United States
Environmental Protection
Agency

ProUCL Version 5.0.00
User Guide

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

RESEARCH AND DEVELOPMENT

EPA/600/R-07/041
September 2013
www.epa.gov

ProUCL Version 5.0.00 User Guide

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

Prepared for:

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

Prepared by:

Anita Singh, Ph.D. and Robert Maichle

Lockheed Martin IS&GS-CIVIL
2890 Woodbridge Ave
Edison NJ 08837

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

129cmb07

NOTICE

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

- ProUCL software was developed by Lockheed Martin, IS&GS - CIVIL under a contract with the EPA and is made available through the EPA Technical Support Center in Atlanta, Georgia.

- 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.

- ProUCL software provided by the EPA was scanned with McAfee VirusScan v4.5.1 SP1 and is certified free of viruses.

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

Minimum Hardware Requirements

ProUCL 5.0.00 will function but will run slowly and page a lot.

- Intel Pentium 1.0 GHz
- 45 MB of hard drive space
- 512 MB of memory (RAM)
- CD-ROM drive or internet connection
- Windows XP (with SP3), Vista (with SP1 or later), and Windows 7.

ProUCL 5.0.00 will function but some titles and some Graphical User Interfaces (GUIs) will need to be scrolled. Definition without color will be marginal.

- 800 by 600 Pixels
- Basic Color is preferred

Preferred Hardware Requirements

- 1 gigahertz (GHz) or faster Processor.
- 1 gigabyte (GB) of memory (RAM)
- 1024 by 768 Pixels or greater color display

Software Requirements

ProUCL 5.0.00 has been developed in the Microsoft .NET Framework 4.0 using the C# programming language. To properly run ProUCL 5.0.00 software, the computer using the program must have the .NET Framework 4.0 pre-installed. The downloadable .NET Framework 4.0 files can be obtained from one of the following websites:

- http://msdn.microsoft.com/netframework/downloads/updates/default.aspx
  http://www.microsoft.com/en-us/download/details.aspx?id=17851
  Quicker site for 32 Bit Operating systems

- http://www.microsoft.com/en-us/download/details.aspx?id=24872
  Use this site if you have a 64 Bit operating system

Installation Instructions when Downloading from the EPA Web Site

- Download the file SETUP.EXE from the EPA Web site and save to a temporary location.

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

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

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

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

Installation Instructions when Copying from a CD

- Create a folder named **ProUCL 5.0** on a local hard drive of the machine you wish to install ProUCL 5.0.

- Extract the zipped file **ProUCL.zip** to the folder you have just created.

- Run **ProUCL.exe**.

Note: If you have extension turned off, the program will show with the name **ProUCL** in your directory and have an Icon with the label **ProUCL**.

Creating a Shortcut for ProUCL 5.0 on Desktop

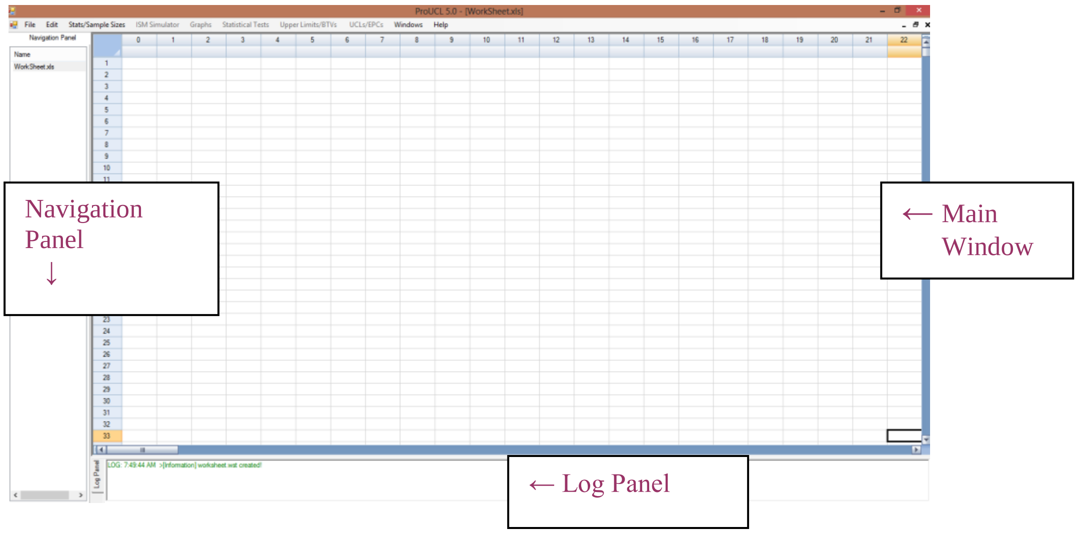
- To create a shortcut of the ProUCL program on your desktop, go to your ProUCL directory and right click on the executable program and send it to desktop. A ProUCL icon will be displayed on your desktop. This shortcut will point to the ProUCL directory consisting of all files required to execute ProUCL 5.0.

