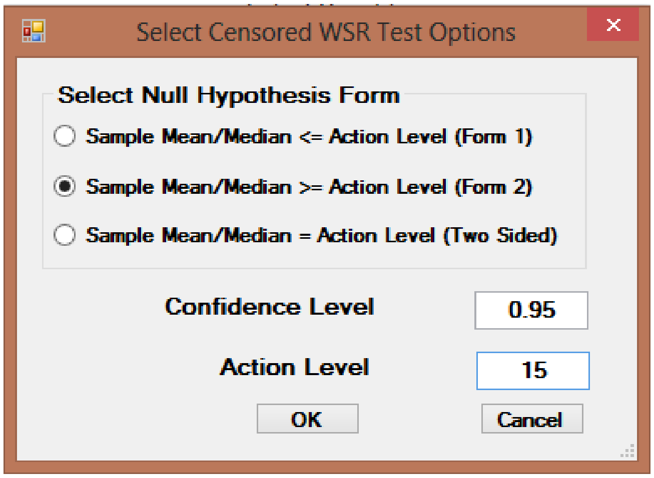

Figure 4-29. Options Related to Performing a Single-Sample Sign Test.

- Specify the **Confidence Level**; default is **0.95**.
- Specify an **Action Level**.
- Select form of Null Hypothesis; default is **Sample Mean/Median <= Action Level (Form 1)**.
- Click on **OK** button to continue or on **Cancel** button to cancel the test.

Example 4-5: (continued). Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed earlier in this chapter. This data set is used here to illustrate one sample Wilcoxon Signed Rank test on a data set with NDs. The output sheet generated by ProUCL is provided as follows.

Table 4-8. Output for Single-Sample Wilcoxon Signed Rank Test (Data with Non-detects)

One Sample Wilcoxon Signed Rank Test for Data Sets with Non-Detects	
User Selected Options	
Date/Time of Computation	3/18/2013 1:48:46 PM
From File	Zn-Cu-ND-data.xls
Full Precision	OFF
Confidence Coefficient	95%
Action Level	15.000
Selected Null Hypothesis	Mean/Median >= Action Level (Form 2)
Alternative Hypothesis	Mean/Median < the Action Level

Table 4-8 (continued). Output for Single-Sample Wilcoxon Signed Rank Test (Data with Nondetects)

Zn (basin trough)			
One Sample Wilcoxon Signed Rank Test			
Raw Statistics			
Number of Valid Data	50		
Number of Distinct Data	20		
Number of Non-Detects	4		
Number of Detects	46		
Percent Non-Detects	8.00%		
Minimum Non-detect	3		
Maximum Non-detect	10	**H0: Sample Median >= 15 (Form 2)**	
Minimum Detect	3		
Maximum Detect	90	Large Sample z-Test Statistic	1.269
Mean of Detects	23.13	Critical Value (0.05)	-1.645
Median of Detects	20	P-Value	0.898
SD of Detects	19.03		
Median of Processed Data used in WSR	18.5	**Conclusion with Alpha = 0.05**	
Number Above Action Level	27	**Do Not Reject H0, Conclude Mean/Median >= 15**	
Number Equal Action Level	1	**P-Value > Alpha (0.05)**	
Number Below Action Level	22		
T-plus	764	**Dataset contains multiple Non-Detect values!**	
T-minus	461	**All NDs are replaced by their respective DL/2**	

4.3.2 Two-Sample Hypothesis Testing Approaches

The two-sample hypotheses testing approaches available in ProUCL are described in this section. Like **Single-Sample Hypothesis**, the **Two-Sample Hypothesis** options are available under the **Statistical Tests** module of ProUCL. These approaches are used to compare the parameters and distributions of two populations (e.g., Background vs. AOC) based upon data sets collected from those populations. Several forms (Form 1, Form 2, and Form 2 with Substantial Difference, S) of the two-sample hypothesis testing approaches are available in ProUCL. The methods are available for full uncensored data sets as well as for data sets with ND observations with multiple detection limits. Some details about this hypothesis form can be found in the background guidance document for CERCLA sites (EPA 2002b).

- **Full (w/o NDs)**—performs parametric and nonparametric hypothesis tests on uncensored data sets consisting of all detected values. The following tests are available:

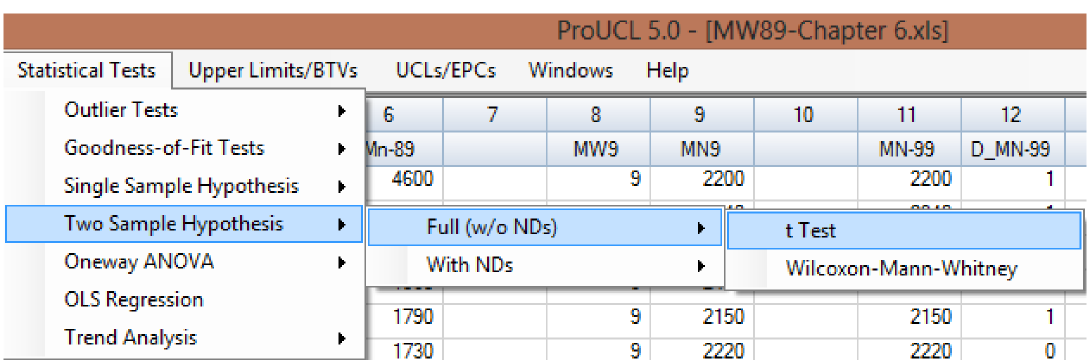

Figure 4-30. Performing Two-Sample Hypothesis Tests.

4.3.2.1 Student's t-Test

Based upon collected data sets, this test is used to compare the mean concentrations of two populations/groups provided the populations are normally distributed. The data sets are represented by independent random observations, X1, X2, . . . , Xn collected from one population (e.g., site), and independent random observations, Y1, Y2, . . . , Ym collected from another (e.g., background) population. The same terminology is used for all other two-sample tests discussed in the following sub-sections of this section.

Student's t-test also assumes that the spreads (variances) of the two populations are approximately equal.

The F-test can be used to the check the equality of dispersions of two populations. A couple of other tests (e.g., Levene 1960) are also available in the literature to compare the variances of two populations. Since the F-test performs fairly well, other tests are not included in the ProUCL software. For more details refer to ProUCL Technical Guides.

4.3.2.2 Two-Sample Nonparametric Wilcoxon-Mann-Whitney Test

This test is used to determine the comparability 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 (mean, median) of the two populations are significantly different.

The Wilcoxon-Mann-Whitney test does not assume that the data are normally or lognormally distributed. For large samples (e.g., ≥ 20), the distribution of the WMW test statistic can be approximated by a normal distribution.

<u>Notes:</u> 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, cleanup and remediation decisions have to be made in the 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.

When the two-sample WMW test is used on a dataset with multiple non-detect limits, all values below the highest ND limit are treated as ND.

4.3.2.3 Gehan Test

The Gehan test is used when many ND observations or multiple DLs are present in the two data sets; therefore, the conclusions derived using this test may not be reliable when dealing with samples of sizes smaller than 10. Furthermore, it has been suggested throughout this guide to have a minimum of 8-10 observations (from each of the populations) to use hypotheses testing approaches, as decisions derived based upon smaller data sets may not be reliable enough to draw important decisions about human health and the environment.

4.3.2.4 Two-Sample t-Test

Click Statistical Tests ► Two Sample Hypothesis ► Full (w/o NDs) ► t Test

The **Select Variables** screen will appear.

- Select variable(s) from the **Select Variables** screen.

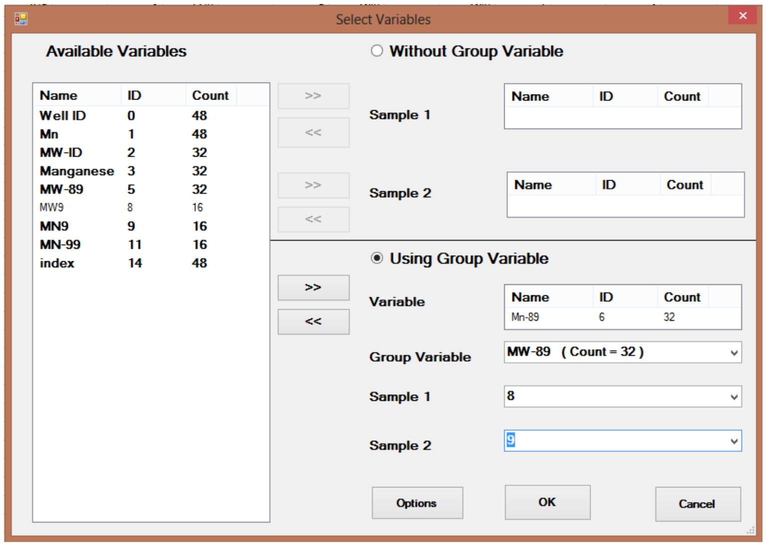

Figure 4-31. Selecting Variables for a Two-Sample t-Test.

- **Without Group Variable**: This option is used when the sampled data of the variable (e.g., lead) for the two populations (e.g., site vs. background) are given in separate columns.
- **With Group Variable**: This option is used when sampled data of the variable (e.g., lead) is composed of two or more populations (e.g., site vs. background) and are given in the same column.
- The values are separated into different populations (groups) by the values of an associated Group ID Variable. The group variable may represent several populations (e.g., background, surface, subsurface, silt, clay, sand, several AOCs, MWs). The user can compare two groups at a time by using this option.
- When the **Group** option is used, the user then selects a variable by using the **Group Variable Option**. The user should select an appropriate variable representing a group variable. The user can use letters, numbers, or alphanumeric labels for the group names.

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

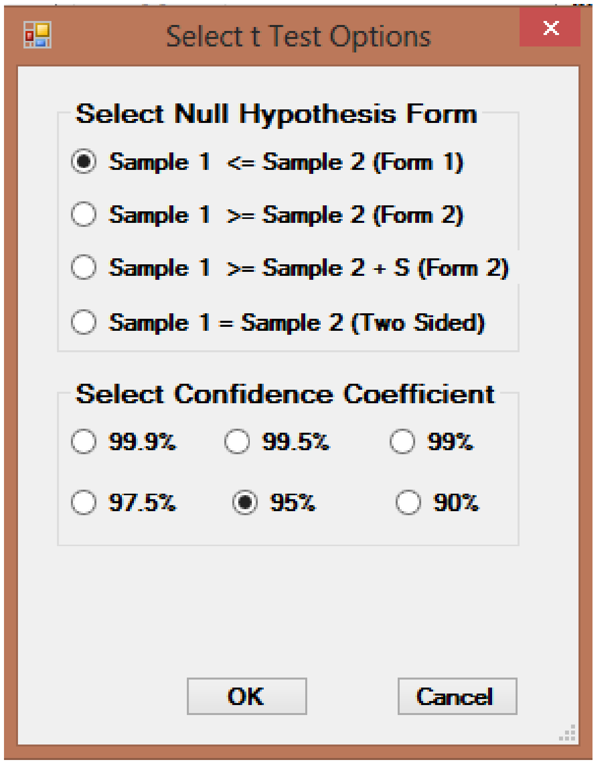

Figure 4-32. Options Related to Performing a Two-Sample t-Test.

- If the 3rd null hypothesis form is selected specify a useful Substantial Difference, S value. The default choice is 0.
- Select the Confidence Coefficient. The default choice is 95%.
- Select the form of Null Hypothesis. The default is Sample 1 <= Sample 2 (Form 1).
- 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 Sample 1 versus Sample 2 Comparison.

Example 4-6. Consider the manganese concentrations data set included with the ProUCL download as MW-1-8-9.xls, the data were collected from three wells: MW1, an upgradient well, and MW8 and MW9, two downgradient wells. The two-sample t-test results, comparing Mn concentrations in MW8 vs. MW9, are described as follows.

Table 4-9. Output for Two-Sample t-Test (Full Data without NDs)

Confidence Coefficient	95%
Substantial Difference (S)	0.000
Selected Null Hypothesis	Sample 1 Mean = Sample 2 Mean (Two Sided Alternative)
Alternative Hypothesis	Sample 1 Mean <> Sample 2 Mean

Sample 1 Data: Mn-89(8)

Sample 2 Data: Mn-89(9)

Raw Statistics

	Sample 1	Sample 2
Number of Valid Observations	16	16
Number of Distinct Observations	16	15
Minimum	1270	1050
Maximum	4600	3080
Mean	1998	1968
Median	1750	2055
SD	838.8	500.2
SE of Mean	209.7	125

Sample 1 vs Sample 2 Two-Sample t-Test

H0: Mean of Sample 1 = Mean of Sample 2

		t-Test	Lower C.Val	Upper C.Val	
Method	DF	Value	t (0.025)	t (0.975)	P-Value
Pooled (Equal Variance)	30	0.123	-2.042	2.042	0.903
Welch-Satterthwaite (Unequal Varian	24.5	0.123	-2.064	2.064	0.903

Pooled SD: 690.548

Conclusion with Alpha = 0.050

Student t (Pooled): Do Not Reject H0, Conclude Sample 1 = Sample 2

Welch-Satterthwaite: Do Not Reject H0, Conclude Sample 1 = Sample 2

Test of Equality of Variances

Variance of Sample 1	703523
Variance of Sample 2	250190

Numerator DF	Denominator DF	F-Test Value	P-Value
15	15	2.812	0.054

Conclusion with Alpha = 0.05

Two variances appear to be equal

For the two-sample t-Test the output also produces values for the Satterthwaite t-Test as well as the F-test. Below provides a brief understanding of their tests and why they are of interest when running a two-sample t-Test. If these tests are not familiar to the user, they should consult a knowledgeable statistician.

4.3.2.5 Satterthwaite t-Test

This test is used to compare the means of two populations when the variances 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 the two populations. However, this test alone is more powerful test to compare the means of two populations.

4.3.2.6 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 sensitive to departures from normality.

4.3.2.7 Two-Sample Wilcoxon-Mann-Whitney Test

Click Statistical Tests ► Two Sample Hypothesis ► Chose whether or not your dataset has NDs ► Wilcoxon-Mann-Whitney

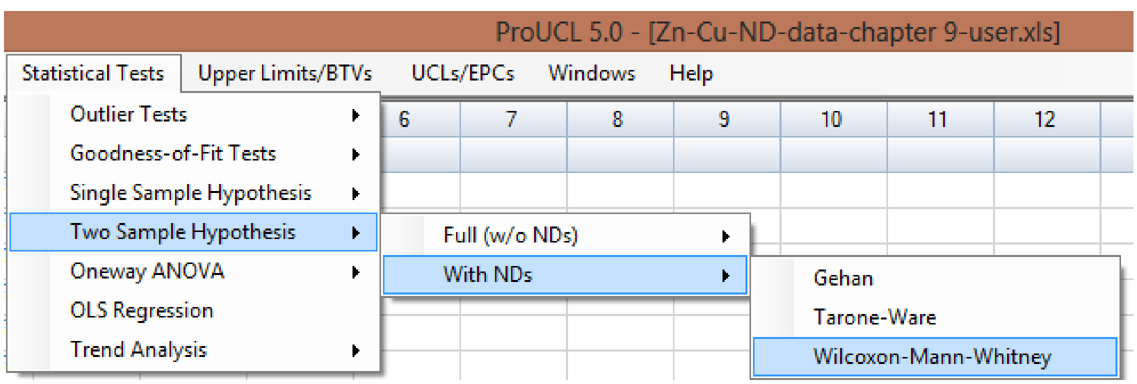

Figure 4-33. Performing a Two-Sample Wilcoxon Mann-Whitney Test.

The **Select Variables** screen shown below will appear.

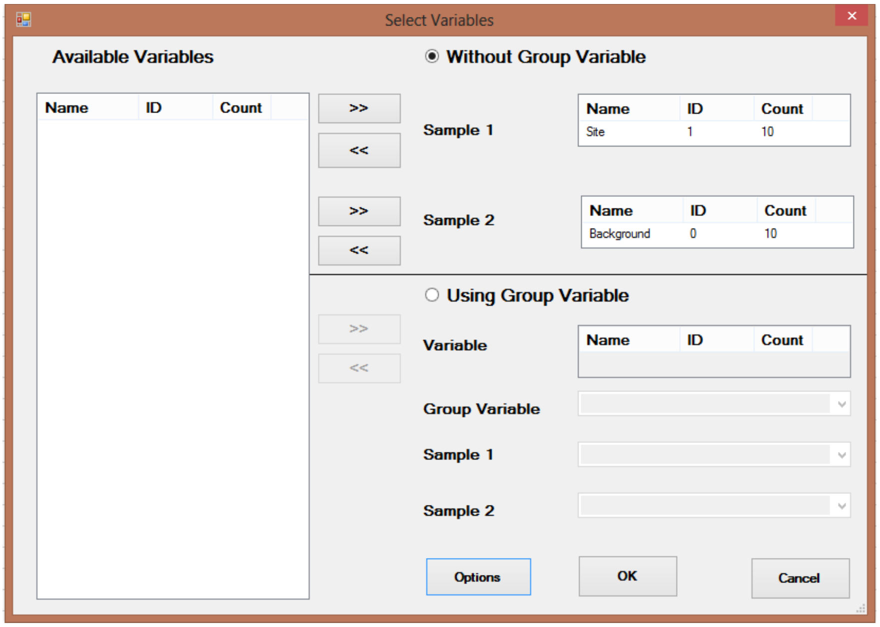

Figure 4-34. Selecting Variables for a Two-Sample Wilcoxon Mann-Whitney Test.

- Select variable(s) from the **Select Variables** screen.
- **Without Group Variable**: This option is used when the data values of the variable (e.g., TCDD 2,3,7,8) for the site and the background are given in separate columns.
- **With Group Variable**: This option is used when data values of the variable (TCDD 2, 3, 7, 8) are given in the same column. The values are separated into different samples (groups) by the values of an associated **Group Variable**. When using this option, the user should select an appropriate variable representing groups such as AOC1, AOC2, AOC3 etc.
- When the **Options** button is clicked, the following window will be shown.

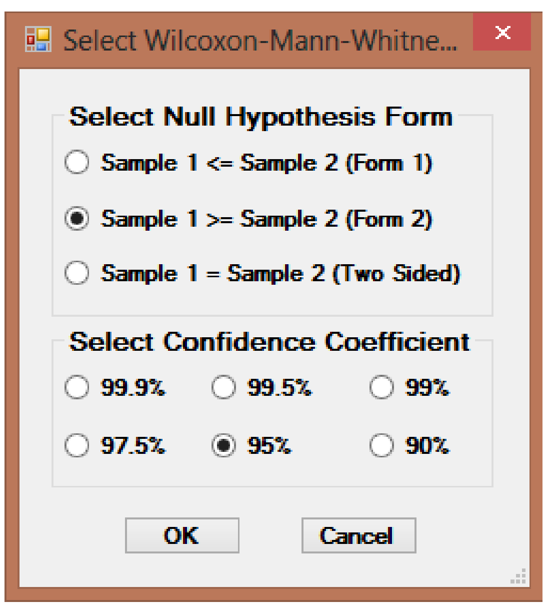

Figure 4-35. Options Related to Performing a Two-Sample Wilcoxon Mann-Whitney Test.

- Choose the **Confidence Coefficient**. The default choice is **95%**.
- Select the form of Null Hypothesis. The default is **Sample 1 <= Sample 2 (Form 1)**.
- Click on **OK** button to continue or on **Cancel** button to cancel the selected options.
- Click on **OK** to continue or on **Cancel** to cancel the Sample 1 vs. Sample 2 comparison.

Example 4-7: Consider a two-sample dataset with non-detects and multiple detection limits included in the ProUCL download as WMW-with NDs.xls. Note that as the data have multiple detection limits, the two sample WMW test will map non-detects to the highest detection limit. It is therefore advised to use this test with caution in the case that the data in question consists of multiple detection limits. The WMW test results are summarized as follows.

Table 4-10. Output for Two-Sample Wilcoxon-Mann-Whitney Test (with Non-detects)

Date/Time of Computation	3/18/2013 6:43:04 PM
From File	WMW-NDs-Chapter 9-user_a.xls
Full Precision	OFF
Confidence Coefficient	95%
Selected Null Hypothesis	Sample 1 Mean/Median >= Sample 2 Mean/Median (Form 2)
Alternative Hypothesis	Sample 1 Mean/Median < Sample 2 Mean/Median

Sample 1 Data: Site
Sample 2 Data: Background

Raw Statistics

	Sample 1	Sample 2
Number of Valid Data	11	11
Number of Non-Detects	3	3
Number of Detect Data	8	8
Minimum Non-Detect	4	4
Maximum Non-Detect	11	9
Percent Non-detects	27.27%	27.27%
Minimum Detect	2	1
Maximum Detect	43	27
Mean of Detects	27	15.5
Median of Detects	29.5	16.5
SD of Detects	13.71	9.196

WMW test is meant for a Single Detection Limit Case
of Gehan or T-W test is suggested when multiple detection limits are pres
All observations <= 11 (Max DL) are ranked the same

WMW test is meant for a Single Detection Limit Case
Use of Gehan or T-W test is suggested when multiple detection limits are present
All observations <= 11 (Max DL) are ranked the same

Wilcoxon-Mann-Whitney (WMW) Test

H0: Mean/Median of Sample 1 >= Mean/Median of Sample 2

Sample 1 Rank Sum W-Stat	144.5
WMW U-Stat	78.5
Mean (U)	60.5
SD(U) - Adj ties	15.22
WMW U-Stat Critical Value (0.05)	35
Standardized WMW U-Stat	1.191
Approximate P-Value	0.883

Conclusion with Alpha = 0.05
Do Not Reject H0, Conclude Sample 1 >= Sample 2

Notes: 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 tends to reduce the associated power of the WMW test considerably. This in turn may lead to an incorrect conclusion.

4.3.2.8 Two-Sample Gehan Test

Click Statistical Tests ► Two Sample Hypothesis ► With NDs ► Gehan

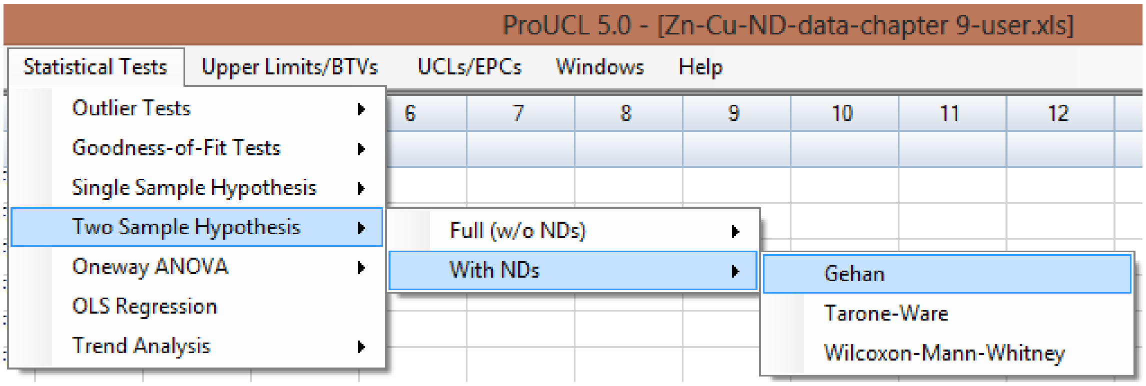

Figure 4-36. Performing a Two-Sample Gehan Test.

The **Select Variables** screen will appear.

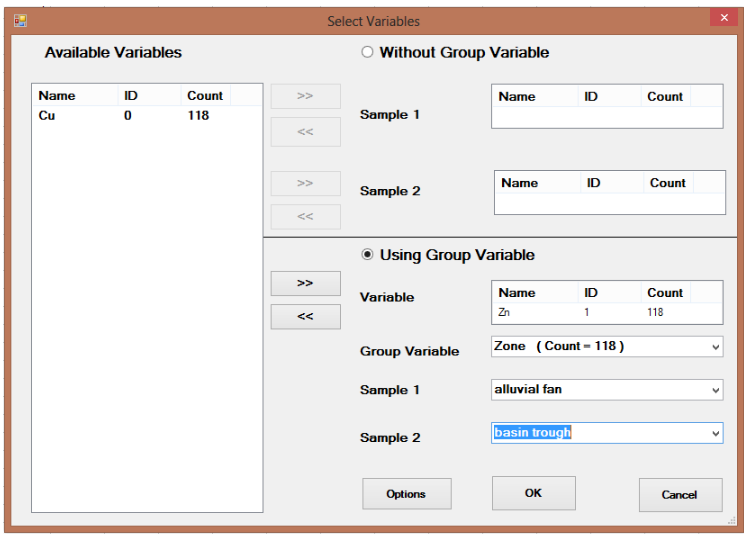

Figure 4-37. Selecting Variables for a Two-Sample Gehan Test.

- Select variable(s) from the **Select Variables** screen.
- **Without Group Variable**: This option is used when the data values of the variable (Zinc) for the two data sets are given in separate columns.
- **With Group Variable**: This option is used when data values of the variable (Zinc) for the two data sets are given in the same column. The values are separated into different samples (groups) by the values of an associated **Group Variable**. When using this option, the user should select a group variable representing groups/populations such as Zone 1, Zone2, Zone3, etc.
- When the **Options** button is clicked, the following window will be shown.

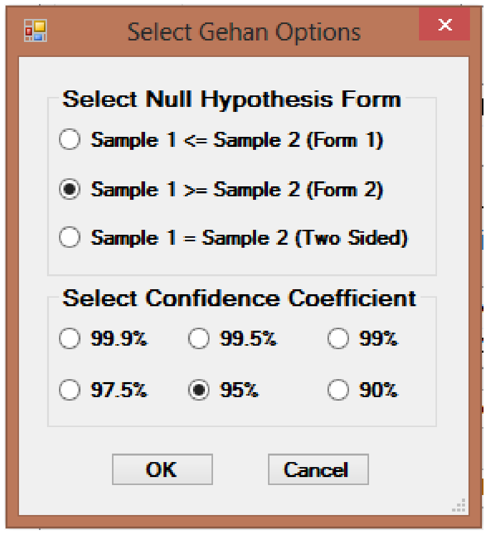

Figure 4-38. Options Related to Performing a Two-Sample Gehan Test.

- Choose the **Confidence Coefficient**. The default choice is **95%**.
- Select the form of Null Hypothesis. The default is **Sample 1 <= Sample 2 (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 Sample 1 vs. Sample 2 Comparison.

Example 4-8: Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed in the literature (Helsel 2012b). This dataset is included in the ProUCL download as Zn-Cu-two-zones-NDs.xls. This data set is used here to illustrate the Gehan two-sample test. The output sheet generated by ProUCL follows.