Caution: It should be noted that since all files in your ProUCL directory are needed to execute the ProUCL software, one needs you generate a shortcut using the process described above. Specifically, 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.

Getting Started

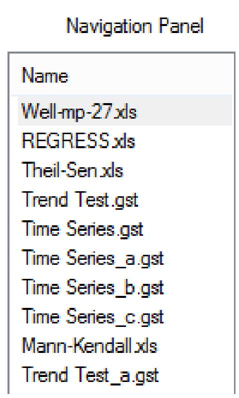
The functionality and the use of the methods and options available in ProUCL 5.0 have been illustrated using Screen shots of output screen generated by ProUCL 5.0. ProUCL 5.0 uses a pull-down menu structure, similar to a typical Windows program.

The screen shown below appears when the program is executed.



The above 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. 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.

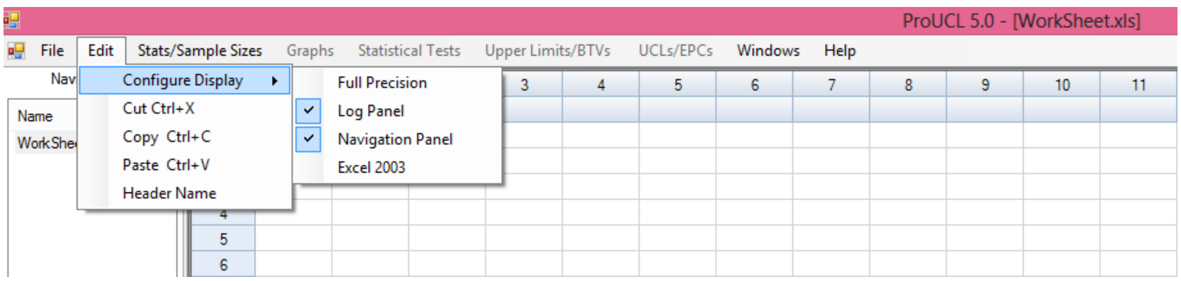


- The user may want to assign names of his choice to these output files when saving them using the "Save" or "Save As" Options.

- The **LOG PANEL** displays transactions in green, warnings in orange, and errors in red. For an example, when one attempts to run a procedure meant for left-censored data sets on a full-uncensored data set, ProUCL 5.0 will print out a warning message in orange in this panel.

    - Should both panels be unnecessary, you can choose **Configure ► Panel ON/OFF**.

The use of this option gives extra 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.



EXECUTIVE SUMMARY

The main objective of the ProUCL software funded by the USEPA is to compute rigorous statistics to help decision makers and project teams in making correct decisions at a polluted site which are cost-effective, and protective of human health and the environment. The ProUCL software is based upon the philosophy that rigorous statistical methods can be used to compute correct estimates of population parameters and decision making statistics including: the upper confidence limit (UCL) of the mean, the upper tolerance limit (UTL), and the upper prediction limit (UPL) to help decision makers and project teams in making correct decisions. A few commonly used text book type methods (e.g., CLT, Student's t-UCL) alone cannot address all scenarios and situations occurring in the various environmental studies. Since many environmental decisions are based upon a 95% UCL (UCL95) of the population mean, it is important to compute correct UCLs of practical merit. The use and applicability of a statistical method (e.g., student's t-UCL, Central Limit Theorem (CLT)-UCL, adjusted gamma-UCL, Chebyshev UCL, bootstrap-t UCL) depend upon data size, data skewness, and data distribution. ProUCL computes decision statistics using several parametric and nonparametric methods covering a wide-range of data variability, distribution, skewness, and sample size. It is anticipated that the availability of the statistical methods in the ProUCL software covering a wide range of environmental data sets will help the decision makers in making more informative and correct decisions at the various Superfund and RCRA sites.

It is noted that for moderately skewed to highly skewed environmental data sets, UCLs based on the CLT and the Student's t-statistic fail to provide the desired coverage (e.g., 0.95) to the population mean even when the sample sizes are as large as 100 or more. The sample size requirements associated with the CLT increases with skewness. It will be naive and incorrect to state that a CLT or Student's statistic based UCLs are adequate to estimate EPC terms based upon skewed data sets. These facts have been described in the published documents summarizing simulation experiments conducted on positively skewed data sets to evaluate the performances of the various UCL computation methods. The use of a parametric lognormal distribution on a lognormally distributed data set yields unstable impractically large UCLs values, especially when the standard deviation (*sd*) of the log-transformed data becomes greater than 1.0 and the data set is of small size less than 30-50. Many environmental data sets can be modeled by a gamma as well as a lognormal distribution. The use of a gamma distribution on gamma distributed data sets tends to yield UCL values of practical merit. Therefore, the use of gamma distribution based decision statistics such as UCLs, UPLs, and UTLs cannot be dismissed by stating that it is easier (than a gamma model) to use a lognormal model to compute these upper limits.

The suggestions made in ProUCL are based upon the extensive experience of the developers in environmental statistical methods, published environmental literature, and procedures described in various EPA guidance documents. The inclusion of outliers in the computation of the various decision statistics tends to yield inflated values of those decision statistics, which can lead to incorrect decisions. Often inflated statistics computed using a few outliers tend to represent those outliers rather than representing the main dominant population of interest (e.g., reference area). It is suggested to identify outliers, observations coming from population(s) other than the main dominant population, before computing the decision statistics needed to address project objectives. The project team may want to perform the statistical evaluations twice, once with outliers and once without outliers. This exercise will help the project team in computing correct and defensible decision statistics needed to make cleanup and remediation decisions at polluted sites.

The initial development during 1999-2000 and all subsequent upgrades and enhancements of the ProUCL software have been funded by USEPA through its Office of Research and Development (ORD).   Initially

ProUCL was developed as a research tool for USEPA scientists and researchers of the Technical Support Center and ORD-NERL, EPA Las Vegas. Background evaluations, groundwater monitoring, exposure and risk management and cleanup decisions in support of the Comprehensive Environmental Recovery, Compensation, and Liability Act (CERCLA) and Resource Conservation and Recovery Act (RCRA) site projects of USEPA are often derived based upon the various test statistics (e.g., Shapiro-Wilk test, t-test, Wilcoxon-Mann-Whitney (WMW) test, analysis of variance [ANOVA], Mann-Kendall [MK] test) and decision statistics including UCLs of mean, UPLs, and UTLs. To address the statistical needs of the environmental projects of the USEPA, over the years ProUCL software has been upgraded and enhanced to include many graphical tools and statistical methods described in the various EPA guidance documents including: EPA 1989a, 1989b, 1991, 1992a, 1992b, 2000 (MARSSIM), 2002a, 2002b, 2002c, 2006a, 2006b, and 2009. Several statistically rigorous methods (e.g., for data sets with NDs) not easily available in the existing guidance documents and in the environmental literature are also available in ProUCL version 5.0.00 (ProUCL 5.0).