Table 4-11. Output for Two-Sample Gehan Test (with Nondetects)

User Selected Options		
Date/Time of Computation	ProUCL 5.17/14/2021 11:45:24 AM	
From File	Zn-Cu-two-zones-NDs.xls	
Full Precision	OFF	
Confidence Coefficient	95%	
Selected Null Hypothesis	Sample 1 Mean/Median >= Sample 2 Mean/Median (Form 2)	
Alternative Hypothesis	Sample 1 Mean/Median < Sample 2 Mean/Median	
Sample 1 Data: Zn(alluvial fan)		
Sample 2 Data: Zn(basin trough)		
Raw Statistics		
	Sample 1	Sample 2
Number of Valid Data	67	50
Number of Missing Observations	1	0
Number of Non-Detects	16	4
Number of Detect Data	51	46
Minimum Non-Detect	3	3
Maximum Non-Detect	10	10
Percent Non-detects	23.88%	8.00%
Minimum Detect	5	3
Maximum Detect	620	90
Mean of Detects	27.88	23.13
Median of Detects	11	20
SD of Detects	85.02	19.03
KM Mean	22.7	21.61
KM SD	74.03	18.77
Sample 1 vs Sample 2 Gehan Test		
H0: Mean of Sample 1 >= Mean of background		
Gehan z Test Value	-3.037	
Critical z (0.05)	-1.645	
P-Value	0.0012	
Conclusion with Alpha = 0.05		
Reject H0, Conclude Sample 1 < Sample 2		
P-Value < alpha (0.05)		

4.3.2.9 Two-Sample Tarone-Ware Test

Like the Gehan test, the T-W test can be used when many ND observations or multiple DLs may be present in the two data sets; conclusions derived using this test may not be reliable when dealing with samples of small sizes (<10). Like the Gehan test, the T-W test described below is based upon the normal approximation of the T-W statistic and should be used when enough (e.g., $m \geq 10$ and $n \geq 10$) site and background (or monitoring well) data are available.

The details of these methods can be found in the ProUCL Technical Guides (2013, 2015) and are also available in EPA (2002b, 2006a, 2009a, 2009b). It is emphasized 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 data sets may consist of NDs with multiple detection limits and observations from multiple populations (e.g., mixture samples collected from various onsite locations) and outliers.

<u>Notes:</u> As mentioned before, when one wants to use two-sample hypotheses tests on data sets with NDs, ProUCL assumes that samples from both of the groups have ND 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; the user will have to provide a ND column (with 0 or 1 entries only) for the selected variable of each of the two samples/groups. Thus, when one of the samples (e.g., site arsenic) has no ND value, the user supplies an associated ND column with all entries set 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.

Click Statistical Tests ► Two Sample Hypothesis Testing ► Two Sample ► With NDs ► Tarone-Ware

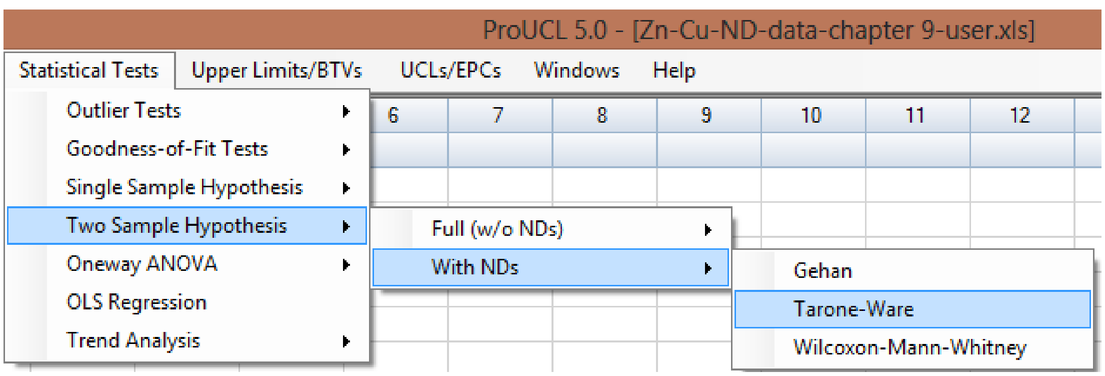

Figure 4-39. Performing a Two-Sample Tarone-Ware Test.

The **Select Variables** screen will appear.

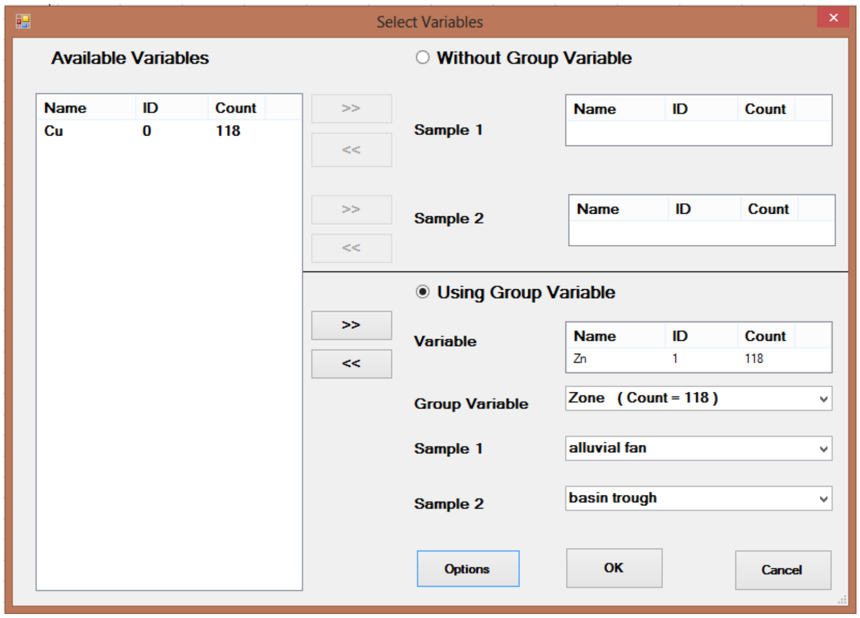

Figure 4-40. Selecting Variables for a Two-Sample Tarone-Ware Test.

- Select variable(s) from the **Select Variables** screen.
- **Without Group Variable**: This option is used when the data values of the variable (Cu) for the two data sets are given in separate columns.
- **With Group Variable**: This option is used when data values of the variable (Cu) for the two data sets are given in the same column. The values are separated into different samples (groups) by the values of an associated **Group Variable**. When using this option, the user should select a group variable/ID by clicking the arrow next to the **Group Variable** option for a drop-down list of available variables. The user selects an appropriate group variable representing the groups to be tested.
- When the **Options** button is clicked, the following window will be shown.

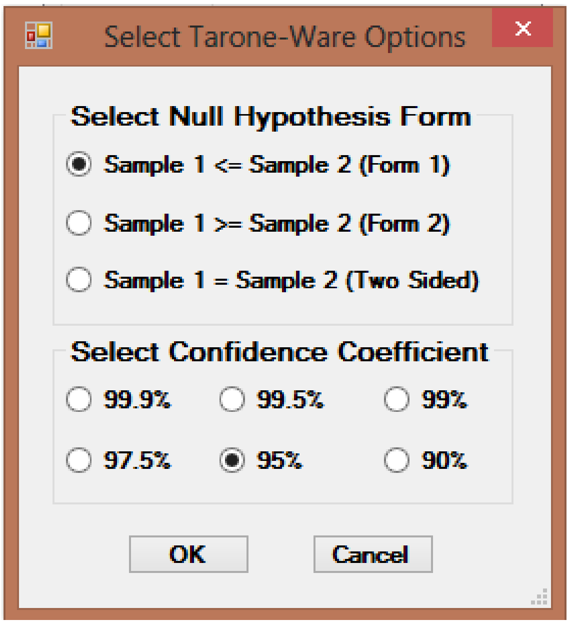

Figure 4-41. Options Related to Performing a Two-Sample Tarone-Ware Test.

Choose the **Confidence Coefficient**. The default choice is **95%**.

- Select the form of Null Hypothesis. The default is **Sample 1 <= Sample 2 (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 Sample 1 vs. Sample 2 Comparison.

Example 4-8: (continued). Consider the copper and zinc data set used earlier (Zn-Cu-two-zones-NDs.xls). The data set is used here to illustrate the T-W two-sample test. The output sheet generated by ProUCL is described as follows.

Table 4-12. Output for Two-Sample Tarone-Ware Test (with Non-detects)

Confidence Coefficient	95%
Selected Null Hypothesis	Sample 1 Mean/Median >= Sample 2 Mean/Median (Form 2)
Alternative Hypothesis	Sample 1 Mean/Median < Sample 2 Mean/Median

Sample 1 Data: Zn(alluvial fan)		
Sample 2 Data: Zn(basin trough)		
Raw Statistics		
	Sample 1	Sample 2
Number of Valid Data	67	50
Number of Missing Observations	1	0
Number of Non-Detects	16	4
Number of Detects	51	46
Minimum Non-Detect	3	3
Maximum Non-Detect	10	10
Percent Non-detects	23.88%	8.00%
Minimum Detect	5	3
Maximum Detect	620	90
Mean of Detects	27.88	23.13
Median of Detects	11	20
SD of Detects	85.02	19.03
KM Mean	22.7	21.61
KM SD	74.03	18.77
Sample 1 vs Sample 2 Tarone-Ware Test		
H0: Mean/Median of Sample 1 >= Mean/Median of Sample 2		
TW Statistic	-2.113	
TW Critical Value (0.05)	-1.645	
P-Value	0.0173	
Conclusion with Alpha = 0.05		
Reject H0, Conclude Sample 1 < Sample 2		
P-Value < alpha (0.05)		

4.4 One-way ANOVA

One-way Analysis of Variance (ANOVA) is a statistical technique that is used to compare the measures of central tendencies: means or medians of more than two populations/groups. One-way ANOVA is often used to perform inter-well comparisons in groundwater monitoring projects. Classical One-way ANOVA is a generalization of the two-sample t-test (Hogg and Craig 1995); and nonparametric ANOVA, the

Kruskal-Wallis test (Hollander and Wolfe 1999), is a generalization of the two-sample Wilcoxon Mann Whitney test. Theoretical details of One-way ANOVA are given in the ProUCL Technical Guide. **One-way ANOVA** is available under the **Statistical Tests** module of ProUCL. It is advised to use these tests on raw data in the original scale without transforming the data (e.g., using a log-transformation).

4.4.1 Classical One-Way ANOVA

Click Statistical Tests ► One-way ANOVA ► Classical

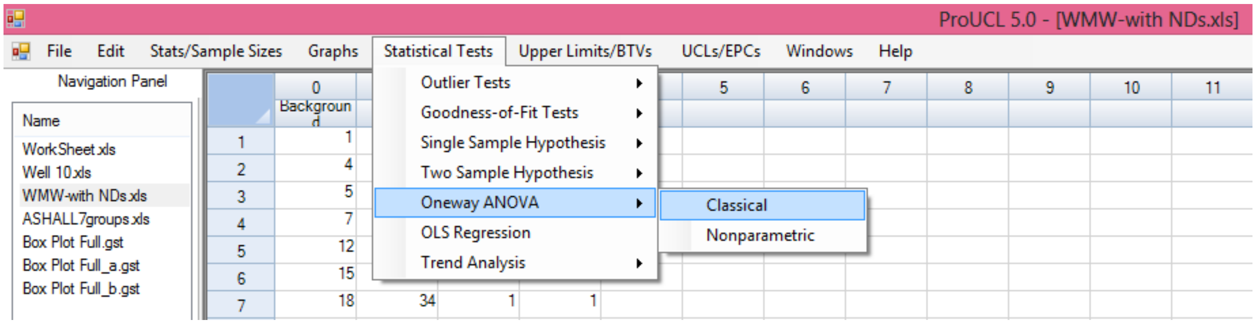

Figure 4-42. Performing Classical One-Way ANOVA.

The data file used should follow the format as shown below; the data file should consist of a group variable defining the various groups (stacked data) to be evaluated using the **One-way ANOVA** module. The **One-way ANOVA** module can process multiple variables simultaneously.

Table 4-13. Data Format for Classical One-Way ANOVA.

Well ID	Mn	As
1	460	3
1	527	5
1	579	6
1	541	1
1	518	3.5
8	1350	50
8	1770	61
8	2050	82
8	2420	91
8	1630	31
8	2810	100
9	2200	67
9	2340	82
9	2340	85
9	2420	97
9	2150	130
9	2220	189

The **Select Variables** screen will appear.

- Select the variables for testing.
- Select a **Group** variable by using the arrow under the **Group Column** option.
- Click **OK** to continue or **Cancel** to cancel the test.

Example 4-9: Consider Fisher's (1936) 3 species (groups) Iris flower data set. Fisher collected data on sepal length, sepal width, petal length and petal width for each of the 3 species. One-way ANOVA results with conclusions for the variable sepal-width (sp-width) are shown as follows:

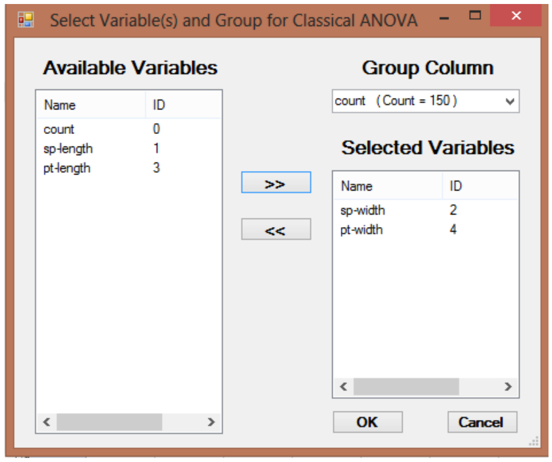

Figure 4-43. Selecting Variables for Classical One-Way ANOVA.

Table 4-14. Output for a Classical One-way ANOVA

	Classical Oneway ANOVA
Date/Time of Computation	3/26/2013 10:45:03 AM
From File	FULLIRIS-nds.xls
Full Precision	OFF

sp-width

Group	Obs	Mean	SD	Variance
1	50	3.428	0.379	0.144
2	50	2.77	0.314	0.0985
3	50	2.974	0.322	0.104
Grand Statistics (All data)	150	3.057	0.436	0.19

Classical One-Way Analysis of Variance Table

Source	SS	DOF	MS	V.R.(F Stat)	P-Value
Between Groups	11.34	2	5.672	49.16	0
Within Groups	16.96	147	0.115		
Total	28.31	149			

Pooled Standard Deviation	0.34
R-Sq	0.401

Note: A p-value <= 0.05 (or some other selected level) suggests that there are significant differences in mean/median characteristics of the various groups at 0.05 or other selected level of significance
A p-value > 0.05 (or other selected level) suggests that mean/median characteristics of the various groups are comparable.

4.4.2 Nonparametric ANOVA

Nonparametric One-way ANOVA or the Kruskal–Wallis (K-W) test is a generalization of the Mann-Whitney two-sample test. This is a nonparametric test and can be used when data from the various groups are not normally distributed.

Click Statistical Tests ► One-way ANOVA ► Nonparametric

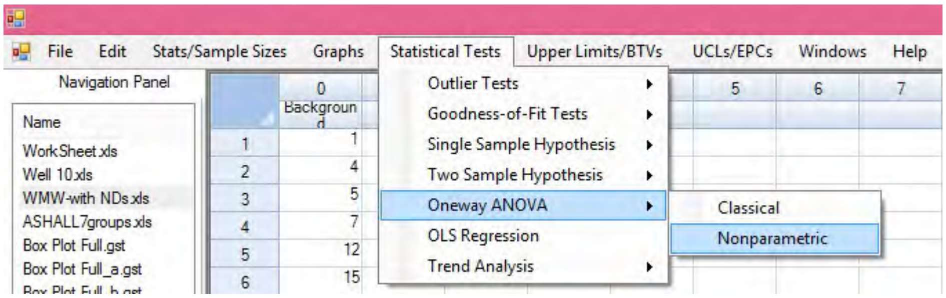

Figure 4-44. Performing One-Way Nonparametric ANOVA.

Like classical One-way ANOVA, nonparametric ANOVA also requires that the data file used should follow the data format as shown above; the data file should consist of a group variable defining the various groups to be evaluated using the **One-way ANOVA** module.

The **Select Variables** screen will appear.

- Select the variables for testing.
- Select the **Group** variable.

Click **OK** to continue or **Cancel** to cancel the test.

Example 4-9: (continued). Nonparametric One-way ANOVA results with conclusion for sepal-length (sp-length) are shown as follows.

Table 4-15. Output for a Nonparametric ANOVA

	Nonparametric Oneway ANOVA (Kruskal-Wallis Test)			
Date/Time of Computation	3/26/2013 11:11:32 AM			
From File	FULLIRIS-nds.xls			
Full Precision	OFF			
sp-length				
Group	Obs	Median	Ave Rank	Z
1	50	5	29.64	-9.142
2	50	5.9	82.65	1.425
3	50	6.5	114.2	7.716
Overall	150	5.8	75.5	
K-W (H-Stat)	DOF	P-Value	(Approx. Chisquare)	
96.76	2	0		
96.94	2	0	(Adjusted for Ties)	
Note: A p-value <= 0.05 (or some other selected level) suggests that there are significant differences in				
mean/median characteristics of the various groups at 0.05 or other selected level of significance				
A p-value > 0.05 (or other selected level) suggests that mean/median characteristics of the various groups are comparable.				

4.5 Trend Analysis

The OLS of regression and trend tests are often used to determine trends potentially present in constituent concentrations at polluted sites, especially in GW monitoring applications. More details about these tests as they apply to GW monitoring can be found in EPA (2009e). The OLS regression and two nonparametric trend tests: Mann-Kendall test and Theil-Sen test are available under the **Statistical Tests** module of ProUCL. The details of these tests can be found in Hollander and Wolfe (1999) and Draper and Smith (1998). Some time series plots, which are useful in comparing trends in analyte concentrations of multiple groups (e.g., monitoring wells), are also available in ProUCL.

The two nonparametric trend tests: M-K test and Theil-Sen test are meant to identify trends in time series data (data collected over a certain period of time such as daily, monthly, quarterly, etc.) with distinct values of the time variable (time of sampling events). If multiple observations are collected/reported at a sampling event (time), one or more pairwise slopes used in the computation of the Theil-Sen test may not be computed (become infinite). Therefore, it is suggested that the Theil-Sen test only be used on data sets with one measurement collected at each sampling event. If multiple measurements are collected at a sampling event, the user may want to use the average (or median, mode, minimum or maximum) of those measurements resulting in a time series with one measurement per sampling time event. Theil-Sen test in ProUCL has an option which can be used to average multiple observations reported for the various sampling events. The use of this option also computes M-K test statistic and OLS statistics based upon averages of multiple observations collected at the various sampling events.

A feature that was new as of ProUCL 5.1 is that in addition to slope and intercept of the nonparametric Theil-Sen (T-S) trend line, ProUCL computes residuals based upon the T-S trend line.

The trend tests in ProUCL software also assume that the user has entered data in chronological order. If the data are not entered properly in chronological order, the graphical trend displays may be meaningless. **Trend Analysis** and **OLS Regression** modules handle missing values in both response variable (e.g., analyte concentrations) as well as the sampling event variable (called independent variable in OLS).

4.5.1 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) Regression is the most advanced method available in ProUCL for trend analysis. OLS R has some underlying assumptions that need to be checked as they provide an idea how good is a regression model and how well it represents the data. These assumptions are all related to residuals, the difference between the observed and fitted value:

- Constant variance of residuals
- Independence
- Normal distribution of residuals.

More information on how to perform OLS regression and how to evaluate the assumptions is available in training:

ProUCL Utilization 2020: Part 2: Trend Analysis
https://clu-in.org/conf/tio/ProUCLAtoZ2/

Click Statistical Tests► OLS Regression.

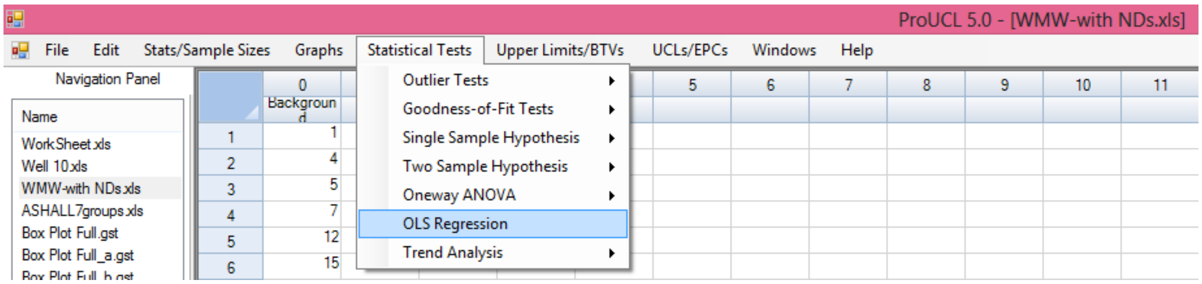

Figure 4-45. Performing OLS Regression.

The Select Regression Variables screen will appear.

- Select the **Dependent Variable** and the **Independent Variable** for the regression analysis.

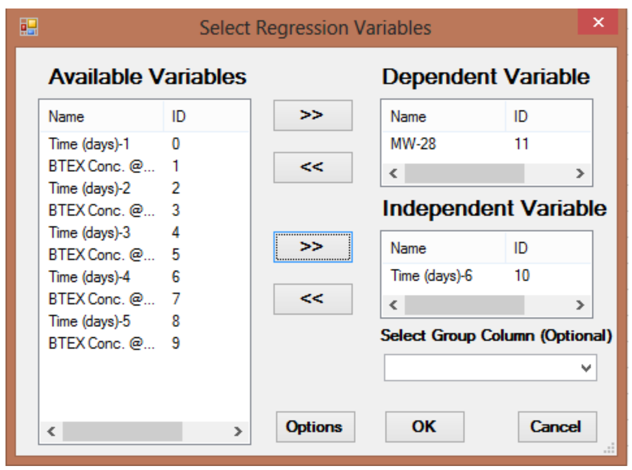

Figure 4-46. Selecting Variables for OLS Regression.

- Select a group variable (if any) by using the arrow below the **Select Group Column (Optional)**. The analysis will be performed separately for each group.
- When the **Options** button is clicked, the following options window will appear.

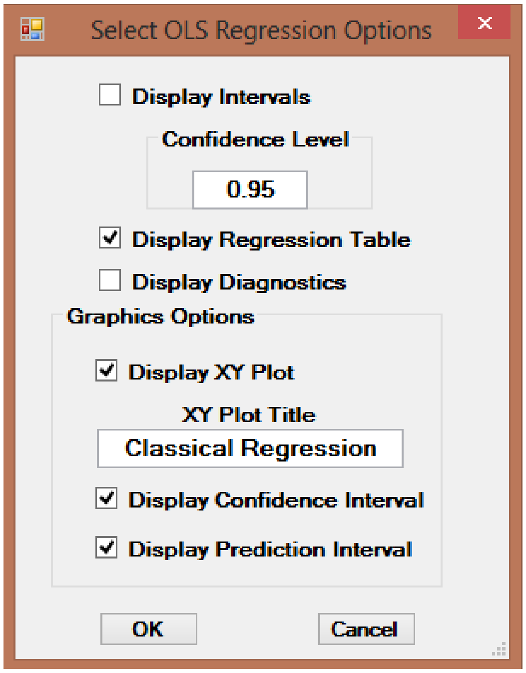

Figure 4-47. Options Related to Performing OLS Regression.

- Select **Display Intervals** for the confidence limits and the prediction limits of each observation to be displayed at the specified **Confidence Coefficient**. The interval estimates will be displayed in the output sheet.
- Select **Display Regression Table** to display Y-hat, residuals and the standardized residuals in the output sheet.
- Select "**XY Plot**" to generate a scatter plot display showing the regression line.
- Select **Confidence Interval** and **Prediction Interval** to display the confidence and the prediction bands around the regression line.
- Click on **OK** button to continue or on **Cancel** button to cancel the option.
- Click **OK** to continue or **Cancel** to cancel the OLS Regression.
- The use of the above options will display the following graph on your computer screen which can be copied using the **Copy Chart (To Clipboard)** in a Microsoft documents (e.g., word document) using the **File ►Paste** combination.
- The above options will also generate an Excel-Type output sheet. A partial output sheet is shown below following the OLS Regression Graph.

Example 4-10: Consider analyte concentrations, X collected from a groundwater (GW) monitoring well, MW-28 over a certain period of time. This dataset is included in the ProUCL download as, Trend-MW-28-Real-data.xls. The objective is to determine if there is any trend in GW concentrations, X of the MW-28. The OLS regression line with inference about slope and intercept are shown in the following figure. The slope and its associated *p*-value suggest that there is a significant downward trend in GW concentrations of MW-28.

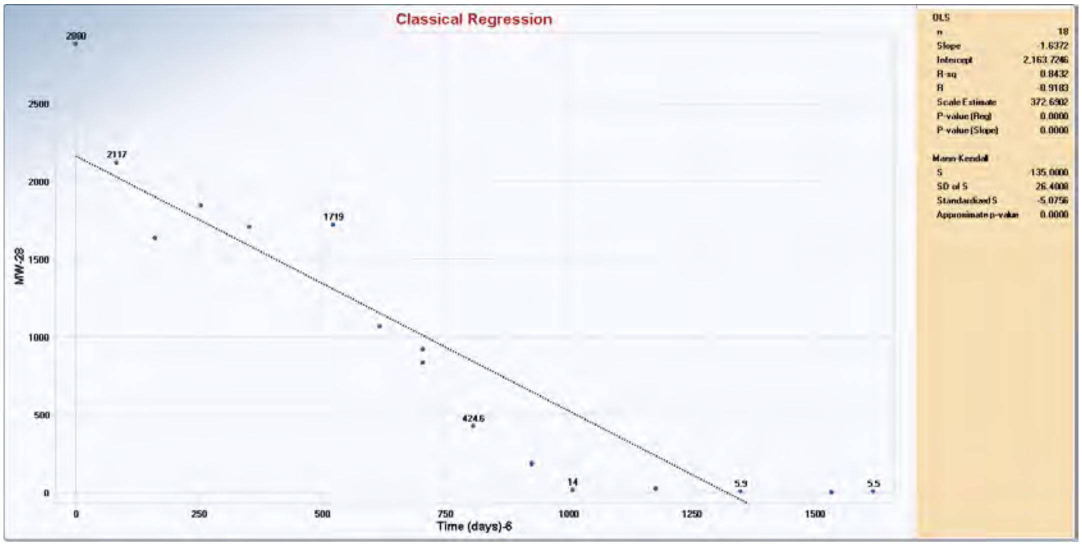

Figure 4-48. OLS Regression Graph without Regression and Prediction Intervals

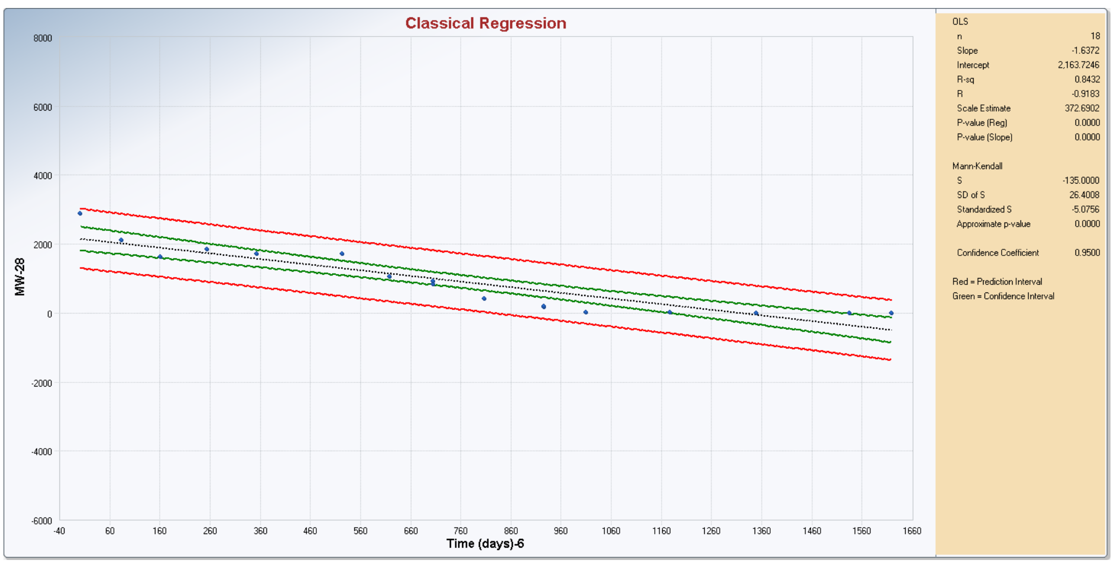

Figure 4-49. OLS Regression Graph with Regression and Prediction Intervals

Table 4-16. Partial Output of OLS Regression Analysis

	Ordinary Least Squares Linear Regression Output Sheet
User Selected Options	
Date/Time of Computation	3/27/2013 11:51:45 AM
From File	Trend-MW-data-use.xls
Full Precision	OFF
Number Reported (x-values)	18
Dependendant Variable	MW-28
Independent Variable	Time (days)-6

Regression Estimates and Inference Table

Paramater	Estimates	Std. Error	T-values	p-values
intercept	2164	165.3	13.09	5.793E-10
ïme (days)-6	-1.637	0.176	-9.276	7.7292E-8

OLS ANOVA Table

Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	11952431	1	11952431	86.05	0.0000
Error	2222368	16	138898		
Total	14174799	17			

R Square	0.843
Adjusted R Square	0.833
Sqrt(MSE) = Scale	372.7

Regression Table

Obs	Y Vector	Yhat	Residuals	Res/Scale
1	2880	2164	716.3	1.922
2	2117	2028	89.17	0.239
3	1633	1900	-267.6	-0.718
4	1845	1748	97.13	0.261
5	1706	1587	118.2	0.317
6	1719	1307	411.1	1.103
7	1065	1154	-88.55	-0.238
8	831.8	1009	-177.7	-0.477
9	920.6	1009	-88.87	-0.238

<u>Verifying Normality of Residuals:</u> As shown in the above partial output, ProUCL displays residuals including standardized residuals on the OLS output sheet. Those residuals can be imported (copying and pasting) in an excel file to assess the normality of those OLS residuals using a histogram. The parametric trend evaluations based upon the OLS slope (significant slope, confidence interval and prediction interval) are valid provided the OLS residuals are normally distributed. Therefore, it is suggested that the user assesses the normality of OLS residuals before drawing trend conclusions using a parametric test based upon the OLS slope estimate. When the assumptions are not met, one can use graphical displays and nonparametric trend tests (e.g., T-S test) to determine potential trends present in a time series data set.