ProUCL 5.0 has graphical, estimation, and hypotheses testing methods for uncensored-full data sets and for left-censored data sets consisting of NDs observations with multiple detection limits (DLs) or reporting limits (RLs). In addition to computing general statistics, ProUCL 5.0 has goodness-of-fit (GOF) tests for normal, lognormal and gamma distributions, parametric and nonparametric methods including bootstrap methods for skewed data sets to compute various decision making statistics such as UCLs of mean (EPA 2002a), percentiles, UPLs for a certain number of future observations (e.g., k with k=1, 2, 3,...), UPLs for mean of future k (≥1) observations, and UTLs (e.g., EPA 1992b, 2002b, and 2009). Many positively skewed environmental data sets can be modeled by a lognormal as well as a gamma model. It is well-known that for moderately skewed to highly skewed data sets, the use of a lognormal distribution tends to yield inflated and unrealistically large values of the decision statistics especially when the sample size is small (e.g., <20-30). For gamma distributed skewed uncensored and left-censored data sets, ProUCL software computes decision statistics including UCLs, percentiles, UPLs for future k (≥1) observations, UTLs, and upper simultaneous limits (USLs).

For data sets with NDs, ProUCL has several estimation methods including the Kaplan-Meier (KM) method, regression on order statistics (ROS) methods and substitution methods (e.g., replacing NDs by DL, DL/2). ProUCL 5.0 can be used to compute upper limits which adjust for data skewness; specifically, for skewed data sets, ProUCL 5.0 computes upper limits using KM estimates in gamma (lognormal) UCL and UTL equations provided the detected observations in the left-censored data set follow a gamma (lognormal) distribution. Some poor performing commonly used and cited methods such as the DL/2 substitution method and H-statistic based UCL computation method have been incorporated in ProUCL for historical reasons, and research and comparison purposes.

The Sample Sizes module of ProUCL can be used to develop data quality objectives (DQOs) based sampling designs and to perform power evaluations needed to address statistical issues associated with the various polluted sites projects. ProUCL provides user friendly options to enter the desired values for the decision parameters such as Type I and Type II error rates, and other DQOs used to determine the minimum sample sizes needed to address project objectives. The Sample Sizes module can compute DQOs based minimum sample sizes needed: to estimate the population mean; to perform single and two-sample hypotheses testing approaches; and in acceptance sampling to accept or reject a batch of discrete items such as a lot of drums consisting of hazardous waste. Both parametric (e.g., t-test) and nonparametric (e.g., Sign test, WMW test, test for proportions) sample size determination methods are available in ProUCL.

ProUCL has exploratory graphical methods for both uncensored data sets and for left-censored data sets consisting of ND observations. Graphical methods in ProUCL include histograms, multiple quantile-quantile (Q-Q) plots, and side-by-side box plots. The use of graphical displays provides additional insight about the information contained in a data set that may not otherwise be revealed by the use of estimates (e.g., 95% upper limits) and test statistics (e.g., two-sample t-test, WMW test). In addition to providing information about the data distributions (e.g., normal or gamma), Q-Q plots are also useful in identifying outliers and the presence of mixture populations (e.g., data from several populations) potentially present in a data set. Side-by-side box plots and multiple Q-Q plots are useful to visually compare two or more data sets, such as: site-versus-background constituent concentrations, surface-versus-subsurface concentrations, and constituent concentrations of several groundwater monitoring wells (MWs). ProUCL also has a couple of classical outlier test procedures, such as the Dixon test and the Rosner test which can be used on uncensored data sets as well as on left-censored data sets consisting of ND observations.

ProUCL has parametric and nonparametric single-sample and two-sample hypotheses testing approaches for uncensored as well as left-censored data sets. Single-sample hypotheses tests: Student's t-test, Sign test, Wilcoxon Signed Rank test, and the Proportion test are used to compare site mean/median concentrations (or some other threshold such as an upper percentile) with some average cleanup standard, C_s (or a not-to-exceed compliance limit, A_0) to verify the attainment of cleanup levels (EPA, 1989a; MARSSIM, 2000; EPA 2006a) at remediated site areas of concern. Single-sample tests such as the Sign test and Proportion test, and upper limits including UTLs and UPLs are also used to perform intra-well comparisons. Several two-sample hypotheses tests as described in EPA guidance documents (e.g., EPA 2002b, 2006b, 2009) are also available in the ProUCL software. The two-sample hypotheses testing approaches in ProUCL include: Student's t-test, WMW test, Gehan test and Tarone-Ware test. The two-sample tests are used to compare concentrations of two populations such as site versus background, surface versus subsurface soils, and upgradient versus downgradient wells.

The Oneway Analysis of Variance (ANOVA) module in ProUCL has both classical and nonparametric Kruskal-Wallis (K-W) tests. Oneway ANOVA is used to compare means (or medians) of multiple groups such as comparing mean concentrations of several areas of concern and to perform inter-well comparisons. In groundwater (GW) monitoring applications, the ordinary least squares (OLS) of regression, trend tests, and time series plots are used to identify upwards or downwards trends potentially present in constituent concentrations identified in GW monitoring wells over a certain period of time. The Trend Analysis module performs Mann-Kendall trend test and Theil-Sen trend test on data sets with missing values; and generates trend graphs displaying a parametric OLS regression line and nonparametric Theil-Sen trend line. The Time Series Plots option can be used to compare multiple time-series data sets.

The use of the incremental sampling methodology (ISM) has been recommended (ITRC, 2012) to collect ISM soil samples needed to compute mean concentrations of the decision units (DUs) and sampling units (SUs) requiring characterization and remediation activities. At many polluted sites, a large amount of discrete onsite and/or offsite background data are already available which cannot be directly compared with newly collected ISM data. In order to provide a tool to compare the existing discrete background data with actual field onsite or background ISM data, a Monte Carlo Background Incremental Sample Simulator (BISS) module has been incorporated in ProUCL 5.0 (blocked for general public use) which may be used on a large existing discrete background data set. The BISS module simulates incremental sampling methodology based equivalent background incremental samples. The availability of a large discrete background data set collected from areas with geological conditions comparable to the DU(s) of interest is a pre-requisite for successful application of this module. The BISS module has been temporarily blocked for use in ProUCL 5.0 as this module is awaiting adequate instructions and guidance for its intended use on discrete background data sets.

ProUCL 5.0 is a user friendly freeware package providing statistical and graphical tools needed to address statistical issues described in the various EPA guidance documents. ProUCL 5.0 can process many constituents (variables) simultaneously to: perform various tests (e.g., ANOVA and trend test statistics) and compute decision statistics including UCLs of mean, UPLs, and UTLs – a capability not available in several commercial software packages such as Minitab 16 and NADA for R (Helsel, 2013). ProUCL 5.0 also has the capability of processing data by group variables. ProUCL 5.0 is easy to use and it does not require any programming skills as needed when using other software packages such as Minitab, SAS, and programs written in R script.

Methods incorporated in ProUCL 5.0 have been tested and verified extensively by the developers and the various researchers, scientists, and users. The results obtained by ProUCL are in agreement with the results obtained by using other software packages including Minitab, SAS, and programs written in R Script. ProUCL 5.0 computes decision statistics (e.g., UPL, UTL) based upon the KM method in a straight forward manner without flipping the data and re-flipping the computed statistics for left-censored data sets; these operations are not easy for a typical user to understand and perform. This can unnecessarily become tedious when computing decision statistics for multiple variables/analytes. Moreover, unlike survival analysis, it is important to compute an accurate estimate of the *sd* which is needed to compute decision making statistics including UPLs and UTLs. For left-censored data sets, ProUCL computes a KM estimate of *sd* directly. These issues are elaborated by examples discussed in this User Guide and in the accompanying ProUCL 5.0 Technical Guide.

Table of Contents

ProUCL 5.0.00

Software ProUCL version 5.0.00 (ProUCL 5.0), its earlier versions: ProUCL version 3.00.01, 4.00.02, 4.00.04, 4.00.05, 4.1.00, and 4.1.01, associated Facts Sheet, User Guides and Technical Guides (e.g., EPA 2010a, 2010b) can be downloaded from the following EPA website:

http://www.epa.gov/osp/hstl/tsc/software.htm
http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

Material for a couple of ProUCL webinars offered in March 2011, and relevant literature used in the development of ProUCL 5.0 can also be downloaded from the above EPA website.