4.5.2 Mann-Kendall Test

Click Statistical Tests ►Trend Analysis ► Mann-Kendall.

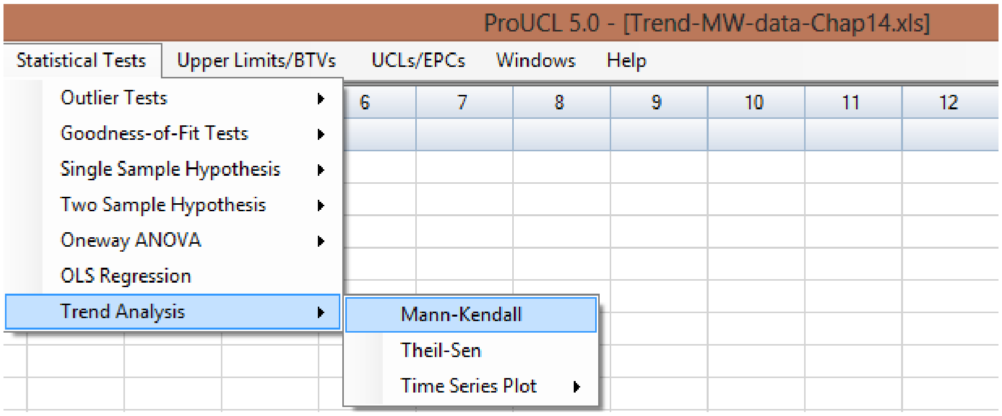

Figure 4-50. Performing the Mann-Kendall Test.

The Select Trend Event Variables screen will appear.

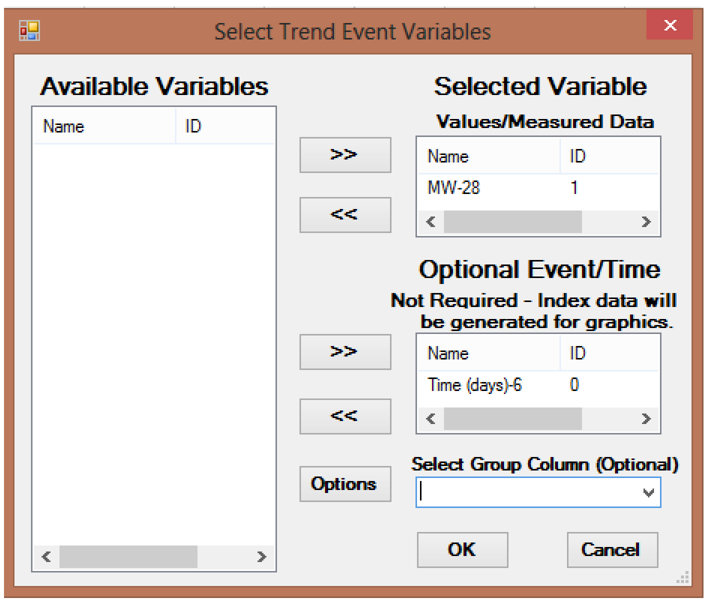

Figure 4-51. Selecting Variables for the Mann-Kendall Test.

- Select the **Event/Time** variable. This variable is optional to perform the Mann-Kendall (M-K) Test; however, for graphical display it is suggested to provide a valid Event/Time variable (continuous numerical values only, such as a Julian date). If the user wants to generate a graphical display without providing an Event/Time variable, ProUCL generates an index variable to represent sampling events, however this will not capture any influence of irregularities in sampling intervals.
- Select the **Values/Measured Data** variable to perform the trend test.
- Select a group variable (if any) by using the arrow below the **Select Group Column (Optional)**. When a group variable is chosen, the analysis is performed separately for each group represented by the group variable.
- When the **Options** button is clicked, the following window will be shown.

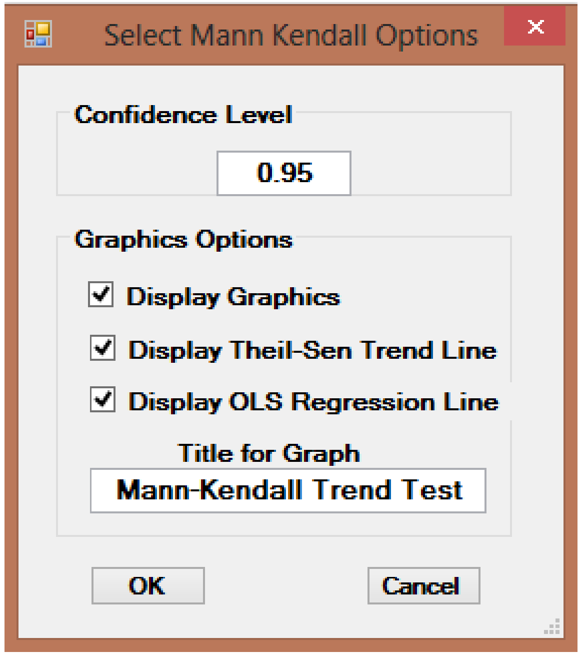

Figure 4-52. Options Related to Performing the Mann-Kendall Test.

- Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
- Select the trend lines to be displayed: **OLS Regression Line** and/or **Theil-Sen Trend Line**. If only **Display Graphics** is chosen, a time series plot will be generated.
- Click on **OK** button to continue or on **Cancel** button to cancel the option.
- Click **OK** to continue or **Cancel** to cancel the Mann-Kendall test.

Example 4-10: (Continued). The M-K test results are shown in the following figure and in the following M-K test output sheet. Based upon the M-K test, it is concluded that there is a statistically significant downward trend in GW concentrations of the MW-28.

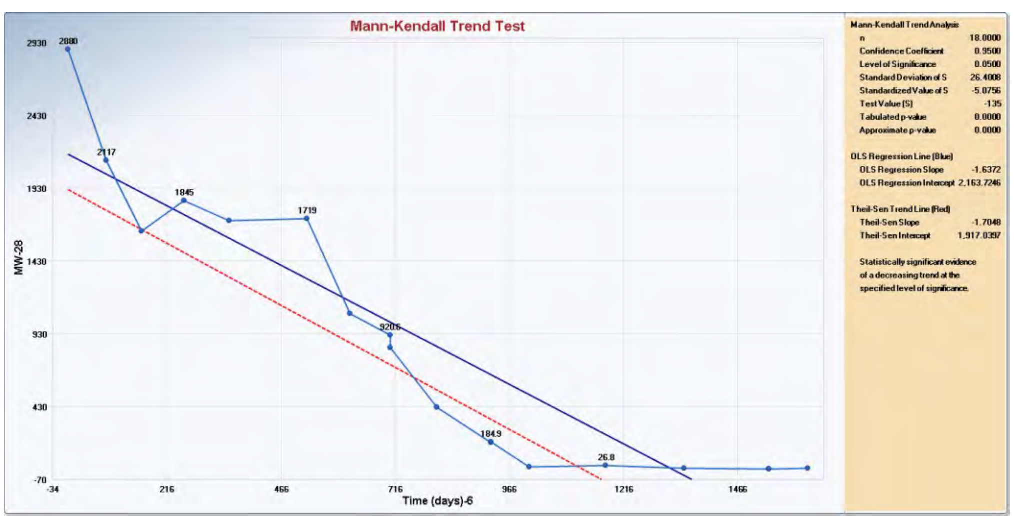

Figure 4-53. Mann Kendall Test Trend Graph displaying all Selected Options

Table 4-17. Mann-Kendall Trend Test Output Sheet

	Mann-Kendall Trend Test Analysis
User Selected Options	
Date/Time of Computation	3/27/2013 12:19:26 PM
From File	Trend-MW-data-Chap14.xls
Full Precision	OFF
Confidence Coefficient	0.95
Level of Significance	0.05
MW-28	
General Statistics	
Number of Events	18
Number Values Reported (n)	18
Minimum	1.7
Maximum	2880
Mean	864.6
Geometric Mean	174.8
Median	628.2
Standard Deviation	913.1
Mann-Kendall Test	
Test Value (S)	-135
Tabulated p-value	0
Standard Deviation of S	26.4
Standardized Value of S	-5.076
Approximate p-value	1.9313E-7
Statistically significant evidence of a decreasing trend at the specified level of significance.	

4.5.3 Theil-Sen Test

To perform the Theil-Sen test, the user is required to provide numerical values for a sampling event variable (numerical values only) as well as values of a characteristic (e.g., analyte concentrations) of interest observed at those sampling events.

Click Statistical Tests ►Trend Analysis ► Theil-Sen.

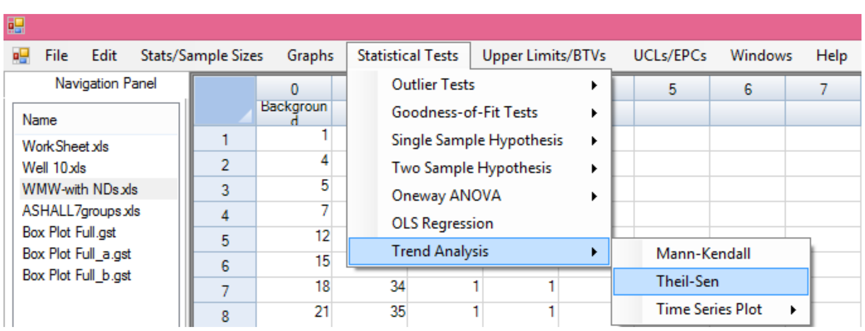

Figure 4-54. Performing the Theil-Shen Test.

The **Select Variables** screen will appear.

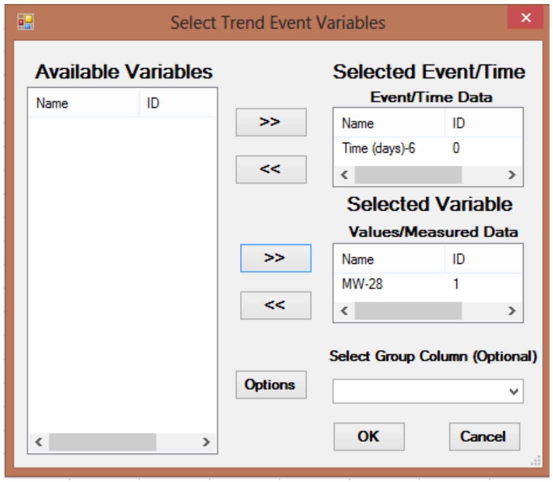

Figure 4-55. Selecting Variables for the Theil-Shen Test.

- Select an **Event/Time Data** variable.
- Select the **Values/Measured Data** variable to perform the test.
- Select a group variable (if any) by using the arrow below the **Select Group Column (Optional)**. When a group variable is chosen, the analysis is performed separately for each group represented by the group variable.
- When the **Options** button is clicked, the following window will be shown.

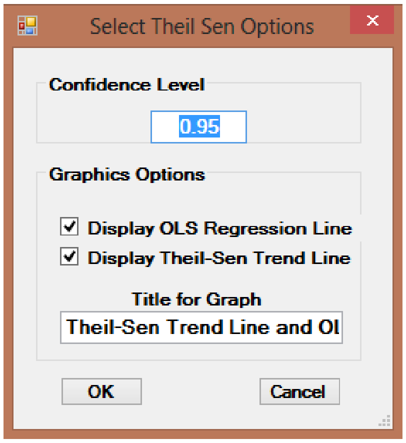

Figure 4-56. Option Related to Performing the Theil-Shen Test.

- Specify the **Confidence Level**; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is **0.95**.
- Select the trend lines to be displayed: **OLS Regression Line** and/or **Theil-Sen Trend Line.**
- Click on **OK** button to continue or on **Cancel** button to cancel the option.
- Click **OK** to continue or **Cancel** to cancel the Theil-Sen Test.

Example 4-10: (continued) The Theil-Sen test results are shown in the following figure and in the following Theil-Sen test Output Sheet. It is concluded that there is a statistically significant downward trend in GW concentrations of MW-28. Theil-Sen test results and residuals are summarized in tables following the trend graph shown below.

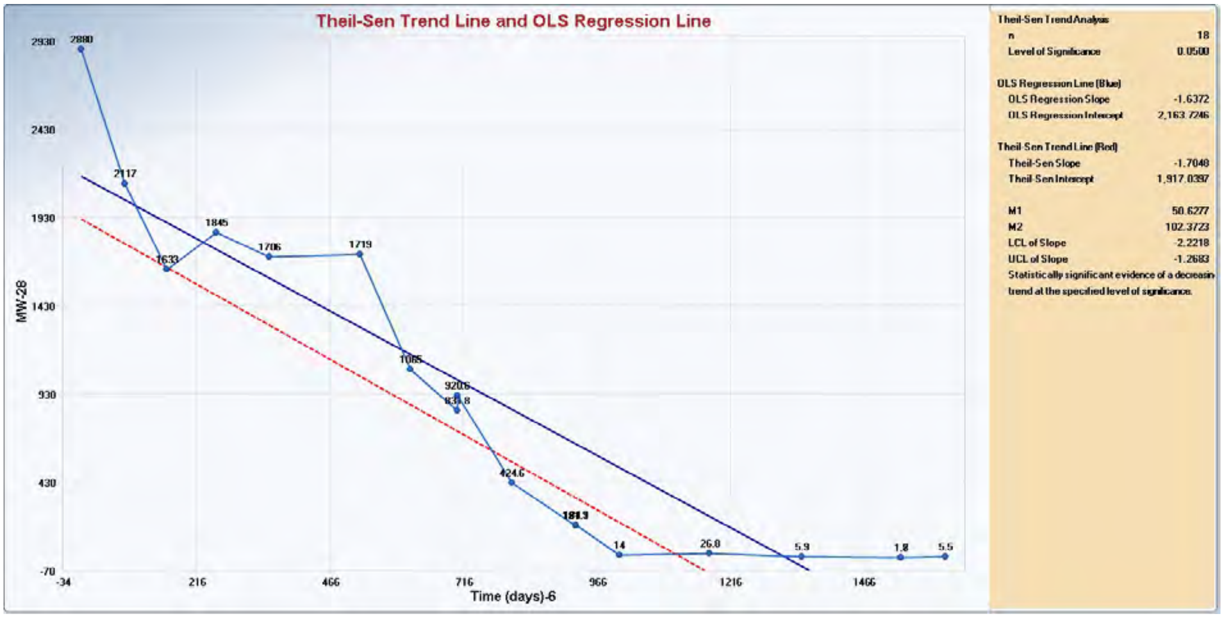

Figure 4-57. Theil-Sen Test Trend Graph displaying all Selected Options

Table 4-18. Theil-Sen Trend Test Output Sheet

Date/Time of Computation	3/27/2013 2:19:55 PM
From File	Trend-MW-data-Chap14.xls
Full Precision	OFF
Confidence Coefficient	0.95
Level of Significance	0.05

MW-28

General Statistics	
Number of Events	18
Number Values Reported (n)	18
Minimum	1.7
Maximum	2880
Mean	864.6
Geometric Mean	174.8
Median	628.2
Standard Deviation	913.1

Approximate inference for Theil-Sen Trend Test	
Mann-Kendall Statistic (S)	-137
Standard Deviation of S	26.4
Standardized Value of S	-5.151
Approximate p-value	1.2930E-7
Number of Slopes	153
Theil-Sen Slope	-1.705
Theil-Sen Intercept	1917
M2'	98.21
One-sided 95% upper limit of Slope	-1.365
95% LCL of Slope (0.025)	-2.222
95% UCL of Slope (0.975)	-1.268

Statistically significant evidence of a decreasing trend at the specified level of significance.

Theil-Sen Trend Test Estimates and Residuals

#	Events	Values	Estimates	Residuals
1	0	2880	1917	963
2	83	2117	1776	341.5
3	161	1633	1643	-10.06
4	254	1845	1484	361
5	352	1706	1317	388.7
6	523	1719	1025	693.2
7	617	1065	865.2	199.8
8	705	831.8	715.1	116.7
9	705	920.6	715.1	205.5
10	807	424.6	541.3	-116.7
11	926	181.1	338.4	-157.3
12	926	184.9	338.4	-153.5
13	1009	14	196.9	-182.9
14	1177	26.8	-89.53	116.3
15	1349	5.9	-382.8	388.7
16	1535	1.7	-699.9	701.6
17	1535	1.8	-699.9	701.7
18	1619	5.5	-843.1	848.6

Notes: As with other statistical test statistics, trend test statistics: M-K test statistic, OLS regression and Theil-Sen slopes may lead to different trend conclusions. In such instances it is suggested that the user supplements statistical conclusions with graphical displays.

Averaging of Multiple Measurements at Sampling Events: In practice, when multiple observations are collected/reported at one or more sampling events (times), one or more pairwise slopes may become infinite, resulting in a failure to compute the Theil-Sen test statistic. In such cases, the user may want to pre-process the data before using the Theil-Sen test. Specifically, to assure that only one measurement is available at each sampling event, the user pre-processes the time series data by computing average, median, mode, minimum, or maximum of the multiple observations collected at those sampling events. The Theil-Sen test in ProUCL provides the option of averaging multiple measurements collected at the various sampling events. This option also computes M-K test and OLS regression statistics using the averages of multiple measurements collected at the various sampling event. The OLS regression and M-K test can be performed on data sets with multiple measurements taken at the various sampling time events. However, often it is desirable to use the averages (or median) of measurements taken at the various sampling events to determine potential trends present in a time-series data set.

Example 4-10: (continued). The data set used in Example 8-10 (Trend-MW-28-Real-data.xls) has some sampling events where multiple observations were taken. Theil-Sen test results based upon averages of multiple observations is shown as follows. The data set is included in the ProUCL Data directory.

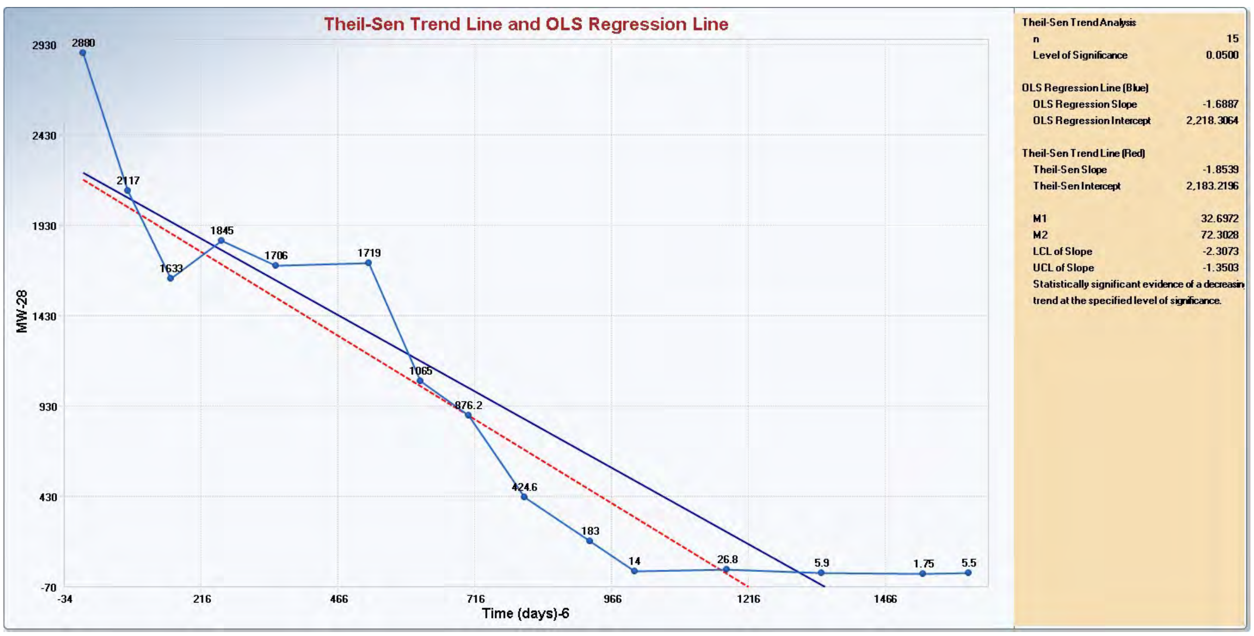

Figure 4-58. Theil-Sen Test Trend Graph displaying all Selected Options Multiple Observations Taken at Some Sampling Events Have Been Averaged

4.5.4 Time Series Plots

This option of the **Trend Analysis** module can be used to determine and compare trends in multiple groups over the same period of time.

This option is specifically useful when the user wants to compare the concentrations of multiple groups (wells) and the exact sampling event dates are not available (data only option). The user may just want to graphically compare the time-series data collected from multiple groups/wells during several quarters (every year, every 5 years, etc.). When the user wants to use this module using the **data/event** option, each group (e.g., well) defined by a group variable must have the same number of observations and should share the same sampling event values. That is the number of sampling events and values (e.g., quarter ID, year ID, etc.) for each group (well) must be the same for this option to work. However, the exact sampling dates (not needed to use this option) in the various quarters (years) do not have to be the same as long as the values of the sampling quarters/years (1,3,5,6,7,9,..) used in generating time-series plots for the various groups (wells) match. Using the geological and hydrological information, this kind of comparison may help the project team in identifying non-compliance wells (e.g., with upward trends in constituent concentrations) and associated reasons.

Click Statistical Tests ► Trend Analysis ► Time Series Plots ► (Data Only or Event/Data)

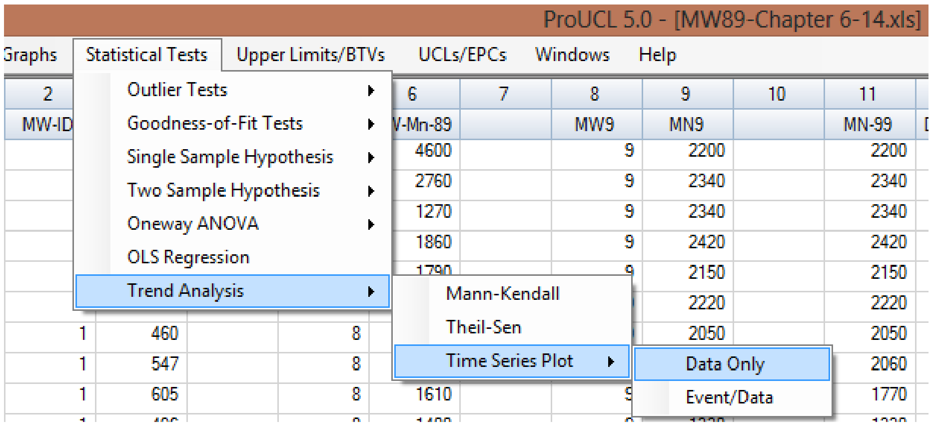

Figure 4-59. Producing Time Series Plots.

When the **Data Only** option is clicked, the following window is shown:

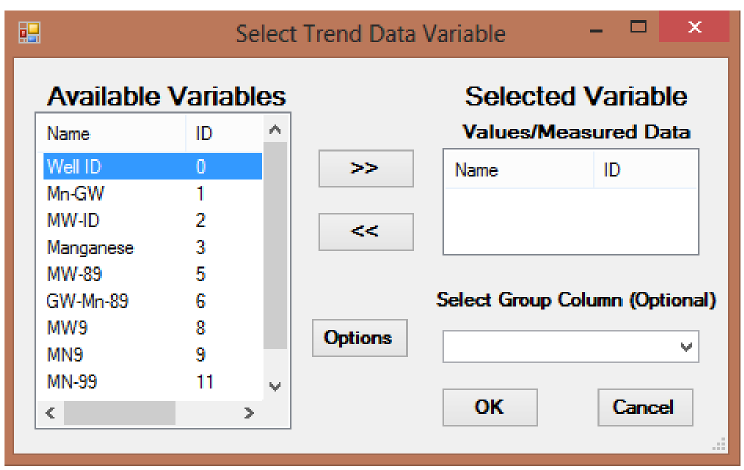

Figure 4-60. Selecting Variables for Time Series Plots – Part One.

This option is used on the measured data only. The user selects a variable with measured values which are used in generating a time series plot. The time series plot option is specifically useful when data come from multiple groups (monitoring wells during the same period of time).

- Select a group variable (is any) by using the arrow shown below the **Group Column (Optional)**.

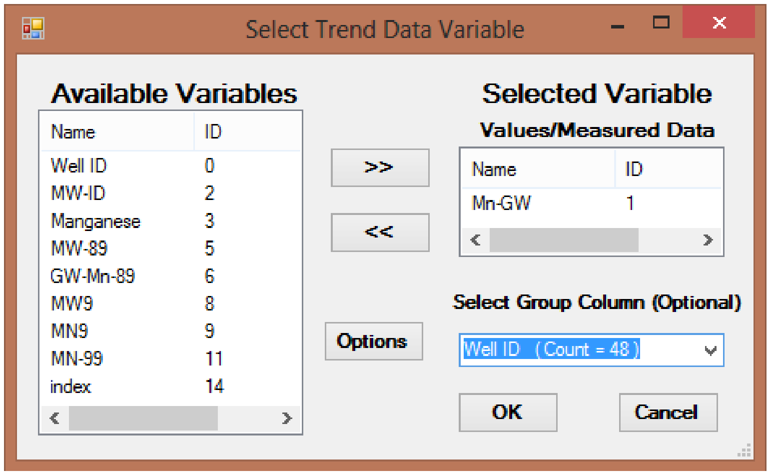

Figure 4-61. Selecting Variables for Time Series Plots – Part Two.