Contact Information for all Versions of ProUCL

The ProUCL software is 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 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

ACRONYMS and ABBREVIATIONS

ACL	alternative compliance or concentration limit
A-D, AD	Anderson-Darling test
AM	arithmetic mean
AOC	area(s) of concern
ANOVA	analysis of variance
A_0	not to exceed compliance limit or specified action level
BCA	bias-corrected accelerated bootstrap method
BIS	Background Incremental Sample
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
COPCs	Contaminants/constituents of potential concern
C_s	cleanup standards
CSM	conceptual site model
CV	coefficient of variation
DL, L	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 half 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
GB	Gigabyte
GHz	Gigahertz
GROS	gamma ROS
GOF, G.O.F.	goodness-of-fit
GOF Q-Q Plot	Quantile-Quantile Plot showing GOF statistics
GUI	graphical user interface
H-UCL	UCL based upon Land's H-statistic
H_A	alternative hypothesis
H_0	null hypothesis
i.i.d.	independently identically distributed
ITRC	Interstate Technology & Regulatory Council
k, K	a positive integer representing future or next k observations
k	number of non-detects in a sample
K	shape parameter of a gamma distribution
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
LPL	lower prediction limit
LROS	logROS; robust ROS
LTL	lower tolerance limit

LSL	lower simultaneous limit
M, *m*	applied to incremental sampling: number of increments in a BISS sample
MAD	median absolute deviation
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
MLE (t)	UCL based upon ML estimates using Student's t-distribution critical value
Multiple Q-Q	multiple quantile-quantile plot
MVUE	minimum variance unbiased estimate
MW	monitoring well
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
PCA	principal component analysis
PDF, pdf	probability density function
Pdf	files in pdf format
PRG	preliminary remediation goals
Q-Q	quantile-quantile
R	applied to incremental sampling: number of replicate ISM
RAGS	Risk Assessment Guidance for Superfund
RCRA	Resource Conservation and Recovery Act
RL	reporting limit
ROS	regression on order statistics
RPM	Remedial Project Manager

RSD	relative standard deviation
S	substantial difference
SCMTSC	Site Characterization and Monitoring Technical Support Center
SD, *Sd*, *sd*	standard deviation
SE	standard error
s_p	pooled standard deviation
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, USEPA	United States Environmental Protection Agency
UTL	upper tolerance limit
USGS	U.S. Geological Survey
USL	upper simultaneous limit
WMW	Wilcoxon-Mann-Whitney
WRS	Wilcoxon Rank Sum
WSR	Wilcoxon Signed Rank
$<$	less than
$>$	Greater than
$\geq$	greater than or equal to
$\leq$	less than or equal to
X_p	p^{th} percentile of a distribution
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
%	represents the percentage symbol
α	Type I error rate

β	Type II error rate
σ	standard deviation of a log-transformed sample
Θ	scale parameter of a gamma distribution

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 EPA scientists including Deana Crumbling, Nancy Rios-Jafolla, Tim Frederick, Dr. Maliha Nash, Kira Lynch, and Marc Stiffleman; James Durant of ATSDR, Dr. Steve Roberts of University of Florida, Dr. Elise A. Striz of NRC, and Drs. Phillip Goodrum and John Samuelian of Integral Consulting Inc. for testing and reviewing ProUCL 5.0 and its associated guidance documents, and for providing helpful comments and suggestions.

Special thanks go to Ms D. Getty and Mr. R. Leuser of Lockheed Martin for providing a thorough technical and editorial review of 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 ProUCL5.0 software and associated guidance documents would not have been possible.

Finally, we wish to dedicate the 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.