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

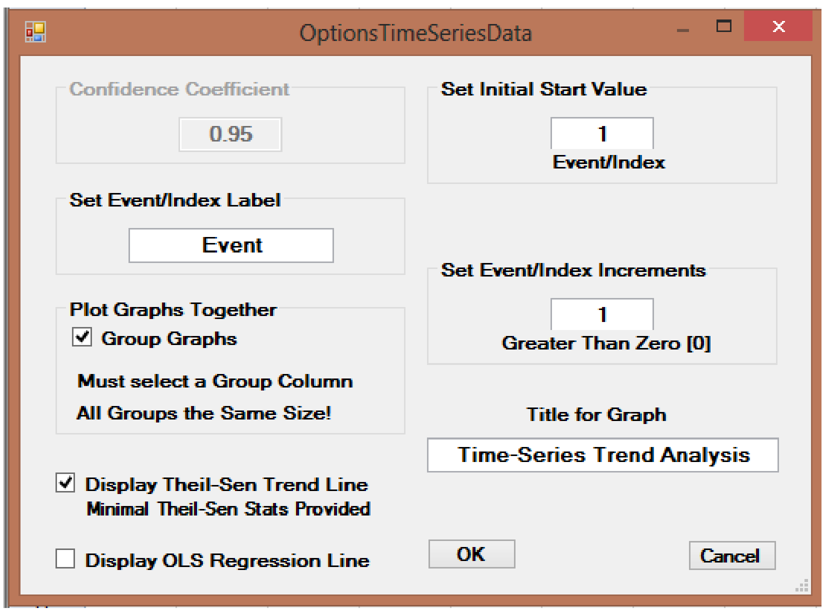

Figure 4-62. Options Related to Producing Time Series Plots.

The user can opt to display graphs for each group individually or for all groups together on the same graph by selecting the **Group Graphs** option. The user can also display the OLS line and/or the Theil-Sen line for all groups displayed on the same graph. The user may pick an initial starting value and an increment value to display the measured data. All statistics will be computed using the data displayed on the graphs (e.g., selected **Event** values).

- Input a starting value for the index of the plot using the **Set Initial Start Value**.
- Input the increment steps for the index of the plot using the **Set Index/Event Increments**.
- Specify the lines (**Regression** and/or **Theil-Sen**) to be displayed on the time series plot.

- Select **Plot Graphs Together** option for comparing the time series trends for more than one group on the same graph.

If this option is not selected but a **Group Variable** is selected, different graphs will be plotted for each group.

- Click on **OK** button to continue or on **Cancel** button to cancel the Time Series Plot.

When the **Event/Data** option is clicked, the following window is shown:

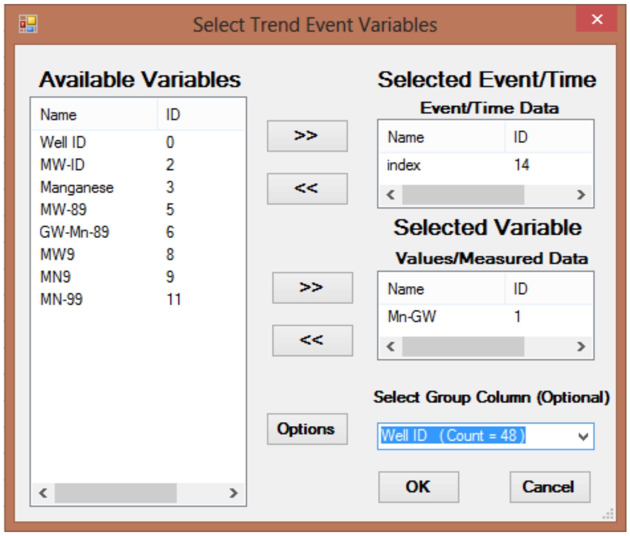

Figure 4-63 Event/Data variable selection screen.

- Select a group variable (if any) by using the arrow shown below the **Group Column (Optional)**.
- This option uses both the Measured Data and the Event/Time Data. The user selects two variables; one representing the Event/Time variable and the other representing the Measured Data values which will be used in generating a time series plot.
- Note that ProUCL has a limitation in dealing with data of a date class. If the user desires to graph the data by time, the best way to do this is to format the data in xcel to have both a readable date column and a separate column with the same data formatted as numeric. Select the numeric date as the Event/Time variable in Figure 4-63.

Example 4-11. The following example shows uranium concentrations graphed according to the date of measurement by first formatting the date data as numeric. This example uses the Trend data-with missing.xls dataset. Note that the user will have to interpret the date axis by comparing the numeric date column in the imported data table with the readable date column.

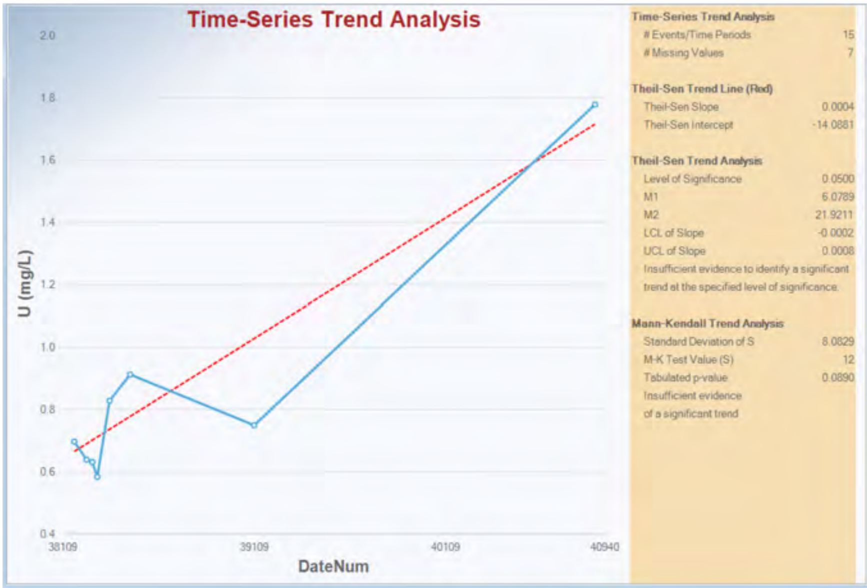

Figure 4-64. Output for a Time Series Plot – Event/Data Option by Date as a Numeric Variable.

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

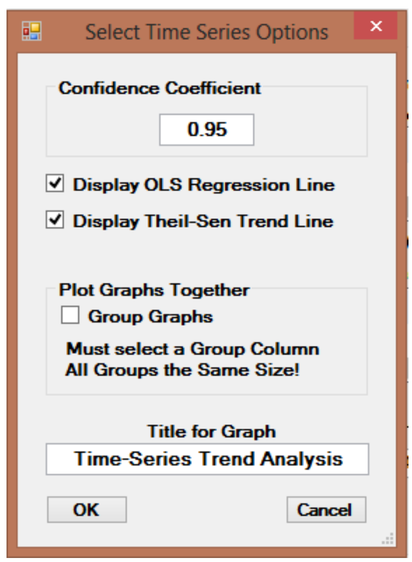

Figure 4-65 Time Series Options Screen.

The user can select to display graphs individually or together for all groups on the same graph by selecting the **Plot Graphs Together** option. The user can also display the OLS line and/or the Theil-Sen line for all groups displayed on the same graph.

- Specify the lines (**Regression** and/or **Theil-Sen**) to be displayed on the time series plot.
- Select **Plot Graphs Together** option for comparing time series trends for more than one group on the same graph.

If this option is not selected but a **Group Variable** is selected, different graphs will be plotted for each group.

- Click on **OK** button to continue or on **Cancel** button to cancel the options.
- Click **OK** to continue or **Cancel** to cancel the Time Series Plot.

<u>Notes:</u> To use this option, each group (e.g., well) defined by a group variable must have the same number of observations and should share the same sampling event values (if available). That is the sampling events (e.g., quarter ID, year ID, etc.) for each group (well) must be the same for this option to work. Specifically, the exact sampling dates within the various quarters (years) do not have to be the same as long as the sampling quarters (years) for the various wells match.

Example 4-12: The following graph has three (3) time series plots comparing manganese concentrations of the three GW monitoring wells (1 upgradient well [MW1] and 2 downgradient wells [MW8 and MW9]) over the period of 4 years (data collected quarterly). This file is included in the ProUCL download as, MW-1-8-9.xls. Some trend statistics are displayed in the side panel.

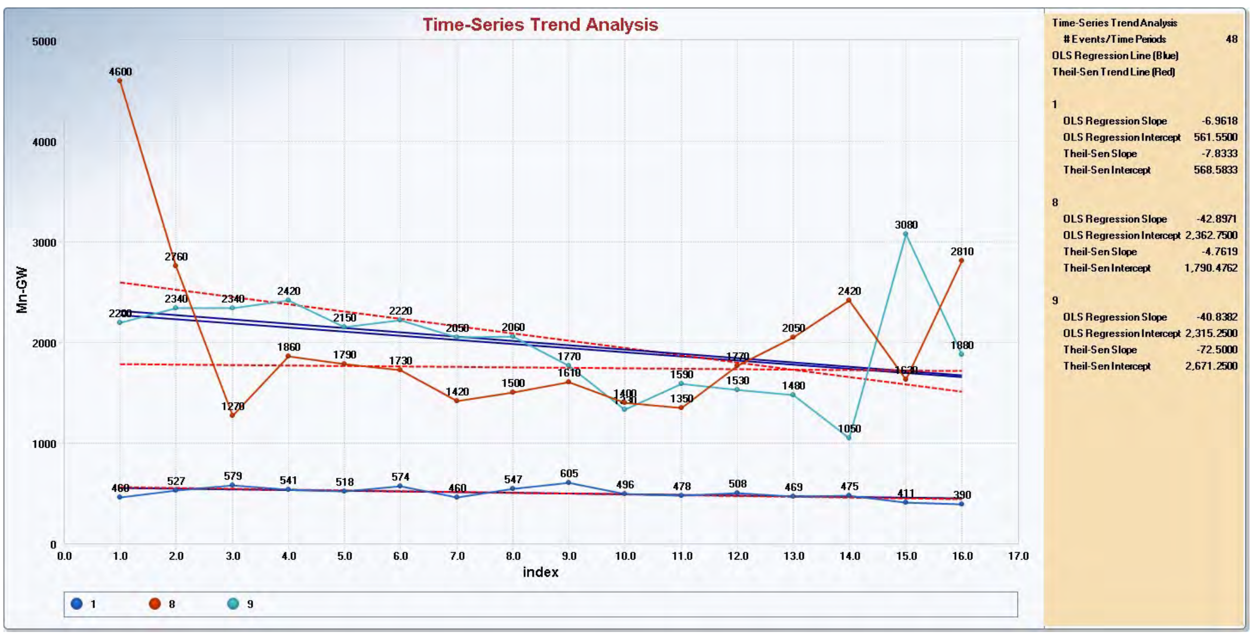

Figure 4-66. Output for a Time Series Plot – Event/Data Option by a Group Variable (1, 8, and 9)

5 Upper Tolerance Limits and Background Threshold Values (UTLs and BTVs)

This chapter illustrates the computations of parametric and nonparametric statistics and upper limits that can be used as estimates of BTVs and other not-to-exceed values. In addition to the information provided in this document users may wish to view the ProUCL training

[ProUCL Utilization 2020: Part 3: Background Level Calculations](#).

The BTV estimation methods are available for data sets with and without ND observations. Technical details about the computation of the various limits and their applicability can be found in the associated ProUCL 5.2 Technical Guide. For each selected variable, this option computes various upper limits such as UPLs, UTLs, USLs and upper percentiles to estimate the BTVs that are used in site versus background evaluations.

Two choices are available to compute background statistics for data sets:

Full (w/o NDs) – computes background statistics for uncensored full data sets without any ND observation.

- With NDs—computes background statistics for data sets consisting of detected as well as non-detected observations with multiple detection limits.

The user specifies the confidence coefficient (probability) associated with each interval estimate. ProUCL accepts a CC value in the interval [0.5, 1), 0.5 inclusive. The default choice is 0.95. For data sets with and without NDs, ProUCL can compute the following upper limits to estimate BTVs:

- Parametric and nonparametric upper percentiles.
- Parametric and nonparametric UPLs for a single observation, future or next k (>1) observations, mean of next k observations. Here future k, or next k observations may represent k observations from another population (e.g., site) different from the sampled (background) population.
- Parametric and nonparametric UTLs.
- Parametric and nonparametric USLs.

<u>Note on Computing Lower Limits:</u> In many environmental applications (e.g., groundwater monitoring), one needs to compute lower limits including: lower prediction limits (LPLs), lower tolerance limits (LTLs), or lower simultaneous limit (LSLs). At present, ProUCL does not directly compute a LPL, LTL, or a LSL. It should be noted that for data sets with and without non-detects, ProUCL outputs several intermediate results and critical values (e.g., khat, nuhat, K, d2max) needed to compute the interval estimates and lower limits. For data sets with and without NDs, except for the bootstrap methods, the same critical value (e.g., normal z value, Chebyshev critical value, or t-critical value) can be used to compute a parametric LPL, LSL, or a LTL (for samples of size >30 to be able to use Natrella's approximation in LTL) as used in the computation of a UPL, USL, or a UTL (for samples of size >30). Specifically, to compute a LPL, LSL, and LTL (n>30) the '+' sign used in the computation of the corresponding UPL, USL, and UTL (n>30) needs to be replaced by the '-' sign in the equations used to compute UPL, USL, and UTL (n>30). For specific details, the user may want to consult a statistician. For data sets *without ND* observations, the user may want to use the Scout 2008 software package (EPA 2009c) to compute the various parametric and nonparametric LPLs, LTLs (all sample sizes), and LSLs.

The examples shown in this user guide will contain non-detect values however in practice all of these calculations can be made on full datasets without non-detect values by simply clicking Upper Limits/BTVs followed by Full (w/o NDs).

5.1 Producing UTLs and BTVs

When constructing UTLs and BTVs in ProUCL the user has access to UTLs and BTVs based on the standard three distributional forms (Normal, Gamma, Lognormal) as well as Non-Parametric options. Any of these options can be selected as shown in the figure below, however the most common and useful choice is to select the All option as shown below. This will give the user access to results from all three distributional options as well as the Non-Parametric approach. The choice of which UTL/BTV to use is obviously a decision that should be made on a site and problem specific level. The following section provides an example of the UTL/BTV process in ProUCL.

Click Upper Limits/BTVs ► Chose whether or not your dataset has NDs ► All

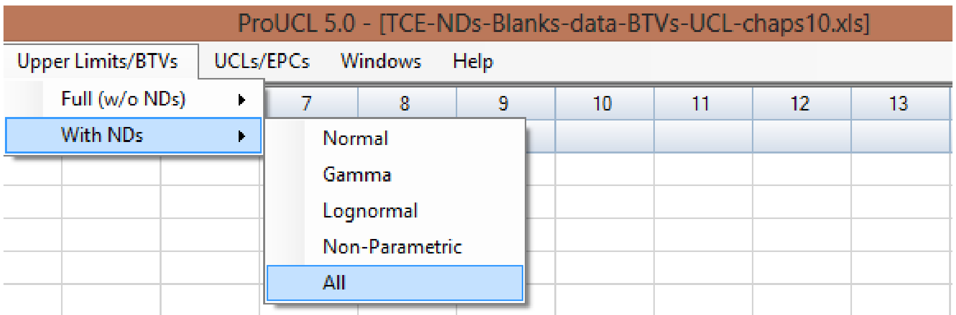

Figure 5-1. Computing Upper Limits and BTVs.

The **Select Variables** screen will appear.

- Select a variable(s) from the **Select Variables** screen.

If needed, select a group variable by clicking the arrow below the **Select Group Column (Optional)** to obtain a drop-down list of variables, and select a proper group variable.

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

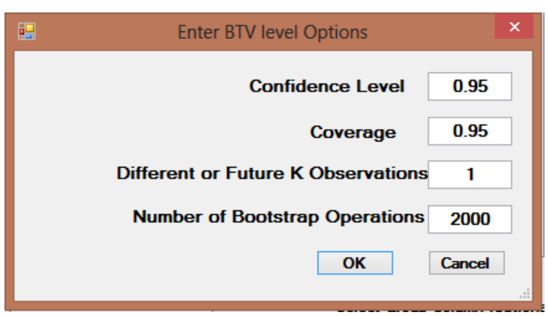

Figure 5-2. Options Related to Computing Upper Limits and BTVs.

- 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 **0.95**.
- Specify the **Future 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** to continue or on **Cancel** button to cancel the Upper Limits/BTVs option.

UTL/BTV Example 5-1: BTV estimates using the **All** option for the TCE data included in the ProUCL download as TCE-NDs-Blanks-data.xls are summarized as follows. The detected data set is of small size (*n*=8) and follows a gamma distribution. The gamma GOF Q-Q plot based upon detected data is shown in the following figure. The relevant statistics have been highlighted in the output table provided after the gamma GOF Q-Q plot.

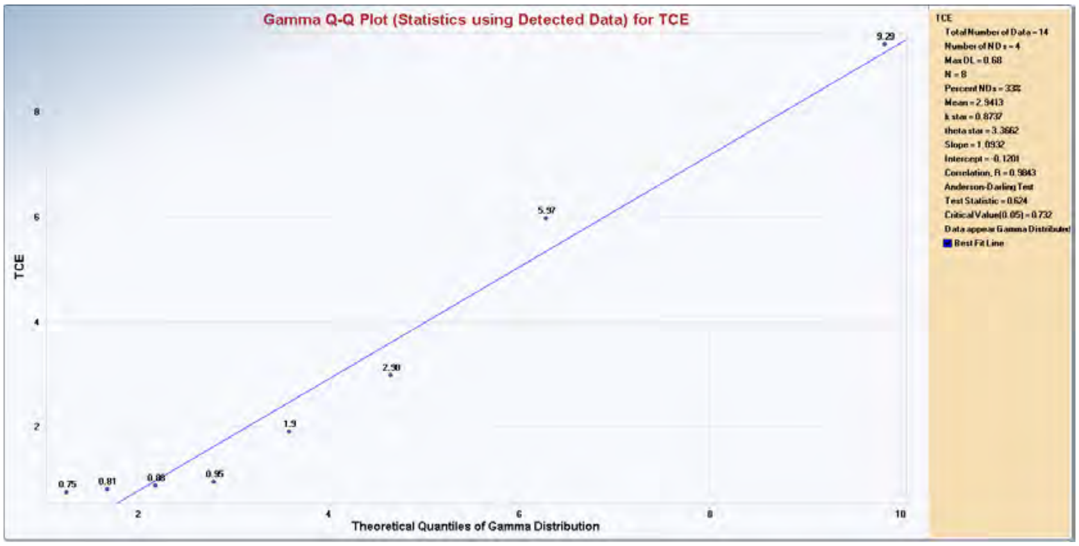

Figure 5-3. Gamma Q-Q Plot for Example 5-1.

Table 5-1. TCE - Output Screen for All BTV Estimates (Left-Censored Data Set with NDs)

General Statistics			
Total Number of Observations	12	Number of Missing Observations	2
Number of Distinct Observations	9		
Number of Detects	8	Number of Non-Detects	4
Number of Distinct Detects	8	Number of Distinct Non-Detects	1
Minimum Detect	0.75	Minimum Non-Detect	0.68
Maximum Detect	9.29	Maximum Non-Detect	0.68
Variance Detected	9.732	Percent Non-Detects	33.33%
Mean Detected	2.941	SD Detected	3.12
Mean of Detected Logged Data	0.634	SD of Detected Logged Data	0.978
Critical Values for Background Threshold Values (BTVs)			
Tolerance Factor K (For UTL)	2.736	d2max (for USL)	2.285
Normal GOF Test on Detects Only			
Shapiro Wilk Test Statistic	0.765	**Shapiro Wilk GOF Test**	
1% Shapiro Wilk Critical Value	0.749	Detected Data appear Normal at 1% Significance Level	
Lilliefors Test Statistic	0.256	**Lilliefors GOF Test**	
1% Lilliefors Critical Value	0.333	Detected Data appear Normal at 1% Significance Level	
Detected Data appear Normal at 1% Significance Level			
Kaplan Meier (KM) Background Statistics Assuming Normal Distribution			
KM Mean	2.188	KM SD	2.61
95% UTL95% Coverage	9.329	95% KM UPL (t)	7.067
90% KM Percentile (z)	5.533	95% KM Percentile (z)	6.481
90% KM Percentile (z)	5.533	95% KM Percentile (z)	6.481
99% KM Percentile (z)	8.26	95% KM USL	8.152
DL/2 Substitution Background Statistics Assuming Normal Distribution			
Mean	2.074	SD	2.799
95% UTL95% Coverage	9.732	95% UPL (t)	7.306
90% Percentile (z)	5.661	95% Percentile (z)	6.678
99% Percentile (z)	8.585	95% USL	8.469
DL/2 is not a recommended method. DL/2 provided for comparisons and historical reasons			
Gamma GOF Tests on Detected Observations Only			
A-D Test Statistic	0.624	**Anderson-Darling GOF Test**	
5% A-D Critical Value	0.732	Detected data appear Gamma Distributed at 5% Significance Level	
K-S Test Statistic	0.274	**Kolmogorov-Smirnov GOF**	
5% K-S Critical Value	0.3	Detected data appear Gamma Distributed at 5% Significance Level	
Detected data appear Gamma Distributed at 5% Significance Level			
Gamma Statistics on Detected Data Only			
k hat (MLE)	1.265	k star (bias corrected MLE)	0.874
Theta hat (MLE)	2.326	Theta star (bias corrected MLE)	3.366
nu hat (MLE)	20.23	nu star (bias corrected)	13.98
MLE Mean (bias corrected)	2.941		
MLE Sd (bias corrected)	3.147	95% Percentile of Chisquare (2kstar)	5.492
Gamma ROS Statistics using Imputed Non-Detects			
GROS may not be used when data set has > 50% NDs with many tied observations at multiple DLs			
GROS may not be used when kstar of detects is small such as <1.0, especially when the sample size is small (e.g., <15-20)			
For such situations, GROS method may yield incorrect values of UCLs and BTVs			
This is especially true when the sample size is small.			
For gamma distributed detected data, BTVs and UCLs may be computed using gamma distribution on KM estimates			
Minimum	0.01	Mean	1.964
Maximum	9.29	Median	0.845
SD	2.877	CV	1.465
k hat (MLE)	0.372	k star (bias corrected MLE)	0.335
Theta hat (MLE)	5.274	Theta star (bias corrected MLE)	5.865
nu hat (MLE)	8.938	nu star (bias corrected)	8.037
MLE Mean (bias corrected)	1.964	MLE Sd (bias corrected)	3.394
95% Percentile of Chisquare (2kstar)	2.956	90% Percentile	5.709
95% Percentile	8.668	99% Percentile	16.26

Table 5-1 (continued) TCE - Output Screen for All BTV Estimates (Left-Censored Data Set with NDs)

The following statistics are computed using Gamma ROS Statistics on Imputed Data					
Upper Limits using Wilson Hilferty (WH) and Hawkins Wixley (HW) Methods					
	WH	HW		WH	HW
95% Approx. Gamma UTL with 95% Coverage	19.62	27.19	95% Approx. Gamma UPL	9.793	11.66
95% Gamma USL	13.95	17.89			

Estimates of Gamma Parameters using KM Estimates			
Mean (KM)	2.188	SD (KM)	2.61
Variance (KM)	6.813	SE of Mean (KM)	0.806
k hat (KM)	0.702	k star (KM)	0.582
nu hat (KM)	16.86	nu star (KM)	13.98
theta hat (KM)	3.115	theta star (KM)	3.757
80% gamma percentile (KM)	3.606	90% gamma percentile (KM)	5.728
95% gamma percentile (KM)	7.957	99% gamma percentile (KM)	13.36

The following statistics are computed using gamma distribution and KM estimates					
Upper Limits using Wilson Hilferty (WH) and Hawkins Wixley (HW) Methods					
	WH	HW		WH	HW
95% Approx. Gamma UTL with 95% Coverage	11.34	11.95	95% Approx. Gamma UPL	6.88	6.896
95% KM Gamma Percentile	5.955	5.902	95% Gamma USL	8.836	9.063

Lognormal GOF Test on Detected Observations Only			
Shapiro Wilk Test Statistic	0.865	**Shapiro Wilk GOF Test**	
10% Shapiro Wilk Critical Value	0.851	Detected Data appear Lognormal at 10% Significance Level	
Lilliefors Test Statistic	0.258	**Lilliefors GOF Test**	
10% Lilliefors Critical Value	0.265	Detected Data appear Lognormal at 10% Significance Level	
Detected Data appear Lognormal at 10% Significance Level			
Background Lognormal ROS Statistics Assuming Lognormal Distribution Using Imputed Non-Detects			
Mean in Original Scale	2.018	Mean in Log Scale	-0.214
SD in Original Scale	2.838	SD in Log Scale	1.512
95% UTL95% Coverage	50.54	95% BCA UTL95% Coverage	9.29
95% Bootstrap (%) UTL95% Coverage	9.29	95% UPL (t)	13.63
90% Percentile (z)	5.606	95% Percentile (z)	9.71
99% Percentile (z)	27.2	95% USL	25.55
Statistics using KM estimates on Logged Data and Assuming Lognormal Distribution			
KM Mean of Logged Data	0.294	95% KM UTL (Lognormal)95% Coverage	15.25
KM SD of Logged Data	0.888	95% KM UPL (Lognormal)	7.06
95% KM Percentile Lognormal (z)	5.784	95% KM USL (Lognormal)	10.21
Background DL/2 Statistics Assuming Lognormal Distribution			
Mean in Original Scale	2.074	Mean in Log Scale	0.0631
SD in Original Scale	2.799	SD in Log Scale	1.149
95% UTL95% Coverage	24.69	95% UPL (t)	9.12
90% Percentile (z)	4.643	95% Percentile (z)	7.048
99% Percentile (z)	15.42	95% USL	14.7
DL/2 is not a Recommended Method. DL/2 provided for comparisons and historical reasons.			
Nonparametric Distribution Free Background Statistics			
Data appear to follow a Discernible Distribution			
Nonparametric Upper Limits for BTVs(no distinction made between detects and nondetects)			
Order of Statistic, r	12	95% UTL with95% Coverage	9.29
Approx, f used to compute achieved CC	0.632	Approximate Actual Confidence Coefficient achieved by UTL	0.46
Approximate Sample Size needed to achieve specified CC	59	95% UPL	9.29
95% USL	9.29	95% KM Chebyshev UPL	14.03

Note: The use of USL tends to yield a conservative estimate of BTV, especially when the sample size starts exceeding 20.
Therefore, one may use USL to estimate a BTV only when the data set represents a background data set free of outliers
and consists of observations collected from clean unimpacted locations.
The use of USL tends to provide a balance between false positives and false negatives provided the data
represents a background data set and when many onsite observations need to be compared with the BTV.