Introduction

Overview of ProUCL Version 5.0.00 Software

The main objective of the ProUCL software funded by the USEPA is to compute rigorous decision statistics to help the decision makers in making correct decisions which are cost-effective, and protective of human health and the environment. The ProUCL software is based upon the philosophy that rigorous statistical methods can be used to compute the correct estimates of population parameters (e.g., site mean, background percentiles) and decision making statistics including the upper confidence limit of (UCL) the mean, the upper tolerance limit (UTL), and the upper prediction limit (UPL) to help the decision makers and project teams in making correct decisions. The use and applicability of a statistical method (e.g., student's t-UCL, Central Limit Theorem (CLT)-UCL, adjusted gamma-UCL, Chebyshev UCL, bootstrap-t UCL) depend upon data size, data variability, data skewness, and data distribution. ProUCL computes decision statistics using several parametric and nonparametric methods covering a wide-range of data variability, skewness, and sample size. A couple of text book methods described in most of the statistical text books (e.g., Hogg and Craig, 1995) based upon the Student's t-statistic and the CLT alone cannot address all scenarios and situations commonly occurring in the various environmental studies. It is naive and incorrect to state or assume that Student's t-statistic and/or CLT based UCLs of mean will provide the desired coverage (e.g., 0.95) to the population mean irrespective of the skewness of the data set/population under consideration. These issues have been discussed in detail in Chapters 2 and 4 of the ProUCL 5.0 Technical guide. Several examples have been discussed throughout this guidance document and also in the accompanying ProUCL 5.0 Technical Guide to elaborate on these issues.

The use of a parametric lognormal distribution on a lognormally distributed data set tends to yield unstable impractically large UCLs values, especially when the standard deviation of the log-transformed data is greater than 1.0 and the data set is of small size such as less than 30-50 (Hardin and Gilbert, 1993; Singh, Singh, and Engelhardt, 1997). Many environmental data sets can be modeled by a gamma as well as a lognormal distribution. Generally, the use of a gamma distribution on gamma distributed data sets yields UCL values of practical merit (Singh, Singh, and Iaci, 2002). Therefore, the use of gamma distribution based decision statistics such as UCLs, UPL, and UTLs cannot be dismissed just because it is easier to use a lognormal model to compute these upper limits or incorrectly assuming that the two distributions behave in a similar manner. The advantages of computing the gamma distribution based decision statistics are discussed in Chapters 2-5 of the ProUCL 5.0 Technical guidance document.

Since many environmental decisions are made based upon a 95% UCL of the population mean, it is important to compute correct UCLs and other decision making statistics of practical merit. In an effort to compute correct UCLs of the population mean and other decision making statistics, in addition to computing the Student's t statistic and the CLT based statistics (e.g., UCLs, UPLs), significant effort has been made to incorporate rigorous statistical methods based UCLs (and other limits) in the ProUCL software covering a wide-range of data skewness and sample sizes (e.g., Singh, Singh, and Engelhardt, 1997; Singh, Singh, and Iaci, 2002; and Singh, Singh, 2003). It is anticipated that the availability of the statistical methods in the ProUCL software covering a wide range of environmental data sets will help the decision makers in making more informative and correct decisions at the various polluted sites.

It is noted that even for skewed data sets, practitioners tend to use the CLT or Student's t-statistic based UCLs of mean based upon samples of sizes 25-30 (large sample rule-of-thumb to use CLT). However, this rule-of-thumb does not apply to moderately skewed to highly skewed data sets, specifically when σ

(standard deviation of the log-transformed data) starts exceeding 1. The large sample requirement associated with the use of the CLT depends upon the skewness of the data distribution under consideration. The large sample requirement for the sample mean to follow an approximate normal distribution increases with the data skewness; and for skewed data sets, even samples of size greater than (>)100 may not be large enough for the sample mean to follow an approximate normal distribution. For moderately skewed to highly skewed environmental data sets, as expected, UCLs based on the CLT and the Student's t- statistic fail to provide the desired coverage to the population mean even when the sample sizes are as large as 100 or more. These facts have been verified in the published simulation experiments conducted on positively skewed data sets (e.g., Singh, Singh, and Engelhardt, 1997; Singh, Singh, and Iaci, 2002; and Singh and Singh, 2003).