UTL/BTV Example 5-1 Conclusion:

The detected data follow a normal distribution based upon the S-W and Lilliefors test. Since the detected data set is of small size (=8), the normal GOF conclusion may be suspect. The detected data also follow gamma as well as a lognormal distribution. It is worth noting in a case that when data follow both Gamma and Lognormal distributions but not a Normal distribution, it is generally preferable to use a Gamma distribution due to instability that can arise due to excessively long right tails for some Lognormal distributions. The various upper limits using Gamma ROS and Lognormal ROS methods and Gamma and Lognormal distribution on KM estimates are summarized as follows.

There are several NDs reported with a low detection limit of 0.68, therefore, GROS method may yield infeasible negative imputed values. Therefore, the use of a gamma distribution on KM estimates is preferred for computing the BTV estimates. The gamma KM UTL95-95 (HW) =11.34, and gamma KM UTL95-95 (WH) = 11.95. Any one of these two limits can be used to estimate the BTV.

Table 5-2. Summary of Upper Limits Computed using Gamma and Lognormal Distribution of Detected Data Sample Size = 12, No. of NDs = 4, % NDs = 33.33, Max Detect = 9.29

	Gamma Distribution		Lognormal Distribution	
Upper Limits	Result	Reference/ Method of Calculation	Result	Reference/ Method of Calculation
Mean (KM)	2.188	--	0.29	Logged
Mean (ROS)	1.964	--	2.018	--
UPL95 (ROS)	9.79	WH- ProUCL(ROS)	13.63	Helsel (2012b), EPA (2009e)- LROS
UTL95-95 (ROS)	19.62	WH- ProUCL(ROS)	50.54	Helsel (2012b), EPA (2009e)- LROS
UPL95 (KM)	6.88	WH - ProUCL (KM- Gamma)	7.06	KM-Lognormal EPA (2009e)
UTL95-95 (KM)	11.34	WH - ProUCL (KM- Gamma)	15.25	KM- Lognormal EPA(2009e)

Note: All computations have been performed using the ProUCL software. In the above table, methods proposed/described in the literature have been cited in the Reference Method of Calculation column. The statistics summarized above demonstrate the merits of using the gamma distribution based upper limits to estimate decision parameters (BTVs) of interest. These results summarized in the above tables suggest that the use of a gamma distribution cannot be dismissed just because it is easier to use a lognormal distribution to model skewed data sets as stated by some practitioners.

6 Upper Confidence Limits and Exposure Point Concentrations (UCLs and EPCs)

Several parametric and nonparametric UCL methods for full-uncensored and left-censored data sets consisting of ND observations with multiple DLs are available in ProUCL . Methods such as the Kaplan-Meier (KM) and regression on order statistics (ROS) methods incorporated in ProUCL can handle multiple detection limits. For details regarding the goodness-of-fit tests and UCL computation methods available in ProUCL, consult the ProUCL Technical Guides, Singh, Singh, and Engelhardt (1997); Singh, Singh, and Iaci (2002); and Singh, Maichle, and Lee (2006).

In addition to the information presented in this document users may wish to view information on producing UCL estimates presented in the third part of the ProUCL 2020 webinar series, located here ProUCL Utilization 2020: Part 3: Background Level Calculations.

In ProUCL, two choices are available for computing UCL statistics:

- Full (w/o NDs): Computes UCLs for full-uncensored data sets without any non-detects.
- With NDs: Computes UCLs for data sets consisting of ND observations with multiple DLs or reporting limits (RLs).

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 a confidence level; a number in the interval [0.5, 1), 0.5 inclusive. The default choice is 0.95.
- The program computes requisite parametric UCLs based on GOF test results.
- The program computes several nonparametric UCLs using the CLT, adjusted 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.

Unless utilizing the ‘All’ option, the user is responsible for selecting an appropriate choice for the data distribution: normal, gamma, lognormal, or nonparametric. It is desirable that the user determines data distribution using the Goodness-of-Fit test option prior to using the UCL option. The UCL output sheet also informs the user if data are normal, gamma, lognormal, or a non-discernible distribution. The 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, ProUCL outputs most of the relevant UCLs available in ProUCL. 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 90%, 95%, 97.5%, and 99% Land's statistic-based H-UCL of the mean. For all other methods, ProUCL can compute a UCL for any confidence coefficient (CC) in the interval (0.5, 1.0), 0.5 inclusive. If you have selected a distribution, then ProUCL will provide a recommended UCL method for 0.95, confidence level. Even though ProUCL can compute UCLs for any confidence coefficient level in the interval (0.5, 1.0), the recommendations are provided only for 95% UCL; as EPC term is estimated by a 95% UCL of the mean.

Notes: Like all other methods, the user may identify a few low probability (coming from extreme tails) outlying observations that may be present in the data set. Refer to Section 4.1 for guidance on dealing with extreme values.

Note on Computing Lower Confidence Limits (LCLs) of the Mean: In several environmental applications, one needs to compute a LCL of the population mean. At present, ProUCL does not directly compute LCLs of mean. It should be pointed out that for data sets with and without NDs, except for the bootstrap methods, gamma distribution (e.g., samples of sizes <50), and H-statistic based LCL of mean, the same critical value (e.g., normal z value, Chebyshev critical value, or t-critical value) are used to compute a LCL of mean as used in the computation of the UCL of mean. Specifically, to compute a LCL, the '+' sign used in the computation of the corresponding UCL needs to be replaced by the '-' sign in the equation used to compute that UCL (excluding gamma, lognormal H-statistic, and bootstrap methods). For specific details, the user may want to consult a statistician. For data sets *without non-detect* observations, the user may want to use the Scout 2008 software package (EPA 2009c) to directly compute the various parametric and nonparametric LCLs of mean.

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 values is useful when using bootstrap methods. Specifically, it is not desirable to use bootstrap methods on data sets with only a few distinct values.

6.1 Producing UCLs and EPCs

Click **UCLs/EPCs** ► Chose whether or not your dataset has NDs

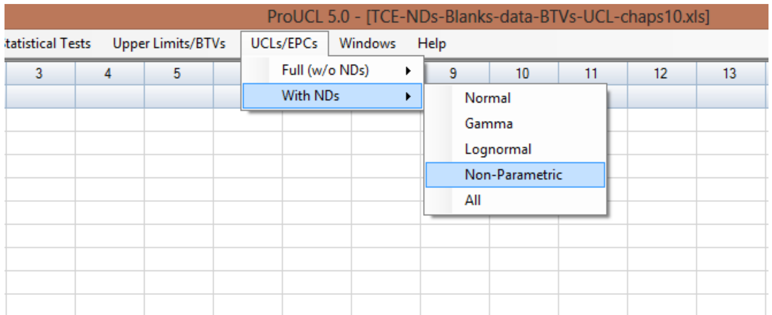

Figure 6-1. Computing UCLs.

Choose the Normal, Gamma, Lognormal, Non-Parametric, or All option.

The **Select Variables** screen will appear.

- Select a variable(s) from the **Select Variables** screen.
- If needed, select a group variable by clicking the arrow below the **Select Group Column (Optional)** 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.

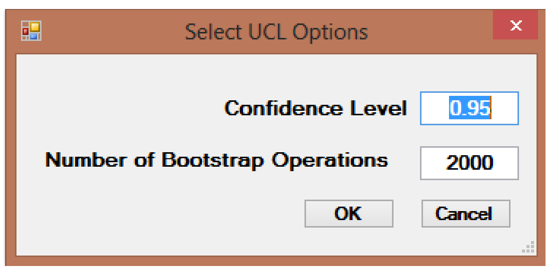

Figure 6-2. Options Related to Computing UCLs.

- 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** to continue or on **Cancel** to cancel the selected UCL computation option.

Example 6-1. This real data set of size n=55 with 18.8% NDs (=10) is also used in Chapters 4 and 5 of the ProUCL Technical Guide. This dataset is included in the ProUCL download as, TRS-Real-data-with-NDs.xls. The minimum detected value is 5.2 and the largest detected value is 79000, *sd* of detected logged data is 2.79 suggesting that the data set is highly skewed. The detected data follow a gamma as well as a lognormal distribution. It is noted that GROS data set with imputed values follows a gamma distribution and LROS data set with imputed values follows a lognormal distribution (results not included). The lognormal Q-Q plot based upon detected data is shown in the following figure. The various UCL output sheets: normal, nonparametric, gamma, and lognormal generated by ProUCL are summarized in tables following the lognormal Q-Q plot on detected data. The main results have been highlighted in the output screen provided after the lognormal GOF Q-Q plot.

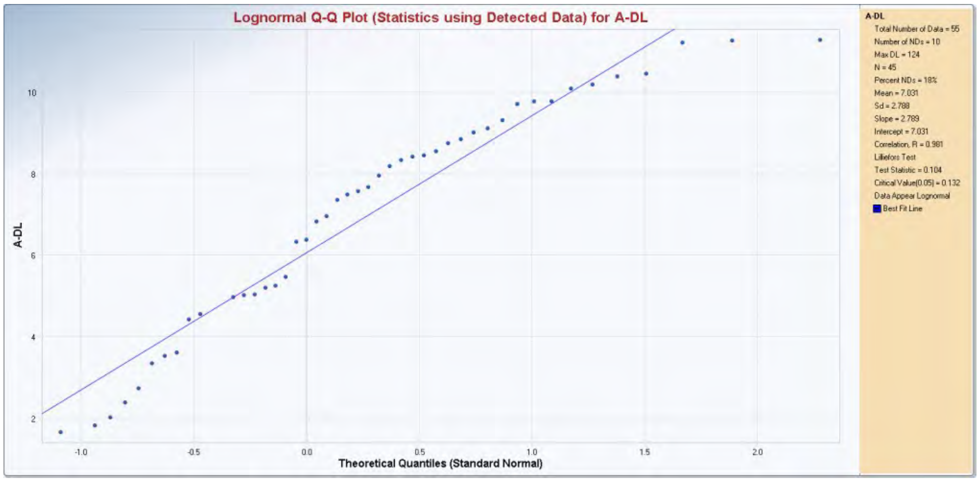

Figure 6-3. Lognormal Q-Q Plot Example.

Table 6-1. Output Screen for UCLs based upon Normal, Lognormal, and Gamma Distributions (of Detects)

A-DL

General Statistics			
Total Number of Observations	55	Number of Distinct Observations	53
Number of Detects	45	Number of Non-Detects	10
Number of Distinct Detects	45	Number of Distinct Non-Detects	8
Minimum Detect	5.2	Minimum Non-Detect	3.8
Maximum Detect	79000	Maximum Non-Detect	124
Variance Detects	3.954E+8	Percent Non-Detects	18.18%
Mean Detects	10556	SD Detects	19886
Median Detects	1940	CV Detects	1.884
Skewness Detects	2.632	Kurtosis Detects	6.496
Mean of Logged Detects	7.031	SD of Logged Detects	2.788
Normal GOF Test on Detects Only			
Shapiro Wilk Test Statistic	0.575	**Shapiro Wilk GOF Test**	
1% Shapiro Wilk Critical Value	0.926	Detected Data Not Normal at 1% Significance Level	
Lilliefors Test Statistic	0.298	**Lilliefors GOF Test**	
1% Lilliefors Critical Value	0.153	Detected Data Not Normal at 1% Significance Level	
Detected Data Not Normal at 1% Significance Level			
Kaplan-Meier (KM) Statistics using Normal Critical Values and other Nonparametric UCLs			
KM Mean	8638	KM Standard Error of Mean	2488
90KM SD	18246	95% KM (BCA) UCL	12625
95% KM (t) UCL	12802	95% KM (Percentile Bootstrap) UCL	12698
95% KM (z) UCL	12731	95% KM Bootstrap t UCL	15088
90% KM Chebyshev UCL	16102	95% KM Chebyshev UCL	19483
97.5% KM Chebyshev UCL	24176	99% KM Chebyshev UCL	33394
Gamma GOF Tests on Detected Observations Only			
A-D Test Statistic	0.591	**Anderson-Darling GOF Test**	
5% A-D Critical Value	0.86	Detected data appear Gamma Distributed at 5% Significance Level	
K-S Test Statistic	0.115	**Kolmogorov-Smirnov GOF**	
5% K-S Critical Value	0.143	Detected data appear Gamma Distributed at 5% Significance Level	
Detected data appear Gamma Distributed at 5% Significance Level			
Gamma Statistics on Detected Data Only			
k hat (MLE)	0.307	k star (bias corrected MLE)	0.302
Theta hat (MLE)	34333	Theta star (bias corrected MLE)	34980
nu hat (MLE)	27.67	nu star (bias corrected)	27.16
Mean (detects)	10556		
Gamma ROS Statistics using Imputed Non-Detects			
GROS may not be used when data set has > 50% NDs with many tied observations at multiple DLs			
GROS may not be used when kstar of detects is small such as <1.0, especially when the sample size is small (e.g., <15-20)			
For such situations, GROS method may yield incorrect values of UCLs and BTVs			
This is especially true when the sample size is small.			
For gamma distributed detected data, BTVs and UCLs may be computed using gamma distribution on KM estimates			
Minimum	0.01	Mean	8637
Maximum	79000	Median	588
SD	18415	CV	2.132
k hat (MLE)	0.18	k star (bias corrected MLE)	0.183
Theta hat (MLE)	47915	Theta star (bias corrected MLE)	47314
nu hat (MLE)	19.83	nu star (bias corrected)	20.08
Adjusted Level of Significance (β)	0.0456		
Approximate Chi Square Value (20.08, α)	10.91	Adjusted Chi Square Value (20.08, β)	10.73
95% Gamma Approximate UCL	15896	95% Gamma Adjusted UCL	16167

Table 6-1 (continued). Output Screen for UCLs based upon Normal, Lognormal, and Gamma Distributions (of Detects)

Estimates of Gamma Parameters using KM Estimates			
Mean (KM)	8638	SD (KM)	18246
Variance (KM)	3.329E+8	SE of Mean (KM)	2488
k hat (KM)	0.224	k star (KM)	0.224
nu hat (KM)	24.66	nu star (KM)	24.64
theta hat (KM)	38539	theta star (KM)	38557
80% gamma percentile (KM)	12016	90% gamma percentile (KM)	26081
95% gamma percentile (KM)	43162	99% gamma percentile (KM)	89358
Gamma Kaplan-Meier (KM) Statistics			
Approximate Chi Square Value (24.64, α)	14.34	Adjusted Chi Square Value (24.64, β)	14.13
95% KM Approximate Gamma UCL	14846	95% KM Adjusted Gamma UCL	15069
Lognormal GOF Test on Detected Observations Only			
Shapiro Wilk Test Statistic	0.939	**Shapiro Wilk GOF Test**	
10% Shapiro Wilk Critical Value	0.953	Detected Data Not Lognormal at 10% Significance Level	
Lilliefors Test Statistic	0.104	**Lilliefors GOF Test**	
10% Lilliefors Critical Value	0.12	Detected Data appear Lognormal at 10% Significance Level	
Detected Data appear Approximate Lognormal at 10% Significance Level			
Lognormal ROS Statistics Using Imputed Non-Detects			
Mean in Original Scale	8638	Mean in Log Scale	5.983
SD in Original Scale	18414	SD in Log Scale	3.391
95% t UCL (assumes normality of ROS data)	12793	95% Percentile Bootstrap UCL	12911
95% BCA Bootstrap UCL	13680	95% Bootstrap t UCL	14942
95% H-UCL (Log ROS)	1855231		
Statistics using KM estimates on Logged Data and Assuming Lognormal Distribution			
KM Mean (logged)	6.03	KM Geo Mean	415.6
KM SD (logged)	3.286	95% Critical H Value (KM-Log)	5.7
KM Standard Error of Mean (logged)	0.449	95% H-UCL (KM -Log)	1173988
KM SD (logged)	3.286	95% Critical H Value (KM-Log)	5.7
KM Standard Error of Mean (logged)	0.449		
DL/2 Statistics			
DL/2 Normal		**DL/2 Log-Transformed**	
Mean in Original Scale	8639	Mean in Log Scale	6.015
SD in Original Scale	18413	SD in Log Scale	3.374
95% t UCL (Assumes normality)	12795	95% H-Stat UCL	1765241
DL/2 is not a recommended method, provided for comparisons and historical reasons			
Nonparametric Distribution Free UCL Statistics			
Detected Data appear Gamma Distributed at 5% Significance Level			
Suggested UCL to Use			
95% KM Approximate Gamma UCL	14846		
The calculated UCLs are based on assumptions that the data were collected in a random and unbiased manner.			
Please verify the data were collected from random locations.			
If the data were collected using judgmental or other non-random methods, contact a statistician to correctly calculate UCLs.			
Note: Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL.			
Recommendations are based upon data size, data distribution, and skewness using results from simulation studies.			
However, simulations results will not cover all Real World data sets; for additional insight the user may want to consult a statistician.			

Detected data follow a gamma as well as a lognormal distribution. It is noted here again that in situations such as this, where data fit a gamma and lognormal distribution, but not the normal distribution, it is generally preferable to use a gamma distribution due to instability that can arise due to excessively long right tails for some lognormal distributions, as demonstrated in Table 6-2. The various upper limits using gamma ROS and lognormal ROS methods and gamma and lognormal distribution on KM estimates are summarized in the following table.

Table 6-2. Upper Confidence Limits Computed using Gamma and Lognormal Distributions of Detected Data Sample Size = 55, No. of NDs=10, % NDs = 18.18%

Upper Limits	Gamma Distribution		Lognormal Distribution	
	Result	Reference/ Method of Calculation	Result	Reference/ Method of Calculation
Min (detects)	5.2	--	1.65	Logged
Max (detects)	79000	--	11.277	Logged
Mean (KM)	8638	--	6.3	Logged
Mean (ROS)	8637	--	8638	--
UCL95 (ROS)	15896	ProUCL 5.0 -GROS	14863	bootstrap-t on LROS, ProUCL 5.0
			12918	percentile bootstrap on LROS, Helsel(2012)
UCL (KM)	14844	ProUCL 5.0 - KM-Gamma	1173988	H-UCL, KM mean and *sd* on logged data, EPA (2009e)

All computations have been performed using the ProUCL software. In the above table, methods proposed/described in the literature have been cited in the Reference Method of Calculation column. The results summarized in the above table reiterate that the use of a gamma distribution cannot be dismissed just because it is easier to use a lognormal distribution to model skewed data sets. These results also demonstrate that for skewed data sets, one should use bootstrap methods which adjust for data skewness (e.g., bootstrap-t method) rather than using percentile bootstrap method.

7 Windows

The Windows tab in ProUCL 5.2 is a simple tab consisting of 3 options to help arrange user files according to their preference. Often this option will not even be used but on occasion it can be helpful.

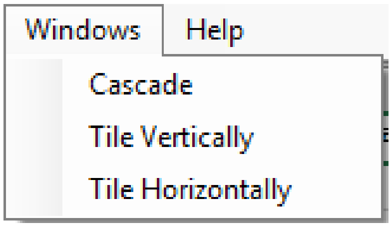

Figure 7-1. Windows options

Cascade creates a cascading flow of open user tabs that can be clicked through at will.

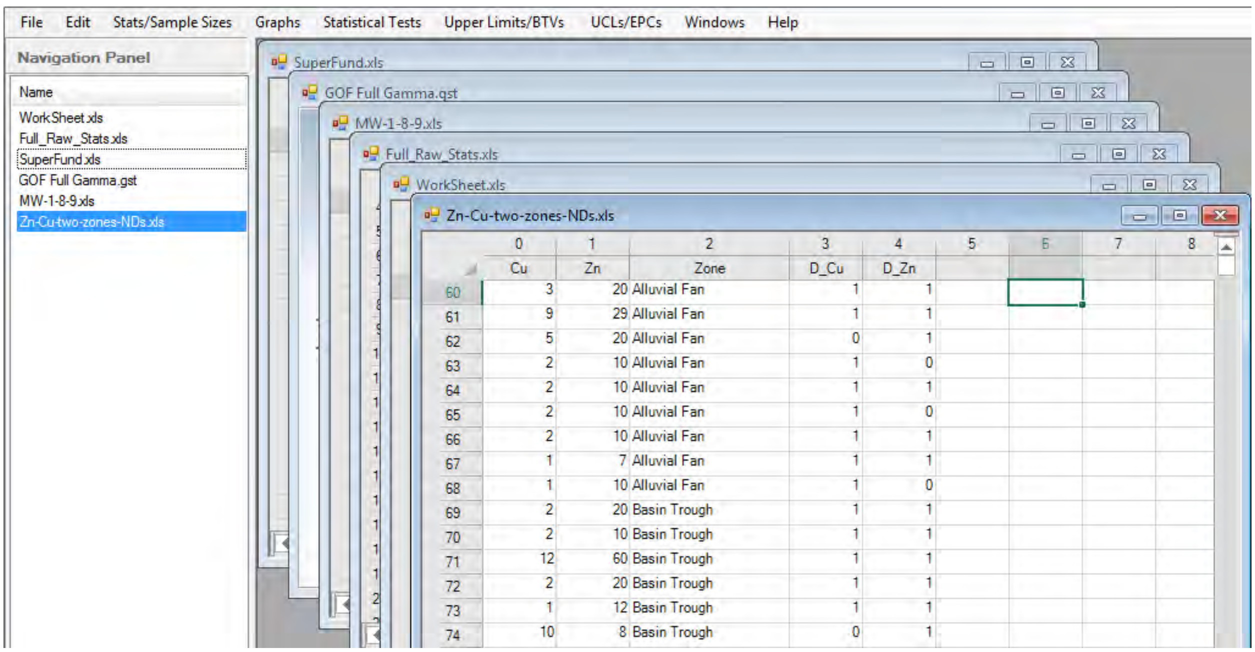

Figure 7-2. Cascade option

The Tile vertically option creates a flat layout of user tabs tiled vertically

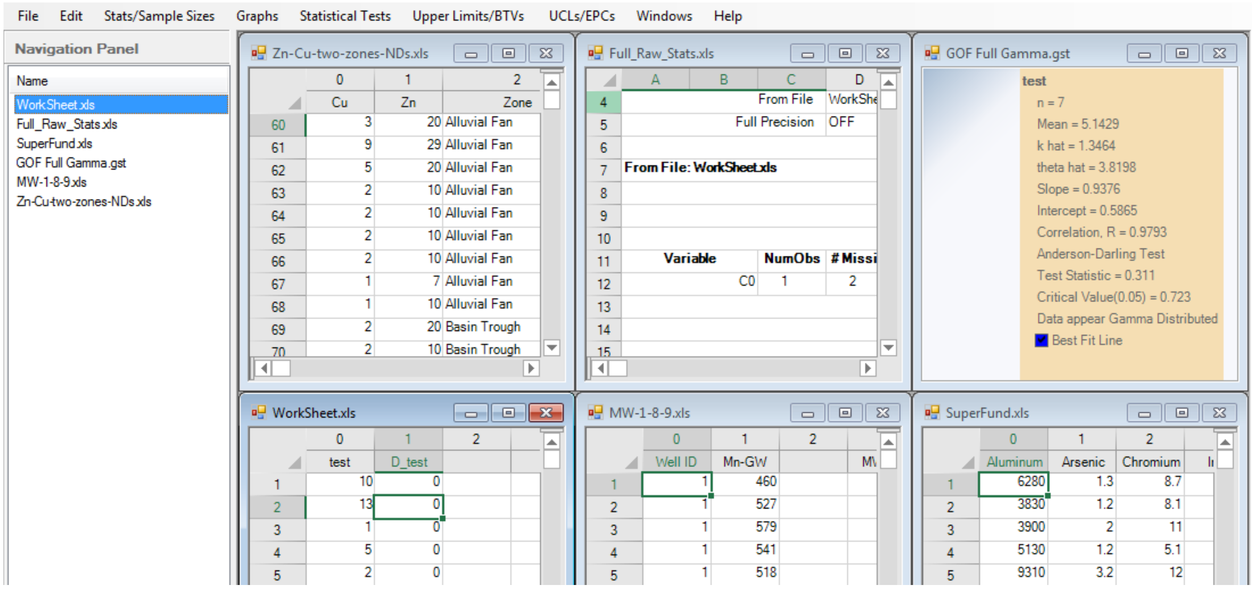

Figure 7-3 Tile Vertically option

Tile Horizontally functions similarly to Tile Horizontally but working in a horizontal rather than vertical tiling direction

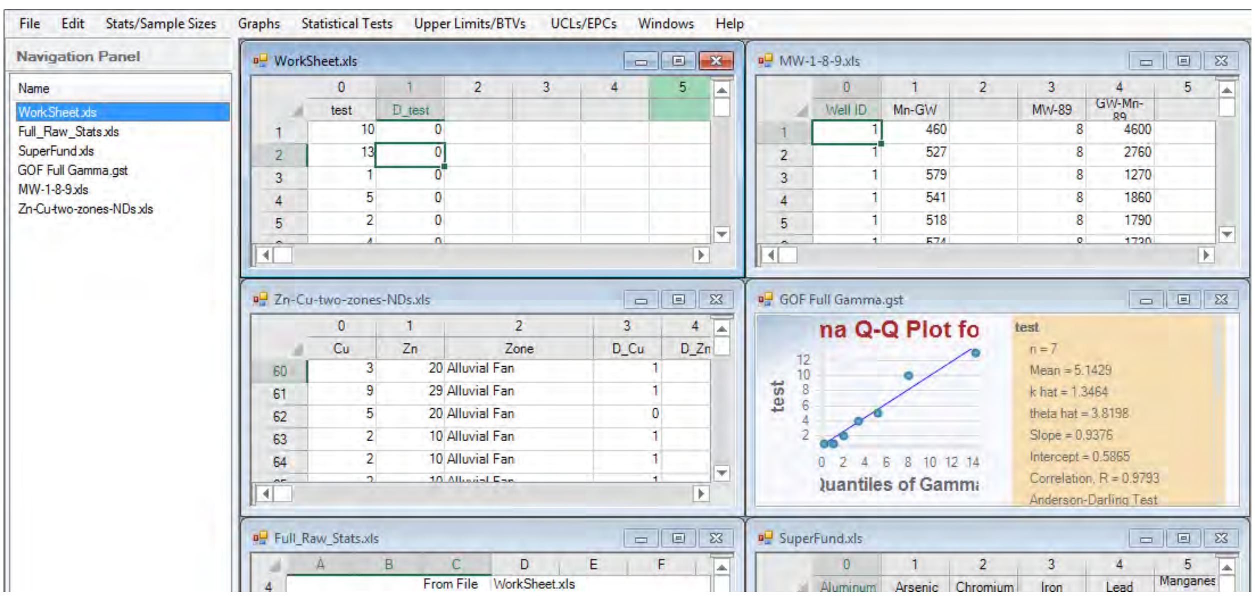

Figure 7-4 Tile Horizontally option

To return to a regular full window view of any user tabs simply click on the full screen icon just to the left of the red x on a given user window.

8 Help

The Help tab provides users with a couple small useful bits of information broken into About ProUCL, Overview, and Technical Support.

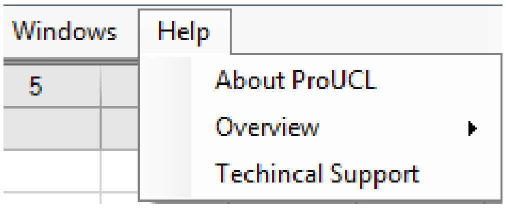

Figure 8-1 Help options

The About ProUCL option will provide the user with a bit of basic ProUCL build information, while the Overview option will give two options discussing in varying depths the updates and changes for the newest version of the build, while the Technical Support option provides users with contact information for support ProUCL support staff if they are in need of assistance.

9 Guidance on the Use of Statistical Methods in ProUCL Software

Decisions based upon statistics computed using discrete data sets of small sizes cannot be considered reliable enough to make decisions that affect human health and the environment. Several U.S. EPA guidance documents (e.g., EPA 2000, 2006a, 2006b) detail DQOs and minimum sample size requirements needed to address statistical issues associated with different environmental applications. In order to obtain reliable statistical results, an adequate amount of data should be collected using project-specified DQOs (i.e., CC, decision error rates). The **Sample Sizes** module (Section 2.3) of ProUCL computes minimum sample sizes based on DQOs specified by the user and described in many guidance documents. In some cases, it may not be possible (e.g., due to resource constraints) to collect the calculated number of samples needed to meet the project-specific DQOs. Under these circumstances one can use the **Sample Sizes** module to assess the power of the test statistic resulting from the reduced number of samples which were collected.

This chapter also describes the differences between the various statistical upper limits including upper confidence limits (UCLs) of the mean, upper prediction limits (UPLs) for future observations, and upper tolerance intervals (UTLs) often used to estimate the environmental parameters of interest including EPC terms and BTVs. The use of a statistical method depends upon the environmental parameter(s) being estimated or compared.

- The measures of central tendency (e.g., means, medians, or their UCLs) are used to compare site mean concentrations with a cleanup standard, C_s, also representing some central tendency measure of a reference area or some other known threshold representing a measure of central tendency.

- The upper threshold values, such as the CLs, alternative concentration limits (ACL), or not-to-exceed values, are used when individual point-by-point observations are compared with those threshold values.

Depending upon whether the environmental parameters (e.g., BTVs, not-to-exceed value, or EPC term) are known or unknown, different statistical methods with different data requirements are needed to compare site concentrations with pre-established or estimated standards and BTVs. Several upper limits, as well as single and two sample hypotheses testing approaches are available in ProUCL for both full-uncensored and left-censored data sets for performing the comparisons described above.

9.1 Summary of the DQO Process

While the purpose of this document is not to detail the DQO process, it is important for users of ProUCL to understand the basics of the process, as it is a recommended planning tool for collection of data of desired quality and a proper sample size for decisions to be made supported by statistical analysis of collected data. The discussion provided here is summarized and a more detailed discussion of the DQO process is located here. https://www.epa.gov/sites/production/files/2015-06/documents/g4-final.pdf

There are seven steps to the DQO process, which each play an important part in providing quality and quantity of data that are input for environmental data analysis. One element of the validity of ProUCL estimates is that seven steps of DQO process were appropriately applied before the data were collected. Outcome of ProUCL calculations need to therefore be critically evaluated against the DQOs set in planning process.

9.1.1 State the Problem

The first step in any systematic planning process, and therefore the DQO Process, is to define the problem that has initiated the study. As environmental problems are often complex combinations of technical, economic, social, and political issues, it is critical to the success of the process to separate each problem, define it completely, and express it in an uncomplicated format. A proven effective approach to formulating a problem and establishing a plan for obtaining information that is necessary to resolve the problem is to involve a team of experts and stakeholders that represent a diverse, multidisciplinary background. Such a team would provide: the ability to develop a concise description of complex problems, and multifaceted experience and awareness of potential data uses.

9.1.2 Identify Goals of the Study

Step 2 of the DQO Process involves identifying the key questions that the study attempts to address, along with alternative actions or outcomes that may result based on the answers to these key questions. For decision-making problems, you should combine the information from these two items to develop a decision statement, which is critical for defining decision performance criteria later in Step 6. For estimation problems, you should frame the study with an estimation statement from which a set of assumptions, inputs, and methods are referenced. On complex decision problems, you may identify multiple decisions that need to be made. These decisions are organized in a sequential or logical fashion within Step 2 and are examined to ensure consistency with the problem statement from Step 1. Similarly, large-scale or complex research studies may involve multiple estimators, and you will begin to determine how the different estimators relate to each other and to the overall study goal.

9.1.3 Identify Information Inputs

The third step of the DQO Process determines the types and sources of information needed to resolve the decision statement or produce the desired estimates; whether new data collection is necessary; the information basis the planning team will need for establishing appropriate analysis approaches and performance or acceptance criteria; and whether appropriate sampling and analysis methodology exists to properly measure environmental characteristics for addressing the problem. Once you have determined what needs to be measured, you may refine the criteria for these measurements in later steps of the DQO Process.

9.1.4 Define Boundaries of the Study

In Step 4 of the DQO Process, you should identify the target population of interest and specify the spatial and temporal features pertinent for decision making or estimation. The target population refers to the total collection or universe of sampling units to be studied and from which samples will be drawn. If the target population consists of "natural" entities (e.g., people, plants, or fish), then the definition of sampling unit is straightforward, it is the entity itself. When the target population consists of continuous media, such as air, water, or soil, the sampling unit must be defined as some area, volume, or mass that may be selected from the target population. When defining sampling units, you should ensure that the sampling units are mutually exclusive (i.e., they do not overlap), and are collectively exhaustive (i.e., the sum of all sampling units covers the entire target population). The actual determination of the appropriate size of a sampling unit, and of an optimal quantity of sample support for environmental data collection efforts can be complicated, and usually will be addressed as a part of the sampling design in Step 7. Here in Step 4, the planning team should be able to provide a first approximation of the sampling unit definition when specifying the target population. Practical constraints that could interfere with sampling should also be identified in this step. A practical constraint is any hindrance or obstacle (such as fences, property access, water bodies) that may interfere with collecting a complete data set. These constraints may limit the spatial and/or temporal boundaries or regions that will be included in the study population and hence, the inferences (conclusions) that can be made with the study data. You also should determine the scale of inference for decisions or estimates. The scale of inference is the area or volume, from which the data will be aggregated to support a specific decision or estimate. For example, a decision about the average concentration of lead in surface soil will depend on area over which the data are aggregated, so you should identify the size of decision units for this problem. A decision or estimate on each piece of land may lead to the recommendation of a specific size such as a half-acre area (equivalent to a semi-urban home area) for the sampling unit.

9.1.5 Develop Analytical Approach

Step 5 of the DQO Process involves developing an analytic approach that will guide how you analyze the study results and draw conclusions from the data. To clarify what you would truly like to learn from the study results, you should imagine in Step 5 that perfect information will be available for making decisions or estimates, thereby allowing you to focus on the underlying "true" conditions of the environment or system under investigation. (This assumption will be relaxed in Step 6, allowing you to manage the practical concerns associated with inherent uncertainty in the data.) The planning team should integrate the outputs from the previous four steps with the parameters (i.e., mean, median, or percentile) developed in this step. For decision problems, the theoretical decision rule is an unambiguous "If...then...else..." statement. For estimation problems, this will result in a clear specification of the estimator (statistical function) to be used to produce the estimate from the data

9.1.6 Specify Performance or Acceptance Criteria

In Step 6 of the DQO Process, you no longer imagine that you have access to perfect information on unlimited data as you did in Step 5. You now face the reality that you will not have perfect information from which to formulate your conclusions. Furthermore, these data are subject to various types of errors due to such factors as how samples were collected, how measurements were made, etc. As a result, estimates or conclusions that you make from the collected data may deviate from what is actually true within the population. Therefore, there is a chance that you will make erroneous conclusions based on your collected data or that the uncertainty in your estimates will exceed what is acceptable to you. In Step 6, you should derive the performance or acceptance criteria that the collected data will need to achieve in order to minimize the possibility of either making erroneous conclusions or failing to keep uncertainty in estimates to within acceptable levels. Performance criteria, together with the appropriate level of Quality Assurance practices, will guide your design of new data collection efforts, while acceptance criteria will guide your design of procedures to acquire and evaluate existing data relative to your intended use. Therefore, the method you use and the type of criteria that you set will, in part, be determined based on the intended use of your data.

9.1.7 Develop Plan for Obtaining Data

By performing Steps 1 through 6 of the DQO Process, you will have generated a set of performance or acceptance criteria that your collected data will need to achieve. The goal of Step 7 is to develop a resource-effective design for collecting and measuring environmental samples, or for generating other types of information needed to address your problem. This corresponds to generating either (a) the most resource-effective data collection process that is sufficient to fulfill study objectives, or (b) a data collection process that maximizes the amount of information available for synthesis and analysis within a fixed budget. In addition, this design will lead to data that will achieve your performance or acceptance criteria. Development of the sampling design is followed by development of the study's QA Project Plan. EPA has developed Guidance for Choosing a Sampling Design for Environmental Data Collection (EPA QA/G-5S) (U.S. EPA, 2002c) which addresses how to create sampling designs for environmental data collection and contains detailed information for six different sampling designs and protocols that are relevant to environmental data collection. In addition, EPA's Data Quality Assessment: Statistical Tools for Practitioners (EPA QA/G-9S) provides examples of common statistical hypothesis tests, approaches to calculating confidence intervals, and sample size formulae that may be relevant for your problem.

9.2 Background Data Sets

Based upon the Conceptual Site Model (CSM) and regional and expert knowledge about the site, the project team selects background or reference areas. Depending upon the site activities and the pollutants, the background area can be site-specific or a general reference area with conditions comparable to the site before contamination due to site related activities. An appropriate random sample of observations should be collected from the background area. A defensible background data set represents a "single" environmental population.

The background data set needs to be evaluated for the presence of data caused by reporting and/or laboratory errors, and extreme values that are suspects of misrepresenting the observed population. Statistical outlier tests give probabilistic evidence for the “misfit” of extreme values. However, their drawback is that they assume a normal distribution of the data without outliers. This is often not the case with environmental data, which tend to be right-skewed, either naturally or due to subsampling error. Therefore, statistical outlier tests available in ProUCL should only be used to identify potential suspect data points that require further investigation to gain an understanding of extreme values in the context of site processes, geology, and historical use. For example, extreme values may represent contamination from the site (hot spots) or high data variability caused by subsampling error. However, it is not unusual for a background to consist of different subpopulations due to the presence of varying soil types, textures, vegetation, historical use of the site, etc. It may have, therefore, have higher variability than expected in the planning process. The same issue of different subpopulations caused by soil types, etc. is also present in site areas.

To obtain representative estimates for the decision-making statistics (e.g., UCLs, UPLs and UTLs), data need to be critically evaluated. Following a five-step process as described in EPA QA/G-9S (2006) Data Quality Assessment: Statistical Methods for Practitioners is recommended:

1. Identify extreme values that may be potential outliers;
2. Apply statistical test;
3. Scientifically review statistical outliers and decide on their disposition;
4. Conduct data analyses with and without statistical outliers; and
5. Document the entire process.

When calculating background threshold value (BTV), the objective is to compute background statistics based upon a data set which is representative of the background population. The occurrence of elevated outliers is not uncommon when background samples are collected from various onsite areas (e.g., large Federal Facilities). The proper disposition of outliers, to include or not include them in statistical computations, should be decided by the project team. The project team may want to compute decision statistics with and without the outliers to evaluate the influence of outliers on the decision making statistics.

A couple of classical outlier tests (Dixon and Rosner tests) are available in ProUCL. These tests assume normal distribution of the data without outliers. Therefore, a distribution of the data needs to be verified before outlier tests are applied. If the data are not normally distributed, they should be normalized by using an appropriate transformation before outlier tests are applied. It is also recommended that these classical outlier tests be supplemented with graphical displays such as a box plot and Q-Q plot. The use of exploratory graphical displays helps in determining the number of outliers potentially present in a data set.

An appropriate background data set of a reasonable size (preferably computed using the DQO process) is needed for the data set to be representative of background conditions and to compute upper limits (e.g., estimates of BTVs) and compare site and background data sets using hypotheses testing approaches. A background data set should have a minimum of 10 observations.

9.3 Site Data Sets

A data set collected from a site population (e.g., AOC, exposure area [EA], DU, group of MWs) should be representative of the population under investigation. Depending upon the 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 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 the planning stages when developing sampling designs. Specifically, the availability of an adequate amount of representative data is required from each of those site sub-populations/strata defined by sample depths, soil types, and other characteristics.

Site data collection requirements depend upon the objective(s) of the study. Specifically, in background versus site comparisons, site data are needed to perform:

- point-by-point onsite comparisons with pre-established ALs or estimated BTVs. Typically, this approach is used when only a small number (e.g., < 6) of onsite observations are compared with a BTV or some other not-to-exceed value. More details can be found in Chapter 3 of the Technical Guide. Alternatively, one can use hypothesis testing approaches (Chapter 6 of ProUCL Technical Guide) provided enough observations (provided by the DQO process preferably, or at least 10) are available.

- single-sample hypotheses tests to compare site data with a pre-established cleanup standard, C_s (e.g., representing a measure of central tendency); proportion test to compare site proportion of exceedances of an AL with a pre-specified allowable proportion, P_0. These hypotheses testing approaches are used on site data when enough site observations are available. Specifically, when at a bare minimum 10 site observations for parametric methods, or 15 for non-parametric methods, 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 both types of error rates (Type 1 and Type 2) 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 a BTV (e.g., 95th percentile) by onsite observations, even when the site and background populations have comparable distributions. The probabilities of these chance exceedances increase as the site sample size increases.

Table 9-1. Probabilities of Exceeding a 95-95 BTV for Various Sample Sizes, When Site and Background Populations Have the Same Distribution

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 tests to compare site data distribution with background data distribution to determine if the site concentrations are comparable to background concentrations. An adequate amount of data needs to be made available from the site as well as the background populations. It is preferable to collect these data via the DQO process as noted in Section 9.1, however at least 10 observations for parametric methods, and 15 from non-parametric methods, need to be collected from each population under comparison.

Notes: From a mathematical point of view, one can perform hypothesis tests on data sets consisting of only 3-4 data values; however, the reliability of the test statistics (and the conclusions derived) thus obtained is questionable. In these situations, it is suggested to supplement the test statistics decisions with graphical displays.

9.4 Discrete Samples or Composite Samples?

ProUCL can be used for discrete sample data sets, as well as on composite sample data sets. However, in a data set (background or site), samples should be either all discrete or all composite, and the background data set should use the same method as the site data set.. 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.

9.5 Upper Limits and Their Use

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. The differences between UCLs and UPLs (or upper percentiles), and UCLs and UTLs should be clearly understood. A UCL with a 95% confidence limit (UCL95) of the mean represents an estimate of the population mean (measure of the central tendency), whereas a UPL95, a UTL95%-95% (UTL95-95), and an upper 95^{th} percentile represent estimates of a threshold from the upper tail of the population distribution such as the 95^{th} percentile. Here, UPL95 represents a 95% upper prediction limit, and UTL95-95 represents a 95% confidence limit of the 95^{th} percentile. For mildly skewed to moderately skewed data sets, the numerical values of these limits tend to follow the order given as follows.

Sample Mean ≤ UCL95 of Mean ≤ Upper 95^{th} Percentile ≤ UPL95 of a Single Observation ≤ UTL95-95

Example 7-1. Consider a real data set collected from a Superfund site (Included in the ProUCL download as superfund.xls). The data set has several inorganic COPCs, including aluminum (Al), arsenic (As), chromium (Cr), iron (Fe), lead (Pb), manganese (Mn), thallium (Tl) and vanadium (V). Iron concentrations follow a normal distribution. This data set has been used in several examples throughout the two ProUCL guidance documents (Technical Guide and User Guide), therefore it is provided as follows.

Table 9-2. Data Set for Example 7-1.

Aluminum	Arsenic	Chromium	Iron	Lead	Manganese	Thallium	Vanadium
6280	1.3	8.7	4600	16	39	0.0835	12
3830	1.2	8.1	4330	6.4	30	0.068	8.4
3900	2	11	13000	4.9	10	0.155	11
5130	1.2	5.1	4300	8.3	92	0.0665	9
9310	3.2	12	11300	18	530	0.071	22
15300	5.9	20	18700	14	140	0.427	32
9730	2.3	12	10000	12	440	0.352	19
7840	1.9	11	8900	8.7	130	0.228	17
10400	2.9	13	12400	11	120	0.068	21
16200	3.7	20	18200	12	70	0.456	32
6350	1.8	9.8	7340	14	60	0.067	15
10700	2.3	14	10900	14	110	0.0695	21
15400	2.4	17	14400	19	340	0.07	28
12500	2.2	15	11800	21	85	0.214	25
2850	1.1	8.4	4090	16	41	0.0665	8
9040	3.7	14	15300	25	66	0.4355	24
2700	1.1	4.5	6030	20	21	0.0675	11
1710	1	3	3060	11	8.6	0.066	7.2
3430	1.5	4	4470	6.3	19	0.067	8.1
6790	2.6	11	9230	13	140	0.068	16
11600	2.4	16.4		98.5	72.5	0.13	
4110	1.1	7.6		53.3	27.2	0.068	
7230	2.1	35.5		109	118	0.095	
4610	0.66	6.1		8.3	22.5	0.07	

Several upper limits for iron are summarized as follows, and it be seen that they follow the order (in magnitude) as described above.

Table 9-3. 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

For highly skewed data sets, these limits may not follow the 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% and the largest sample value. 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 and the sample size is large (e.g., > 50, 100)There is a great deal of confusion about the appropriate use of these upper limits. A brief discussion about the differences between the applications and uses of the statistical limits described above is provided as follows.

- A UCL represents an average value that is compared with a threshold value also representing an average value, such as a mean $C_{s.}$ For example, a site 95% UCL exceeding a $C_{s,}$ may lead to the conclusion that the cleanup standard, C_s has not been attained by the average site area concentration. It should also be noted that UCLs of means are typically computed from the site data set.

- 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 a site average or a site median with a specified or pre-established cleanup standard, or with the background population average or median.

- A UPL, an upper percentile, or a 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 point-by-point onsite observations are compared with those limits. A site observation exceeding a background UTL may lead to the conclusion that the constituent is present at the site at levels greater than the background concentrations level.

- Single-sample hypotheses testing approaches should be used to compare a site mean or median against a known threshold comparison; and two-sample hypotheses testing approaches should be used to compare a site population with a background population. Several parametric (typically testing the mean) and nonparametric (typically testing the median) single and two-sample hypotheses testing approaches are available in ProUCL.

It is re-emphasized that only averages should be compared with averages, and individual site observations should be compared with UPLs, upper percentiles, UTLs, or USLs. For 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.

9.6 Point-by-Point Comparison of Site Observations with BTVs, and Other Threshold Values

The point-by-point observation comparison method is used when a small number (e.g., < 6) of site observations are compared with pre-established or estimated BTVs, screening levels, or preliminary remediation goals (PRGs). Typically, a single exceedance of the BTV by a site observation may be considered an indication of the presence of contamination at the site area under investigation. The conclusion of an exceedance by a site value is sometimes confirmed by re-sampling at the site location exhibiting constituent concentrations in excess of the BTV. If all collocated sample observations (or all sample observations collected during the same time period) from the same site location exceed the BTV or PRG, then it may be concluded that the location requires further investigation (e.g., continuing treatment and monitoring) and possibly cleanup.

When BTV constituent concentrations are not known or pre-established, one has to collect or extract a background data set of an appropriate size that can be considered representative of the site background. Statistical upper limits are computed using the background data set thus obtained, which are used as estimates of BTVs. To compute reasonably reliable estimates of BTVs, sample size should be established via the DQO process as stated in Section 9.1 but a minimum of 10 background observations should be collected if that is infeasible.

The point-by-point comparison method is also useful when quick turnaround comparisons are required in real time. Specifically, when decisions have to be made in real time by a sampling/screening crew, or when only a few site samples are available, then individual point-by-point site concentrations are compared either with pre-established cleanup goals or with estimated BTVs. The sampling crew can use these comparisons to:

1. screen and identify the COPCs
2. identify the potentially polluted site AOCs
3. continue or stop remediation or excavation at an onsite area of concern.

If a larger number of samples (e.g., >10) are available from the AOC, then the use of hypotheses testing approaches (both single-sample and a two-sample) is preferred. The use of hypothesis testing approaches tends to control the error rates more tightly and efficiently than the individual point-by-point site comparisons.

9.7 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 constituent concentrations of two (e.g., site versus background) or more populations (e.g., MWs).

9.7.1 Single Sample Hypothesis Testing

When pre-established BTVs are used such as the U.S. Geological Survey (USGS) background values (Shacklette and Boerngen 1984), or thresholds obtained from similar sites, there is no need to extract, establish, or collect a background data set. When the BTVs and cleanup standards are known, one-sample hypotheses are used to compare site data with known and pre- established threshold values. As mentioned earlier, when the number of available site samples is < 6, one might perform point-by-point site observation comparisons with a BTV; and when enough site observations (at least 10 for parametric, and 15 for non-parametric methods) are available, it is desirable to use single-sample hypothesis testing approaches. Depending upon the parameter (μ_0, A_0), represented by the known threshold value, one can use single-sample hypotheses tests for population mean or median (t-test, sign test), or use single-sample tests for proportions and percentiles. The details of the single-sample hypotheses testing approaches can be found in EPA (2006b) guidance document and in Chapter 6 of ProUCL Technical Guide.

One-Sample t-Test: This test is used to compare the site mean, μ, with some specified cleanup standard, C_s, where the C_s represents an average threshold value, μ_0. The Student's t-test (or a UCL of the mean) may be used to verify 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 ND observations, provided the detection limits of all NDs fall below the specified threshold value, C_s. These tests are used to compare the site location (e.g., median, mean) with some specified 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 constituent concentration distribution rather than the mean threshold value, μ_0, then a test for proportion or a test for percentile (equivalently UTL 95-95 UTL 95-90) may be used to compare site proportion (or site percentile) with the specified threshold or action level, A_0.

9.7.2 Two-Sample Hypothesis Testing

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, two-sample hypothesis testing approaches are 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 from each of the two populations, as mentioned in Section 9.1 this is best established from the DQO process, or when that is infeasible a minimum of 10 samples for parametric methods, and 15 for non-parametric methods should be taken in each of both the site and background datasets. 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) in sample size determination. That is, a larger number (>15-20) of representative background (and site) samples should be collected from larger background (and site) areas; every effort should be made to collect as many samples as determined by the DQOs-based sample sizes.

The two-sample hypotheses testing approaches incorporated in ProUCL 5.2 are listed as follows:

1. Student t-test (with equal and unequal variances)—Parametric test assumes normality
2. Wilcoxon-Mann-Whitney (WMW) test—Nonparametric test handles data with NDs with one DL—assumes two populations have comparable shapes and variability
3. Gehan test—Nonparametric test handles data sets with NDs and multiple DLs - assumes comparable shapes and variability
4. Tarone-Ware (T-W) test—Nonparametric test handles data sets with NDs and multiple DLs - assumes comparable shapes and variability

The Gehan and T-W tests are meant to be used on left-censored data sets with multiple DLs. For best results, the samples collected from the two (or more) populations should all be of the same type obtained using similar analytical methods and apparatus; the collected site and background samples should all be discrete or all composite (obtained using the same design and pattern), and be collected from the same medium 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 analytical methods. Good sample collection methods and sampling strategies are given in EPA (1996, 2003) guidance documents.

9.8 Sample Size Requirements and Power Evaluations

Due to resource limitations, it may not be possible to sample the entire population (e.g., background area, site area, AOCs, EAs) under study. Statistics is used to draw inferences about the populations and their known or unknown statistical parameters based upon much smaller data samples, collected from those populations. To determine and establish BTVs and site-specific screening levels, defensible data sets of appropriate sizes representing the background population (e.g., site-specific, general reference area, or historical data) need to be collected. 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 area and boundaries based upon all current and intended future uses, and the objectives of data collection. 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 constituent concentrations are comparable; or if the site concentrations exceed the background threshold concentration level; or if an adequate amount of remediation approaching the BTV or some cleanup level has been performed at polluted site AOCs.

To perform these statistical tests, determine the number of samples that need to be collected from the populations (e.g., site and background) under investigation using appropriate DQOs processes (EPA 2000, 2006a, 2006b). ProUCL has the **Sample Sizes** module which can be used to develop DQOs based sampling designs needed to address statistical issues associated with polluted sites projects. ProUCL provides user-friendly options to enter the desired/pre-specified values of decision parameters (e.g., Type I and Type II error rates) to determine minimum sample sizes for the selected statistical applications including: estimation of mean, single and two-sample hypothesis testing approaches, and acceptance sampling. Sample size determination methods are available for the sampling of continuous characteristics (e.g., lead or Radium 226), as well as for attributes (e.g., proportion of occurrences exceeding a specified threshold). Both parametric (e.g., t-tests) and nonparametric (e.g., Sign test, test for proportions, WRS test) sample size determination methods are available in ProUCL. ProUCL also has sample size determination methods for acceptance sampling of lots of discrete objects such as a batch of drums containing hazardous waste (e.g., RCRA applications, U.S. EPA 2002c).

However, due to budgetary or logistical constraints, it may not be possible to collect the same number of samples as determined by applying a DQO process. For example, the data might have already been collected (as often is the case) without using a DQO process, or due to resource constraints, it may not have been possible to collect as many samples as determined by using a DQO-based sample size formula.

In practice, the project team and the decision makers tend not to collect enough background samples. It is suggested to collect at least 10 background observations before using statistical methods to perform background evaluations based upon data collected using discrete samples. In case data are collected without using a DQO process, the **Sample Sizes** module can be used to assess the power of the test statistic in retrospect. Specifically, one can use the standard deviation of the computed test statistic (EPA 2006b) and compute the sample size needed to meet the desired DQOs. If the computed sample size is greater than the size of the data set used, the project team may want to collect additional samples to meet the desired DQOs.

Note: From a mathematical point of view, the statistical methods incorporated in ProUCL and described in this guidance document for estimating EPC terms and BTVs, and comparing site versus background concentrations can be performed on small site and background data sets (e.g., of sizes as small as 3). However, those statistics may not be considered representative and reliable enough to make important cleanup and remediation decisions which will potentially impact human health and the environment. ProUCL provides messages when the number of detects is <4-5, and suggests collecting at least 10 observations. Based upon professional judgment, as a rule-of-thumb, ProUCL guidance documents recommend collecting a minimum of 10 observations when data sets of a size determined by a DQOs process (EPA 2006) cannot be collected. This, however, should not be interpreted as the general recommendation and every effort should be made to collect DQOs based number of samples. Some recent guidance documents (e.g., EPA 2009e) have also adopted this rule-of-thumb and suggest collecting a minimum of about 10 samples in the circumstance that data cannot be collected using a DQO-based process. However, the project team needs to make these determinations based upon their comfort level and knowledge of site conditions.

- To allow users to compute decision statistics using data from ISM (ITRC, 2020) samples, ProUCL 5.2 will compute decision statistics (e.g., UCLs, UPLs, UTLs) based upon samples of sizes as small as 3. The user is referred to the ITRC ISM Technical Regulatory Guide (2020) to determine what sample size is appropriate, and which UCL (e.g., Student's t-UCL or Chebyshev UCL) should be used to estimate the EPC term. However, note that the Chebyshev UCL may grossly overestimate the mean.

Table 9-4. Sample size requirements at a glance.

Minimum number of Background and Site Samples when using Non-Parametric methods	Should be developed on a case-by-case basis using the DQO process. (Bare minimum 15 samples in each of the background and Site datasets)
Minimum number of Background and Site Samples when using Parametric methods	Should be developed on a case-by-case basis using the DQO process. (Bare minimum 10 samples in each of the background and Site datasets)
Site samples to be individually compared to a background threshold value	<6

9.8.1 Why a Data Set of Minimum Size, n = 10?

Typically, the computation of parametric upper limits (UPL, UTL, UCL) depends upon three values: the sample mean, sample variability (standard deviation) and a critical value. A critical value depends upon sample size, data distribution, and confidence level. For samples of small size (< 10), the data distribution of the population from which the data derive is uncertain, and the critical values are large and unstable, and upper limits (e.g., UTLs, UCLs) based upon a data set with fewer than 10 observations are mainly driven by those critical values. The differences in the corresponding critical values tend to stabilize when the sample size becomes larger than 10 (see tables below, where degrees of freedom [*df*] = sample size - 1). This is one of the reasons ProUCL guidance documents suggest a minimum data set size of 10 when the number of observations determined from sample-size calculations based upon EPA DQO process exceed the logistical/financial/temporal/constraints of a project. For samples of sizes 2-11, 95% critical values used to compute upper limits (UCLs, UPLs, UTLs, and USLs) based upon a normal distribution are summarized in the subsequent tables. In general, a similar pattern is followed for critical values used in the computation of upper limits based upon other distributions.

For the normal distribution, Student's t-critical values are used to compute UCLs and UPLs which are summarized as follows.

9.9 Critical Values of t-Statistic

Table 9-5. Critical Values of t-Statistic. *df*= sample size-1= *(n-1)*.

	Upper-tail probability *p*				
df	.10	.05	.025	.02	.01
1	3.078	6.314	12.71	15.89	31.82
2	1.886	2.920	4.303	4.849	6.965
3	1.638	2.353	3.182	3.482	4.541
4	1.533	2.132	2.776	2.999	3.747
5	1.476	2.015	2.571	2.757	3.365
6	1.440	1.943	2.447	2.612	3.143
7	1.415	1.895	2.365	2.517	2.998
8	1.397	1.860	2.306	2.449	2.896
9	1.383	1.833	2.262	2.398	2.821
10	1.372	1.812	2.228	2.359	2.764

One can see that once the sample size starts exceeding 9-10 (*df* = 8, 9), the difference between the critical values starts stabilizing. For example, for upper tail probability (= level of significance) of 0.05, the difference between critical values for *df* = 9 and *df* =10 is only 0.021, whereas the difference between critical values for *df*= 4 and 5 is 0.117; similar patterns are noted for other levels of significance. For the normal distribution, critical values used to compute UTL90-95, UTL95-95, USL90, and USL95 are described as follows. One can see that once the sample size starts exceeding 9-10, the difference between the critical values starts decreasing significantly.

Table 9-6. UTLs and USLs for Various Sample Sizes and Confidence Levels.

n	UTL90-95	UTL95-95	USL90	USL95
3	6.155	7.656	1.148	1.153
4	4.162	5.144	1.425	1.462
5	3.407	4.203	1.602	1.671
6	3.006	3.708	1.729	1.822
7	2.755	3.399	1.828	1.938
8	2.582	3.187	1.909	2.032
9	2.454	3.031	1.977	2.11
10	2.355	2.911	2.036	2.176
11	2.275	2.815	2.088	2.234

Note: Nonparametric upper limits (UPLs, UTLs, and USLs) are computed using higher order statistics (i.e., the maximum, second largest, third largest, and so on) of a data set. To achieve the desired confidence coefficient, samples of sizes much greater than 10 are required. It should be noted that critical values of USLs are significantly lower than critical values for UTLs. Critical values associated with UTLs decrease as the sample size increases. Since, as the sample size increases the maximum of the data set also increases, and critical values associated with USLs increase with the sample size.

9.9.1 Sample Sizes for Non-Parametric Bootstrap Methods

Several nonparametric methods including bootstrap methods for computing UCL, UTL, and other limits for both full-uncensored data sets and left-censored data sets with NDs are available in ProUCL. Bootstrap resampling methods are useful when not too few (e.g., < 15-20) and not too many (e.g., > 500- 1000) 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 are drawn with replacement from the same data set. Therefore, to obtain bootstrap resamples with at least 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 15-20. Also, it is not necessary 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., > 500) 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.

Note: Rules-of-thumb about minimum sample size requirements described in this section are based upon professional experience of the developers. ProUCL software is not a policy software. It is recommended that the users/project teams/agencies make determinations about the minimum number of observations and minimum number of detects that should be present in a data set before using a statistical method.

9.10 Statistical Analyses by a Group ID

In environmental applications data are commonly categorized by a group ID variable such as:

1. Surface vs. Subsurface
2. AOC1 vs. AOC2
3. Site vs. Background
4. Upgradient vs. Downgradient monitoring wells

The **Group Option** provides a tool for performing separate statistical tests and for generating separate graphical displays for each member/category of the group (samples from different populations) that may be present in a data set. The graphical displays (e.g., box plots, quantile-quantile plots) and statistics (e.g., background statistics, UCLs, hypotheses tests) of interest can be computed separately for each group by using this option. Moreover, using the **Group Option**, graphical methods can display multiple graphs (e.g., Q-Q plots) on the same graph providing graphical comparison of multiple groups.

It should be pointed out that it is the user's responsibility to provide an adequate amount of data to perform the group operations (see section 2.3). For example, if the user desires to produce a graphical Q-Q plot (e.g., using only detected data) with regression lines displayed, then there should be at least two detected data values (to compute slope, intercept, *sd*) in the data set. Similarly, if the graphs are desired for each group specified by the group ID variable, there should be at least two observations in each group specified by the group variable. When ProUCL data requirements are not met, ProUCL does not perform any computations, and generates a warning message (colored orange) in the lower Log Panel of the output screen of ProUCL.

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

BTVs and not-to-exceed values represent upper threshold values from 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 largest or the second largest detected value. A nonparametric UPL, UTL, and USL are often estimated by higher order statistics such as the maximum value or the second largest value (EPA 1992b, 2009, Hahn and Meeker 1991). The use of higher order statistics to estimate the UTLs depends upon the sample size. For data sets of size: 1) 59 to 92 observations, a nonparametric UTL95-95 is given by the maximum detected value; 2) 93 to 123 observations, a nonparametric UTL95-95 is given by the second largest maximum detected value; and 3) 124 to 152 observations, a UTL95-95 is given by the third largest detected value in the sample, and so on.

9.12 Use of Maximum Detected Value 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 value. Specifically, EPA (1992c) suggests the use of the maximum detected value as the EPC term when a 95% UCL (e.g., the H-UCL) exceeds the maximum value in a data set and "additional data cannot be practically obtained." ProUCL computes 95% UCLs of the mean using several methods based upon normal, gamma, lognormal, and non-identified distributions. In the past, a lognormal distribution was used as the default distribution to model positively skewed environmental data sets. Additionally, only two methods were used to estimate the EPC term based upon: 1) normal distribution and Student's t-statistic, and 2) lognormal distribution and Land's H-statistic (Land 1971, 1975). The use of the H-statistic can yield unstable and impractically large UCL95 for the mean (Singh, Singh, and Engelhardt 1997; Singh, Singh, and Iaci 2002), particularly when the data are not truly lognormal. For highly skewed data sets of smaller sizes (< 30, < 50), H-UCL often exceeds the maximum detected value. ProUCL 5.2 no longer recommends the H-UCL when the sample size is small (n < 75) and the true distribution cannot be reliably determined. Rather than defaulting to lognormality, ProUCL 5.2 tests normality first ($\alpha = 0.01$) due to the stability and robustness of the Student's t-UCL. Gamma UCLs are well-behaved and are recommended in cases where the data are non-normal ($\alpha = 0.05$) but appear to follow a gamma distribution ($\alpha = 0.05$). Lognormality is tested last due to the poor behavior of the H-UCL, and lognormality is rejected with comparatively less evidence against the null hypothesis of lognormality ($\alpha = 0.10$). For details on the changes to recommendations in ProUCL 5.2, refer to Chapter 2 of the Technical Guide.

It should be pointed out that in some cases, the maximum observed value actually might represent an impacted location. It is not desirable to use an observation potentially representing an impacted location to estimate the EPC for an AOC because the EPC represents the average exposure contracted by an individual over an EA during a long period of time. As such, 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. Singh and Singh (2003) studied the performance of the max test (using the maximum observed value to estimate the EPC) via Monte Carlo simulation experiments. They noted that for skewed data sets of small sizes (e.g., < 10-20), even 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 lead to unnecessary further remediation.

Several methods, some of which are described in EPA (2002a) and other EPA documents, are available in ProUCL for estimating the EPC terms. It is unlikely that the UCLs based upon those methods will exceed the maximum detected value, unless some outliers are present in the data set.

9.13 Alternative UCL95 Computations

ProUCL displays a warning message when the suggested 95% UCL (e.g., Hall's or bootstrap-t UCL with outliers) of the mean exceeds the detected maximum concentration. When a 95% UCL does exceed the maximum observed value, ProUCL suggests the use of an alternative UCL computation method. The choice of alternative UCL will depend on the particular data set and may require professional judgement. Practitioners are encouraged to contact a statistician for guidance.

<u>Notes:</u> Using the maximum observed value to estimate the EPC term representing the average exposure contracted by an individual over an EA is not recommended. 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. For such scenarios (when a 95% UCL does exceed the maximum observed value), an alternative UCL computation method should be used. Note that ProUCL no longer recommends the use of the Chebyshev UCL.

9.14 Samples with Nondetect Observations

ND observations are inevitable in most environmental data sets. Singh, Maichle, and Lee (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 ND observations. They concluded that the UCLs obtained using the substitution methods, including the replacement of NDs by DL/2; do not perform well even when the percentage of ND observations is low, such as less than 5% to 10%. They recommended avoiding the use of substitution methods for computing UCL95 based upon data sets with ND observations.

9.14.1 Avoid the Use of the DL/2 Substitution Method to Compute UCL95

Based upon the results of the report by Singh, Maichle, and Lee (2006), it is recommended to avoid the use of the DL/2 substitution method when performing a GOF test, and when computing the summary statistics and various other limits (e.g., UCL, UPL, UTLs) often used to estimate the EPC terms and BTVs. Until recently, the substitution method has been the most commonly used method for computing various statistics of interest for data sets which include NDs. The main reason for this has been the lack of the availability of the other rigorous methods and associated software programs that can be used to estimate the various environmental parameters of interest. Today, several methods (e.g., using KM estimates) with better performance, including the Chebyshev inequality and bootstrap methods, are available for computing the upper limits of interest. Several of those parametric and nonparametric methods are available in ProUCL 4.0 and higher versions. The DL/2 method is included in ProUCL for historical reasons as it had been the most commonly used and recommended method until recently (EPA 2006b). EPA scientists and several reviewers of the ProUCL software had suggested and requested the inclusion of the DL/2 substitution method in ProUCL for comparison and research purposes.

<u>Notes:</u> Even though the DL/2 substitution method has been incorporated in ProUCL, its use is **not recommended** due to its poor performance. The DL/2 substitution method has been retained in ProUCL for historical and comparison purposes. NERL-EPA, Las Vegas strongly recommends avoiding the use of this method even when the percentage of NDs is as low as 5% to 10%.

9.14.2 ProUCL Does Not Distinguish between Detection Limits, Reporting limits, or Method Detection Limits

ProUCL 5.1 (and all previous versions) does not make distinctions between method detection limits (MDLs), adjusted MDLs, sample quantitation limits (SQLs), reporting limits (RLs), or DLs. Multiple DLs (or RLs) in ProUCL mean different values of the detection limits. It is user's responsibility to understand the differences between these limits and use appropriate values (e.g., DLs) for nondetect values below which the laboratory cannot reliably detect/measure the presence of the analyte in collected samples (e.g., soil samples). A data set consisting of values less than the DLs (or MDLs, RLs) is considered a left-censored data set. ProUCL uses statistical methods available in the statistical literature for left-censored data sets for computing statistics of interest including mean, *sd*, UCL, and estimates of BTVs.

The user determines which qualifiers (e.g., J, U, UJ) will be considered as nondetects. Typically, all values with U or UJ qualifiers are considered as nondetect values. It is the user's responsibility to enter a value which can be used to represent a ND value. For NDs, the user enters the associated DLs or RLs (and not zeros or half of the detection limits). An indicator column/variable, D_x taking a value, 0, for all nondetects and a value, 1, for all detects is assigned to each variable, x, with NDs. It is the user's responsibility to supply the numerical values for NDs (should be entered as reported DLs) not qualifiers (e.g., J, U, B, UJ). For example, for thallium with nondetect values, the user creates an associated column labeled as D_thallium to tell the software that the data set will have nondetect values. This column, D_thallium consists of only zeros (0) and ones (1); zeros are used for all values reported as NDs and ones are used for all values reported as detects.

9.14.3 Samples with Low Frequency of Detection

When all of the sampled values are reported as NDs, the EPC term and other statistical limits should also be reported as a ND value, perhaps by the maximum RL or the maximum RL/2. The project team will need to make this determination. Statistics (e.g., UCL95) based upon only a few detected values (e.g., < 4) cannot be considered reliable enough to estimate EPCs which can have a potential impact on human health and the environment. When the number of detected values is small, it is preferable to use ad hoc methods rather than using statistical methods to compute EPCs and other upper limits. Specifically, for data sets consisting of < 4 detects and for small data sets (e.g., size < 10) with low detection frequency (e.g., < 10%), the project team and the decision makers should decide, on a site-specific basis, how to estimate the average exposure (EPC) for the constituent and area under consideration. For 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, when most (e.g., > 95%) of the observations for a constituent lie below the DLs, the sample median or the sample mode (rather than the sample average) may be used as an estimate of the EPC. Note that when the majority of the data are NDs, the median and the mode may also be represented by a ND value. The uncertainty associated with such estimates will be high. 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.

9.15 Some Other Applications of Methods in ProUCL

In addition to performing background versus site comparisons for CERCLA and RCRA sites, performing trend evaluations based upon time-series data sets, and estimating EPCs in exposure and risk evaluation studies, the statistical methods in ProUCL can be used to address other issues dealing with environmental investigations that are conducted at Superfund or RCRA sites.

9.15.1 Identification of COPCs

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

To identify the COPCs, point-by-point site observations are compared with some pre-established soil 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 onsite. The project team should decide the type of site samples (discrete or composite) and the number of site observations that should be collected and compared with the screening levels or the BTVs. In case BTVs or screening levels are not known, the availability of a defensible site-specific background or reference data set of reasonable size (e.g., at least 10) is required for computing reliable and representative estimates of BTVs and screening levels. The constituents with concentrations exceeding the respective screening values or BTVs may be considered COPCs, whereas constituents with concentrations (e.g., in all collected samples) lower than the screening values or BTVs may be omitted from all future evaluations.

9.15.2 Identification of Non-Compliance Monitoring Wells

In MW compliance assessment applications, individual (often discrete) constituent concentrations from a MW are compared with some pre-established limits such as an ACL or a maximum concentration limit (MCL). An exceedance of the MCL or the BTV (e.g., estimated by a UTL95-95 or a UPL95) by a MW concentration may be considered an indication of contamination in that MW. For individual concentration comparisons, the presence of contamination may have to be confirmed by re-sampling from that MW. If concentrations of constituents in the original sample and re-samples exceed the MCL or BTV, then that MW may require further scrutiny, perhaps triggering remediation activities. If the concentration data from a MW for a designated time period determined by the project team are below the MCL or BTV level, then that MW may be considered as complying with the pre-established or estimated standards.

9.15.3 Verification of the Attainment of Cleanup Standards, C_s

Hypothesis testing approaches are used to verify the attainment of the cleanup standard, C_s, at site AOCs after conducting remediation and cleanup at those site AOCs (EPA 1989a, 1994). In order to assess the attainment of cleanup levels, a representative data set of adequate size perhaps obtained using the DQO process needs to be made available from the remediated/excavated areas of the site under investigation. The sample size should also account for the size of the remediated site areas: meaning that larger site areas should be sampled more (with more observations) to obtain a representative sample of the remediated areas under investigation. Typically, the null hypothesis of interest is H_0: Site Mean, $\mu_s \geq C_s$ versus the alternative hypothesis, H_1: Site Mean, $\mu_s < C_s$, where the cleanup standard, C_s, is known *a priori*.

9.15.4 Using BTVs (Upper Limits) to Identify Hot Spots

The use of upper limits (e.g., UTLs) to identify hot spots 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 are compared with a pre-established or estimated BTV. Exceedances of the BTV by site observations may represent impacted locations with elevated concentrations.

9.16 Some General Issues, Suggestions and Recommendations made by ProUCL

9.16.1 Handling of Field Duplicates

ProUCL does not pre-process field duplicates. The project team determines how field duplicates will be handled and pre-processes the data accordingly. For an example, if the project team decides to use average values for field duplicates, then averages need to be computed and field duplicates need to be replaced by their respective average values. It is the user's responsibility to feed in appropriate values (e.g., averages, maximum) for field duplicates. The user is advised to refer to the appropriate EPA guidance documents related to collection and use of field duplicates for more information.

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

For data sets with NDs, ProUCL can compute point estimates of population mean and standard deviation using the KM and ROS methods (and also using the DL/2 substitution method, though it is not recommended). ProUCL uses Chebyshev inequality, bootstrap methods, and normal, gamma, and lognormal distribution-based equations on KM (or ROS) estimates to compute upper limits (e.g., UCLs, UTLs). The simulation study conducted by Singh, Maichle, and Lee (2006) demonstrated that the KM method yields accurate estimates of the population mean. They also demonstrated that for moderately skewed to highly skewed data sets, UCLs based upon KM estimates with BCA bootstrap (mild skewness), KM estimates with Chebyshev inequality (moderate to high skewness), and KM estimates with bootstrap-t method (moderate to high skewness) yield better estimates of EPCs, in terms of coverage probability, than other UCL methods based upon the Student's t- statistic on KM estimates, percentile bootstrap method on KM or ROS estimates.

10 REFERENCES

Aitchison, J. and Brown, J.A.C. 1969. The Lognormal Distribution, Cambridge: Cambridge University Press.

Anderson, T.W. and Darling, D. A. 1954. Test of goodness-of-fit. Journal of American Statistical Association, Vol. 49, 765-769.

Bain, L.J., and Engelhardt, M. 1991. Statistical Analysis of Reliability and Life Testing Models, Theory and Methods. 2nd Edition. Dekker, New York.

Bain, L.J. and Engelhardt, M. 1992. Introduction to probability and Mathematical Statistics. Second Edition. Duxbury Press, California.

Barber, S. and Jennison, C. 1999. Symmetric Tests and Confidence Intervals for Survival Probabilities and Quantiles of Censored Survival Data. University of Bath, Bath, BA2 7AY, UK.

Barnett, V. 1976. Convenient Probability Plotting Positions for the Normal Distribution. Appl. Statist., 25, No. 1, pp. 47-50, 1976.

Barnett, V. and Lewis T. 1994. Outliers in Statistical Data. Third edition. John Wiley & Sons Ltd. UK.

Bechtel Jacobs Company, LLC. 2000. Improved Methods for Calculating Concentrations used in Exposure Assessment. Prepared for DOE. Report # BJC/OR-416.

Best, D.J. and Roberts, D.E. 1975. The Percentage Points of the Chi-square Distribution. Applied Statistics, 24: 385-388.

Best, D.J. 1983. A note on gamma variate generators with shape parameters less than unity. Computing, 30(2):185-188, 1983.

Blackwood, L. G. 1991. Assurance Levels of Standard Sample Size Formulas, Environmental Science and Technology, Vol. 25, No. 8, pp. 1366-1367.

Blom, G. 1958. Statistical Estimates and Transformed Beta Variables. John Wiley and Sons, New York.
Bowman, K. O. and Shenton, L.R. 1988. Properties of Estimators for the Gamma Distribution, Volume 89. Marcel Dekker, Inc., New York.

Bradu, D. and Mundlak, Y. 1970. Estimation in Lognormal Linear Models. Journal of the American Statistical Association, 65, 198-211.

Chen, L. 1995. Testing the Mean of Skewed Distributions. Journal of the American Statistical Association, 90, 767-772.

Choi, S. C. and Wette, R. 1969. Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. Technometrics, Vol. 11, 683-690.Cochran, W. 1977. Sampling Techniques, New York: John Wiley.

Cohen, A. C., Jr. 1950. Estimating the Mean and Variance of Normal Populations from Singly Truncated and Double Truncated Samples. Ann. Math. Statist., Vol. 21, pp. 557-569.

Cohen, A. C., Jr. 1959. Simplified Estimators for the Normal Distribution When Samples Are Singly Censored or Truncated. Technometrics, Vol. 1, No. 3, pp. 217-237.

Cohen, A. C., Jr. 1991. Truncated and Censored Samples. 119, Marcel Dekker Inc. New York, NY 1991. Conover W.J.. 1999. Practical Nonparametric Statistics, 3rd Edition, John Wiley & Sons, New York.

D'Agostino, R.B. and Stephens, M.A. 1986. Goodness-of-Fit Techniques. Marcel Dekker, Inc. Daniel, Wayne W. 1995. Biostatistics. 6th Edition. John Wiley & Sons, New York.

David, H.A. and Nagaraja, H.N. 2003. Order Statistics. Third Edition. John Wiley.

Department of Navy. 2002a. Guidance for Environmental Background Analysis. Volume 1 Soil. Naval Facilities Engineering Command. April 2002.

Department of Navy. 2002b. Guidance for Environmental Background Analysis. Volume 2 Sediment. Naval Facilities Engineering Command. May 2002.

Dixon, W.J. 1953. Processing Data for Outliers. Biometrics 9: 74-89.

Draper, N.R. and Smith, H. 1998. Applied Regression Analysis (3rd Edition). New York: John Wiley & Sons.

Dudewicz, E.D. and Misra, S.N. 1988. Modern Mathematical Statistics. John Wiley, New York.

Efron, B. 1981. Censored Data and Bootstrap. Journal of American Statistical Association, Vol. 76, pp. 312-319.

Efron, B. 1982. The Jackknife, the Bootstrap, and Other Resampling Plans, Philadelphia: SIAM. Efron, B. and Tibshirani, R.J. 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

El-Shaarawi, A.H. 1989. Inferences about the Mean from Censored Water Quality Data. Water Resources Research, 25, pp. 685-690.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7 (2): 179–188.

Fleischhauer, H. and Korte, N. 1990. Formation of Cleanup Standards Trace Elements with Probability Plot. Environmental Management, Vol. 14, No. 1. 95-105.

Gehan, E.A. 1965. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Sample.

Biometrika 52, 203-223.Gerlach, R. W., and J. M. Nocerino. 2003. Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples. EPA/600/R-03/027. www.epa.gov/esd/tsc/images/particulate.pdf.

Gibbons. 1994. Statistical Methods for Groundwater Monitoring. John Wiley &Sons.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Gilliespie, B.W., Chen, Q., Reichert H., Franzblau A., Hedgeman E., Lepkowski J., Adriaens P., Demond A., Luksemburg W., and Garabrant DH. 2010. Estimating population distributions when some data are below a limit of detection by using a reverse Kaplan-Meier estimator. Epidemiology, Vol. 21, No. 4.

Gleit, A. 1985. Estimation for Small Normal Data Sets with Detection Limits. Environmental Science and Technology, 19, pp. 1206-1213, 1985.

Grice, J.V., and Bain, L. J. 1980. Inferences Concerning the Mean of the Gamma Distribution. Journal of the American Statistical Association. Vol. 75, Number 372, 929-933.

Gu, M.G., and Zhang, C.H. 1993. Asymptotic properties of self-consistent estimators based on doubly censored data. Annals of Statistics. Vol. 21, 611-624.

Hahn, J. G. and Meeker, W.Q. 1991. Statistical Intervals. A Guide for Practitioners. John Wiley.

Hall, P. 1988. Theoretical comparison of bootstrap confidence intervals. Annals of Statistics, 16, 927- 953.

Hall, P. 1992. On the Removal of Skewness by Transformation. Journal of Royal Statistical Society, B 54, 221-228.

Hardin, J.W. and Gilbert, R.O. 1993. Comparing Statistical Tests for Detecting Soil Contamination Greater Than Background. Pacific Northwest Laboratory, Battelle, Technical Report # DE 94-005498.

Hawkins, D. M., and Wixley, R. A. J. 1986. A Note on the Transformation of Chi-Squared Variables to Normality. The American Statistician, 40, 296–298.

Hayes, A. F. 2005. Statistical Methods for Communication Science, Lawrence Erlbaum Associates, Publishers.

Helsel, D.R. 2005. Nondetects and Data Analysis. Statistics for Censored Environmental Data. John Wiley and Sons, NY.

Helsel, D.R. 2102a. Practical Stats Webinar on ProUCL v4. The Unofficial User Guide; October 15, 2012.

Helsel, D.R. 2012b. Statistics for Censored Environmental Data Using Minitab and R. Second Edition. John Wiley and Sons, NY.

Helsel, D.R. 2013. Nondetects and Data Analysis for Environmental Data, NADA in R

Helsel, D.R. and E. J. Gilroy. 2012. The Unofficial Users Guide to ProUCL4. Amazon, Kindle Edition.

Hinton, S.W. 1993. □ Log-Normal Statistical Methodology Performance. ES&T Environmental Sci. Technol., Vol. 27, No. 10, pp. 2247-2249.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W. 1983. Understanding Robust and Exploratory Data Analysis. John Wiley, New York.

Holgresson, M. and Jorner U. 1978. Decomposition of a Mixture into Normal Components: a Review.

Journal of Bio-Medicine. Vol. 9. 367-392.

Hollander M & Wolfe DA (1999). Nonparametric Statistical Methods (2nd Edition). New York: John Wiley & Sons.

Hogg, R.V. and Craig, A. 1995. Introduction to Mathematical Statistics; 5th edition. Macmillan. Huber, P.J. 1981, Robust Statistics, John Wiley and Sons, NY.

Hyndman, R. J. and Fan, Y. 1996. Sample quantiles in statistical packages, American Statistician, 50, 361–365.

Interstate Technology Regulatory Council (ITRC). 2012. Incremental Sampling Methodology. Technical and Regulatory Guidance, 2012.

Interstate Technology Regulatory Council (ITRC). 2013 Groundwater Statistics and Monitoring Compliance. Technical and Regulatory Guidance, December 2013.

Interstate Technology Regulatory Council (ITRC). 2015. Decision Making at Contaminated Sites.

Interstate Technology Regulatory Council (ITRC). 2020. Updated Incremental Sampling Methodology. Technical and Regulatory Guidance, 2020.

Johnson, N.J. 1978. Modified-t-Tests and Confidence Intervals for Asymmetrical Populations. The American Statistician, Vol. 73, 536-544.

Johnson, N.L., Kotz, S., and Balakrishnan, N. 1994. Continuous Univariate Distributions, Vol. 1. Second Edition. John Wiley, New York.

Johnson, R.A. and D. Wichern. 2002. Applied Multivariate Statistical Analysis. 6th Edition. Prentice Hall.

Kaplan, E.L. and Meier, O. 1958. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, Vol. 53. 457-481.

Kleijnen, J.P.C., Kloppenburg, G.L.J., and Meeuwsen, F.L. 1986. Testing the Mean of an Asymmetric Population: Johnson's Modified-t Test Revisited. Commun. in Statist.-Simula., 15(3), 715-731.

Krishnamoorthy, K., Mathew, T., and Mukherjee, S. 2008. Normal distribution based methods for a Gamma distribution: Prediction and Tolerance Interval and stress-strength reliability. Technometrics, 50, 69-78.

Kroese, D.P., Taimre, T., and Botev Z.I. 2011. Handbook of Monte Carlo Methods. John Wiley & Sons.

Kruskal, W. H., and Wallis, A. 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47, 583-621.

Kupper, L. L. and Hafner, K. B. 1989, How Appropriate Are Popular Sample Size Formulas? The American Statistician, Vol. 43, No. 2, pp. 101-105

Kunter, M. J., C. J. Nachtsheim, J. Neter, and Li W. 2004. Applied Linear Statistical Methods. Fifth Edition. McGraw-Hill/Irwin.

Laga, J., and Likes, J. 1975, Sample Sizes for Distribution-Free Tolerance Intervals Statistical Papers. Vol. 16, No. 1. 39-56

Land, C. E. 1971. Confidence Intervals for Linear Functions of the Normal Mean and Variance. Annals of Mathematical Statistics, 42, pp. 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., pp. 385-419.

Levene, Howard. 1960. Robust tests for equality of variances. In Olkin, Harold, et alia. Stanford University Press. pp. 278–292.

Lilliefors, H.W. 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. Journal of the American Statistical Association, 62, 399-404.

Looney and Gulledge. 1985. Use of the Correlation Coefficient with Normal Probability Plots. The American Statistician, 75-79.

Manly, B.F.J. 1997. Randomization, Bootstrap, and Monte Carlo Methods in Biology. Second Edition. Chapman Hall, London.

Maronna, R.A., Martin, R.D., and Yohai, V.J. 2006, Robust Statistics: Theory and Methods, John Wiley and Sons, Hoboken, NJ.

Marsaglia, G. and Tsang, W. 2000. A simple method for generating gamma variables. ACM Transactions on Mathematical Software, 26(3):363-372.

Millard, S. P. and Deverel, S. J. 1988. Nonparametric statistical methods for comparing two sites based on data sets with multiple nondetect limits. Water Resources Research, 24, pp. 2087-2098.

Millard, S.P. and Neerchal, M.K. 2002. Environmental Stats for S-PLUS. Second Edition. Springer. Minitab version 16. 2012. Statistical Software.

Molin, P., and Abdi H. 1998. New Tables and numerical approximations for the Kolmogorov-Smirnov/Lilliefors/ Van Soest’s test of normality. In Encyclopedia of Measurement and Statistics, Neil Salkind (Editor, 2007). Sage Publication Inc. Thousand Oaks (CA).

Natrella, M.G. 1963. Experimental Statistics. National Bureau of Standards, Hand Book No. 91, U.S. Government Printing Office, Washington, DC.

Noether, G.E. 1987 Sample Size Determination for some Common Nonparametric Tests, Journal American Statistical Assoc., 82, 645-647

Perrson, T., and Rootzen, H. 1977. Simple and Highly Efficient Estimators for A Type I Censored Normal Sample. Biometrika, 64, pp. 123-128.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. 1990. Numerical Recipes in C, The Art of Scientific Computing. Cambridge University Press. Cambridge, MA.

R Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rosner, B. 1975. On the detection of many outliers. Technometrics, 17, 221-227.

Rosner, B. 1983. Percentage points for a generalized ESD many-outlier procedure. Technometrics, 25, 165-172.

Rousseeuw, P.J. and Leroy, A.M. 1987. Robust Regression and Outlier Detection. John Wiley.

Royston, P. 1982a. Algorithm AS 181: The W test for Normality. Applied Statistics, 31, 176–180.

Royston, P. 1982b. An extension of Shapiro and Wilk's W test for normality to large samples. Applied Statistics, 31, 115–124.

Shacklette, H.T, and Boerngen, J.G. 1984. Element Concentrations in Soils and Other Surficial Materials in the Conterminous United States, U.S. Geological Survey Professional Paper 1270.

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.

Scheffe, H., and Tukey, J.W. 1944. A formula for Sample Sizes for Population Tolerance Limits. The Annals of Mathematical Statistics. Vol 15, 217.

Schneider, B.E. and Clickner, R.P. 1976. On the Distribution of the Kolmogorov-Smirnov Statistic for the Gamma Distribution with Unknown Parameters. Mimeo Series Number 36, Department of Statistics, School of Business Administration, Temple University, Philadelphia, PA.

Schneider, B. E. 1978. Kolmogorov-Smirnov Test Statistic for the Gamma Distribution with Unknown Parameters, Dissertation, Department of Statistics, Temple University, Philadelphia, PA.

Schneider, H. 1986. Truncated and Censored Samples from Normal Populations. Vol. 70, Marcel Dekker Inc., New York, 1986.

She, N. 1997. Analyzing Censored Water Quality Data Using a Nonparametric Approach. Journal of the American Water Resources Association 33, pp. 615-624.

Shea, B. 1988. Algorithm AS 239: Chi-square and Incomplete Gamma Integrals. Applied Statistics, 37: 466-473.

Shumway, A.H., Azari, A.S., Johnson, P. 1989. Estimating Mean Concentrations Under Transformation for Environmental Data with Detection Limits. Technometrics, Vol. 31, No. 3, pp. 347-356.

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.

Sinclair, A.J. 1976. Applications of Probability Graphs in Mineral Exploration. Association of Exploration Geochemists, Rexdale Ontario, p 95.

Singh, A. 1993. Omnibus Robust Procedures for Assessment of Multivariate Normality and Detection of Multivariate Outliers. In Multivariate Environmental Statistics, Patil G.P. and Rao, C.R., Editors, pp. 445-488. Elsevier Science Publishers.

Singh, A. 2004. Computation of an Upper Confidence Limit (UCL) of the Unknown Population Mean Using ProUCL Version 3.0. Part I. Download from: www.epa.gov/nerlesd1/tsc/issue.htm

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. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

Singh, A. and Nocerino, J.M. 1995. Robust Procedures for the Identification of Multiple Outliers. Handbook of Environmental Chemistry, Statistical Methods, Vol. 2.G, pp. 229-277. Springer Verlag, Germany.

Singh, A. and Nocerino, J.M. 1997. Robust Intervals for Some Environmental Applications." The Journal of Chemometrics and Intelligent Laboratory Systems, Vol 37, 55-69.

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.K. and Ananda. M. 2002. Rank kriging for characterization of mercury contamination at the East Fork Poplar Creek, Oak Ridge, Tennessee. Environmetrics, Vol. 13, pp. 679-691.

Singh, A. and Singh, A.K. 2007. ProUCL Version 4 Technical Guide (Draft). Publication EPA/600/R-07/041. January, 2007. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

Singh, A. and Singh, A.K. 2009. ProUCL Version 4.00.04 Technical Guide (Draft). Publication EPA/600/R-07/041. February, 2009. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

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, December 1997.

Singh, A., Singh A.K., and Engelhardt, M. 1999, Some Practical Aspects of sample Size and Power Computations for Estimating the Mean of Positively Skewed Distributions in Environmental Applications.

Office of Research and Development. EPA/006/s-99/006. November 1999. http://www.epa.gov/esd/tsc/images/325cmb99rpt.pdf

Singh, A., Singh, A.K., and Flatman, G. 1994. Estimation of Background Levels of Contaminants. Math Geology, Vol. 26, No, 3, 361-388.

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.

Stephens, M. A. 1970. Use of Kolmogorov-Smirnov, Cramer-von Mises and Related Statistics Without Extensive Tables. Journal of Royal Statistical Society, B 32, 115-122.

Sutton, C.D. 1993. Computer-Intensive Methods for Tests About the Mean of an Asymmetrical Distribution. Journal of American Statistical Society, Vol. 88, No. 423, 802-810.

Tarone, R. and Ware, J. 1978. On Distribution-free Tests for Equality of Survival Distributions.

Biometrika, 64, 156-160.

Thom, H.C.S. 1968. Direct and Inverse Tables of the Gamma Distribution. Silver Spring, MD; Environmental Data Service.

U.S. Environmental Protection Agency (EPA). 1989a. Methods for Evaluating the Attainment of Cleanup Standards, Vol. 1, Soils and Solid Media. Publication EPA 230/2-89/042. Available at https://clu-in.org/download/stats/vol1soils.pdf

U.S. Environmental Protection Agency (EPA). 1989b. Statistical Analysis of Ground-water Monitoring Data at RCRA Facilities. Interim Final Guidance. Washington, DC: Office of Solid Waste. April 1989. Available at

https://epa.ohio.gov/Portals/30/hazwaste/GW/1989%20USEPA%20RCRA_interim_final_guidance.pdf

U.S. Environmental Protection Agency (EPA). 1992a. Methods for Evaluating the Attainment of Cleanup Standards, Volume 2: Ground Water. Publication 230-R-92-014. Available at https://semspub.epa.gov/work/HQ/175643.pdf

U.S. Environmental Protection Agency (EPA). 1992b. Statistical Analysis of Ground-water Monitoring Data at RCRA Facilities. Addendum to Interim Final Guidance. Washington DC: Office of Solid Waste. July 1992. Available at https://www.wipp.energy.gov/library/Information_Repository_A/Supplemental_Information/EPA%201992.pdf

U.S. Environmental Protection Agency (EPA). 1992c. Supplemental Guidance to RAGS: Calculating the Concentration Term. Publication EPA 9285.7-081, May 1992. Available at https://semspub.epa.gov/work/10/500011427.pdf

U.S. Environmental Protection Agency (EPA). 1996. Soil Screening Guidance: Technical Background Document. Second Edition, Publication EPA/540/R95/128. Available at https://semspub.epa.gov/work/HQ/207.pdf

U.S. Environmental Protection Agency (EPA). MARSSIM. 2000. U.S. Nuclear Regulatory Commission, et al. Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM). Revision 1. EPA 402-R-97-016. Available at http://www.epa.gov/radiation/marssim/ or from http://bookstore.gpo.gov/index.html (GPO Stock Number for Revision 1 is 052-020-00814-1).

U.S. Environmental Protection Agency (EPA). 2002a. Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites. OSWER 9285.6-10. December 2002. Available at https://www.epa.gov/sites/default/files/2016-03/documents/upper-conf-limits.pdf

U.S. Environmental Protection Agency (EPA). 2002b. Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. EPA 540-R-01-003-OSWER 9285.7-41. September 2002. Available at Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites (epa.gov)

U.S. Environmental Protection Agency (EPA). 2002c. RCRA Waste Sampling, Draft Technical Guidance

– Planning, Implementation and Assessment. EPA 530-D-02-002, 2002. Available at RCRA Waste Sampling Draft Technical Guidance (epa.gov)

U.S. Environmental Protection Agency (EPA). 2004. ProUCL Version 3.1, Statistical Software. National Exposure Research Lab, EPA, Las Vegas Nevada, October 2004. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2006a, Guidance on Systematic Planning Using the Data Quality Objective Process, C, EPA/240/B-06/001. Office of Environmental Information, Washington, DC. Download from: https://www.epa.gov/sites/default/files/2015-06/documents/g4-final.pdf

U.S. Environmental Protection Agency (EPA). 2006b. Data Quality Assessment: Statistical Methods for Practitioners, EPA QA/G-9S. EPA/240/B-06/003. Office of Environmental Information, Washington, DC. Download from: https://www.epa.gov/sites/default/files/2015-08/documents/g9s-final.pdf

U.S. Environmental Protection Agency (EPA). 2007. ProUCL Version 4.0 Technical Guide. EPA 600-R-07-041, January 2007. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2009a. ProUCL Version 4.00.05 User Guide (Draft). Statistical Software for Environmental Applications for Data Sets with and without nondetect observations. National Exposure Research Lab, EPA, Las Vegas. EPA/600/R-07/038, February 2009. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2009b. ProUCL Version 4.00.05 Technical Guide (Draft). Statistical Software for Environmental Applications for Data Sets with and without nondetect observations. National Exposure Research Lab, EPA, Las Vegas. EPA/600/R-07/038, February 2009. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2009c. ProUCL4.00.05 Facts Sheet. Statistical Software for Environmental Applications for Data Sets with and without nondetect observations. National Exposure Research Lab, EPA, Las Vegas, Nevada, 2009. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2009d. Scout 2008 – A Robust Statistical Package, Office of Research and Development, February 2009. http://archive.epa.gov/esd/archive-scout/web/html/

U.S. Environmental Protection Agency (EPA). 2009e. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance. EPA 530-R-09-007, 2009. Available at https://archive.epa.gov/epawaste/hazard/web/pdf/unified-guid-toc.pdf

U.S. Environmental Protection Agency (EPA). 2010a. A Quick Guide to the Procedures in Scout (Draft),Office of Research and Development, April 2010. http://archive.epa.gov/esd/archive-scout/web/html/

U.S. Environmental Protection Agency (EPA). 2010b. ProUCL Version 4.00.05 User Guide. EPA/600/R-07/041, May 2010. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2010c. ProUCL Version 4.00.05 Technical Guide. EPA/600/R-07/041, May, 2010. https://www.epa.gov/land-research/proucl-software#references U.S. Environmental Protection Agency (EPA). 2010d. ProUCL 4.00.05, Statistical Software for Environmental Applications for Data Sets with and without nondetect observations. National Exposure Research Lab, EPA, Las Vegas Nevada, May 2010. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2011. ProUCL 4.1.00, Statistical Software for Environmental Applications for Data Sets with and without nondetect observations. National Exposure Research Lab, EPA, Las Vegas Nevada, June 2011. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2013a. ProUCL 5.0.00 Technical Guide EPA/600/R-07/041. September 2013. Office of Research and Development. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2013b. ProUCL 5.0.00 User Guide EPA/600/R- 07/041. September 2013. Office of Research and Development. https://www.epa.gov/land-research/proucl-software#references

U.S. Environmental Protection Agency (EPA). 2014. ProUCL 5.0.00 Statistical Software for Environmental Applications for Datasets with and without Nondetect Observations, Office of Research and Development, August 2014. https://www.epa.gov/land-research/proucl-software#references

Wald, A. 1943. An Extension of Wilks' Method for Setting Tolerance Intervals. Annals of Mathematical Statistics. Vol. 14, 44-55.

Whittaker, J. 1974. Generating Gamma and Beta Random Variables with Non-integral Shape Parameters. Applied Statistics, 23, No. 2, 210-214.

Wilks, S.S. 1941. Determination of Sample Sizes for Setting Tolerance Limits. Annals of Mathematical Statistics, Vol. 12, 91-96.

Wilks, S.S. 1963. Multivariate statistical outliers. Sankhya A, 25: 407-426.

Wilson, E.B., and Hilferty, M.M. 1931, "The Distribution of Chi-Squares," Proceedings of the National Academy of Sciences, 17, 684–688.

Wong, A. 1993. A Note on Inference for the Mean Parameter of the Gamma Distribution. Statistics Probability Letters, Vol. 17, 61-66.

ProUCL UTILIZATION TRAINING

A three-part ProUCL Utilization training was performed in 2020 to help users familiarize with ProUCL functionalities. Each section is approximately 2 hours long and can be played back on demand.

Recordings of this training are available on the EPA CLU-IN web site:

ProUCL Utilization 2020: Part 1: ProUCL A to Z
https://clu-in.org/conf/tio/ProUCLAtoZ1/

Topics:

- Navigating ProUCL
- Starting ProUCL and loading data
- Organizing data
    - Nondetects
    - Missing data
- Exploratory Data Analysis (EDA)
    - Box plot
    - Q-Q plot
- Evaluating the distribution of the data
- Outliers
- Hypothesis testing

ProUCL Utilization 2020: Part 2: Trend Analysis
https://clu-in.org/conf/tio/ProUCLAtoZ2/

Topics:

- Dealing with nondetects in trend analysis
- Time series plot
- Trend Analysis
    - Mann-Kendall
    - Thei-Sen
- Ordinary Least Square Regression

ProUCL Utilization 2020: Part 3: Background Level Calculations
https://clu-in.org/conf/tio/ProUCLAtoZ3/

Topics:

- Coverage vs confidence
- Background Treshold Values (BTV)
    - Upper percentiles
    - Upper prediction limits (UPL)
    - Upper confidence limits (UCL)
    - Upper tolerance limits (UTL)
    - Upper simultaneous limits (USL)

GLOSSARY

Anderson-Darling (A-D) test: The Anderson-Darling test assesses whether known data come from a specified distribution. In ProUCL the A-D test is used to test the null hypothesis that a sample data set, $x_1,..., x_n$ came from a gamma distributed population.

Background Measurements: Measurements that are not site-related or impacted by site activities. Background sources can be naturally occurring or anthropogenic (man-made).

Bias: The systematic or persistent distortion of a measured value from its true value (this can occur during sampling design, the sampling process, or laboratory analysis).

Bootstrap Method: The bootstrap method is a computer-based method for assigning measures of accuracy to sample estimates. This technique allows estimation of the sample distribution of almost any statistic using only very simple methods. Bootstrap methods are generally superior to ANOVA for small data sets or where sample distributions are non-normal.

Central Limit Theorem (CLT): The central limit theorem states that given a distribution with a mean, μ, and variance, σ^2, the sampling distribution of the mean approaches a normal distribution with a mean (μ) and a variance σ^2/N as N, the sample size, increases.

Censored Data Sets: Data sets that contain one or more observations which are nondetects.

Coefficient of Variation (CV): A dimensionless quantity used to measure the spread of data relative to the size of the numbers. For a normal distribution, the coefficient of variation is given by s/xBar. It is also known as the relative standard deviation (RSD).

Confidence Coefficient (CC): The confidence coefficient (a number in the closed interval [0, 1]) associated with a confidence interval for a population parameter is the probability that the random interval constructed from a random sample (data set) contains the true value of the parameter. The confidence coefficient is related to the significance level of an associated hypothesis test by the equality: level of significance = 1 – confidence coefficient.

Confidence Interval: Based upon the sampled data set, a confidence interval for a parameter is a random interval within which the unknown population parameter, such as the mean, or a future observation, x_0, falls.

Confidence Limit: The lower or an upper boundary of a confidence interval. For example, the 95% upper confidence limit (UCL) is given by the upper bound of the associated confidence interval.

Coverage, Coverage Probability: The coverage probability (e.g., = 0.95) of an upper confidence limit (UCL) of the population mean represents the confidence coefficient associated with the UCL.

Critical Value: The critical value for a hypothesis test is a threshold to which the value of the test statistic is compared to determine whether or not the null hypothesis is rejected. The critical value for any hypothesis test depends on the sample size, the significance level, α at which the test is carried out, and whether the test is one-sided or two-sided.

Data Quality Objectives (DQOs): Qualitative and quantitative statements derived from the DQO process that clarify study technical and quality objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

Detection Limit: A measure of the capability of an analytical method to distinguish samples that do not contain a specific analyte from samples that contain low concentrations of the analyte. It is the lowest concentration or amount of the target analyte that can be determined to be different from zero by a single measurement at a stated level of probability. Detection limits are analyte and matrix-specific and may be laboratory-dependent.

Empirical Distribution Function (EDF): In statistics, an empirical distribution function is a cumulative probability distribution function that concentrates probability $1/n$ at each of the n numbers in a sample.

Estimate: A numerical value computed using a random data set (sample), and is used to guess (estimate) the population parameter of interest (e.g., mean). For example, a sample mean represents an estimate of the unknown population mean.

Expectation Maximization (EM): The EM algorithm is used to approximate a probability density function (PDF). EM is typically used to compute maximum likelihood estimates given incomplete samples.

Exposure Point Concentration (EPC): The constituent concentration within an exposure unit to which the receptors are exposed. Estimates of the EPC represent the concentration term used in exposure assessment.

Extreme Values: Values that are well-separated from the majority of the data set coming from the far/extreme tails of the data distribution.

Goodness-of-Fit (GOF): In general, the level of agreement between an observed set of values and a set wholly or partly derived from a model of the data.

Gray Region: A range of values of the population parameter of interest (such as mean constituent concentration) within which the consequences of making a decision error are relatively minor. The gray region is bounded on one side by the action level. The width of the gray region is denoted by the Greek letter delta, Δ, in this guidance.

H-Statistic: Land's statistic used to compute UCL of mean of a lognormal population

H-UCL: UCL based on Land's H-Statistic.

Hypothesis: Hypothesis is a statement about the population parameter(s) that may be supported or rejected by examining the data set collected for this purpose. There are two hypotheses: a null hypothesis, (H_0), representing a testable presumption (often set up to be rejected based upon the sampled data), and an alternative hypothesis (H_A), representing the logical opposite of the null hypothesis.

Jackknife Method: A statistical procedure in which, in its simplest form, estimates are formed of a parameter based on a set of N observations by deleting each observation in turn to obtain, in addition to the usual estimate based on N observations, N estimates each based on N-1 observations.

Kolmogorov-Smirnov (KS) test: The Kolmogorov-Smirnov test is used to decide if a data set comes from a population with a specific distribution. The Kolmogorov-Smirnov test is based on the empirical distribution function (EDF). ProUCL uses the KS test to test the null hypothesis if a data set follows a gamma distribution.

Left-censored Data Set: An observation is left-censored when it is below a certain value (detection limit) but it is unknown by how much; left-censored observations are also called nondetect (ND) observations. A data set consisting of left-censored observations is called a left-censored data set. In environmental applications trace concentrations of chemicals may indeed be present in an environmental sample (e.g., groundwater, soil, sediment) but cannot be detected and are reported as less than the detection limit of the analytical instrument or laboratory method used.

Level of Significance (α): The error probability (also known as false positive error rate) tolerated of falsely rejecting the null hypothesis and accepting the alternative hypothesis.

Lilliefors test: A goodness-of-fit test that tests for normality of large data sets when population mean and variance are unknown.

Maximum Likelihood Estimates (MLE): MLE is a popular statistical method used to make inferences about parameters of the underlying probability distribution of a given data set.

Mean: The sum of all the values of a set of measurements divided by the number of values in the set; a measure of central tendency.

Median: The middle value for an ordered set of n values. It is represented by the central value when n is odd or by the average of the two most central values when n is even. The median is the 50th percentile.

Minimum Detectable Difference (MDD): The MDD is the smallest difference in means that the statistical test can resolve. The MDD depends on sample-to-sample variability, the number of samples, and the power of the statistical test.

Minimum Variance Unbiased Estimates (MVUE): A minimum variance unbiased estimator (MVUE or MVU estimator) is an unbiased estimator of parameters, whose variance is minimized for all values of the parameters. If an estimator is unbiased, then its mean squared error is equal to its variance.

Nondetect (ND) values: Censored data values. Typically, in environmental applications, concentrations or measurements that are less than the analytical/instrument method detection limit or reporting limit.

Nonparametric: A term describing statistical methods that do not assume a particular population probability distribution, and are therefore valid for data from any population with any probability distribution, which can remain unknown.

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.

Outlier: Measurements (usually larger or smaller than the majority of the data values in a sample) that are not representative of the population from which they were drawn. The presence of outliers distorts most statistics if used in any calculations.

Probability - Values (*p*-value): In statistical hypothesis testing, the p-value associated with an observed value, t_{observed} of some random variable T used as a test statistic is the probability that, given that the null hypothesis is true, T will assume a value as or more unfavorable to the null hypothesis as the observed value t_{observed}. The null hypothesis is rejected for all levels of significance, α greater than or equal to the p- value.

Parameter: A parameter is an unknown or known constant associated with the distribution used to model the population.

Parametric: A term describing statistical methods that assume a probability distribution such as a normal, lognormal, or a gamma distribution.

Population: The total collection of N objects, media, or people to be studied and from which a sample is to be drawn. It is the totality of items or units under consideration.

Prediction Interval: The interval (based upon historical data, background data) within which a newly and independently obtained (often labeled as a future observation) site observation (e.g., onsite, compliance well) of the predicted variable (e.g., lead) falls with a given probability (or confidence coefficient).

Probability of Type II (2) Error (β): The probability, referred to as β (beta), that the null hypothesis will not be rejected when in fact it is false (false negative).

Probability of Type I (1) Error = Level of Significance (α): The probability, referred to as α (alpha), that the null hypothesis will be rejected when in fact it is true (false positive).

pth Percentile or pth Quantile: The specific value, X_p of a distribution that partitions a data set of measurements in such a way that the p percent (a number between 0 and 100) of the measurements fall at or below this value, and (100-p) percent of the measurements exceed this value, X_p.

Quality Assurance (QA): An integrated system of management activities involving planning, implementation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the client.

Quality Assurance Project Plan: A formal document describing, in comprehensive detail, the necessary QA, quality control (QC), and other technical activities that must be implemented to ensure that the results of the work performed will satisfy the stated performance criteria.

Quantile Plot: A graph that displays the entire distribution of a data set, ranging from the lowest to the highest value. The vertical axis represents the measured concentrations, and the horizontal axis is used to plot the percentiles/quantiles of the distribution.

Range: The numerical difference between the minimum and maximum of a set of values.

Regression on Order Statistics (ROS): A regression line is fit to the normal scores of the order statistics for the uncensored observations and is used to fill in values imputed from the straight line for the observations below the detection limit.

Resampling: The repeated process of obtaining representative samples and/or measurements of a population of interest.

Reliable UCL: see Stable UCL.

Robustness: Robustness is used to compare statistical tests. A robust test is the one with good performance (that is not unduly affected by outliers and underlying assumptions) for a wide variety of data distributions.

Resistant Estimate: A test/estimate which is not affected by outliers is called a resistant test/estimate

Sample: Represents a random sample (data set) obtained from the population of interest (e.g., a site area, a reference area, or a monitoring well). The sample is supposed to be a representative sample of the population under study. The sample is used to draw inferences about the population parameter(s).

Shapiro-Wilk (SW) test: Shapiro-Wilk test is a goodness-of-fit test that tests the null hypothesis that a sample data set, $x_1, ..., x_n$ came from a normally distributed population.

Skewness: A measure of asymmetry of the distribution of the parameter under study (e.g., lead concentrations). It can also be measured in terms of the standard deviation of log-transformed data. The greater the standard deviation, the greater is the skewness.

Stable UCL: The UCL of a population mean is a stable UCL if it represents a number of practical merit (e.g., a realistic value which can actually occur at a site), which also has some physical meaning. That is, a stable UCL represents a realistic number (e.g., constituent 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.

Standard Deviation (*sd*, sd, SD): A measure of variation (or spread) from an average value of the sample data values.

Standard Error (SE): A measure of an estimate's variability (or precision). The greater the standard error in relation to the size of the estimate, the less reliable is the estimate. Standard errors are needed to construct confidence intervals for the parameters of interests such as the population mean and population percentiles.

Substitution Method: The substitution method is a method for handling NDs in a data set, where the ND is replaced by a defined value such as 0, DL/2 or DL prior to statistical calculations or graphical analyses. This method has been included in ProUCL 5.1 for historical comparative purposes but **is not recommended**

for use. The **bias** introduced by applying the substitution method **cannot be quantified** with any certainty. ProUCL 5.1 will provide a warning when this option is chosen.

Uncensored Data Set: A data set without any censored (nondetects) observations.

Unreliable UCL, Unstable 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 a 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 an impractically large inflated UCL value. Some other 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.

Upper Confidence Limit (UCL): The upper boundary (or limit) of a confidence interval of a parameter of interest such as the population mean.

Upper Prediction Limit (UPL): The upper boundary of a prediction interval for an independently obtained observation (or an independent future observation).

Upper Tolerance Limit (UTL): A confidence limit on a percentile of the population rather than a confidence limit on the mean. For example, a 95% one-sided UTL for 95% coverage represents the value below which 95% of the population values are expected to fall with 95 % confidence. In other words, a 95% UTL with coverage coefficient 95% represents a 95% UCL for the 95^{th} percentile.

Upper Simultaneous Limit (USL): The upper boundary of the largest value.

xBar: arithmetic average of computed using the sampled data values