The initial development and all subsequent upgrades and enhancement of the ProUCL software have been funded by the USEPA through its Office of Research and Development (ORD). Initially ProUCL was developed as a research tool for scientists and researchers of the Technical Support Center and ORD-NERL, EPA Las Vegas. During 1999-2001, the initial intent and objectives of developing the ProUCL software (Version 1.0 and Version 2.0) were to provide a statistical research tool to EPA scientists which can be used to compute theoretically sound 95% upper confidence limits (UCL95s) of the mean routinely used in exposure assessment, risk management and cleanup decisions made at various CERCLA and RCRA sites (EPA 1992a, 2002a). During 2002, the peer-reviewed ProUCL version 2.1 (with Chebyshev inequality based UCLs) was released for public use. Several researchers have developed rigorous parametric and nonparametric statistical methods (e.g., Johnson, 1978; Grice and Bain, 1980; Efron (1981, 1982); Efron and Tibshirani, 1993; Hall (1988, 1992); Sutton, 1993; Chen, 1995; Singh, Singh, and Engelhardt, 1997; Singh, Singh, and Iaci, 2002) to compute upper limits (e.g., UCLs) which adjust for data skewness. Since Student's t-UCL, CLT-UCL, and percentile bootstrap UCL fail to provide the desired coverage to the population mean of skewed distributions, several parametric (e.g., gamma distribution based) and nonparametric (e.g., BCA bootstrap and bootstrap-t, Chebyshev UCL) UCL computation methods which adjust for data skewness were incorporated in ProUCL versions 3.0 and 3.00.02 during 2003-2004. ProUCL version 3.00.02 also had graphical quantile-quantile (Q-Q) plots and GOF tests for normal, lognormal, and gamma distributions; capabilities to statistically analyze multiple variables simultaneously were also incorporated in ProUCL 3.00.02 (EPA 2004).

It is important to compute decision statistics (e.g., UCLs, UTLs) which are cost-effective and protective of human health and the environment (balancing between Type I and Type II errors), therefore, one cannot dismiss the use of the better [better than t-UCL, CLT-UCL, ROS and KM percentile bootstrap UCL, KM-UCL (t)] performing UCL computation methods including gamma UCLs and the various bootstrap UCLs which adjust for data skewness. During 2004-2007, ProUCL was upgraded to versions 4.00.02, and 4.00.04. These upgrades included exploratory graphical (e.g., Q-Q plots, box plots) and statistical (e.g., maximum likelihood estimation [MLE], KM, and ROS) methods for left-censored data sets consisting of nondetect (NDs) observations with multiple DLs or RLs. For uncensored and left-censored data sets, these upgrades provide statistical methods to compute upper limits: percentiles, UPLs and UTLs needed to estimate site-specific background level constituent concentrations or background threshold values (BTVs). To address statistical needs of background evaluation projects (e.g., MARSSIM, 2000; EPA 2002b), several single-sample and two-sample hypotheses testing approaches were also included in these ProUCL upgrades.

During 2008-2010, ProUCL was upgraded to ProUCL 4.00.05. The upgraded ProUCL was enhanced by including methods to compute gamma distribution based UPLs and UTLs (Krishnamoorthy, Mathew, and Mukherjee, 2008). The Sample Size module to compute DQOs based minimum sample sizes needed to address statistical issues associated with the various environmental projects (e.g., MARSSIM, 2000; EPA [2002c, 2006a, 2006b]) was also incorporated in ProUCL 4.00.05.

During 2009-2011, ProUCL 4.00.05 was upgraded to ProUCL 4.1 and 4.1.01. ProUCL 4.1 (2010) and 4.1.01 (2011) retain all capabilities of the previous versions of ProUCL software. Two new modules: Oneway ANOVA and Trend Analysis were included in ProUCL 4.1. The Oneway ANOVA module has both parametric and nonparametric ANOVA tests to perform inter-well comparisons. The Trend Analysis module can be used to determine potential upward or downward trends present in constituent concentrations identified in GW monitoring wells (MWs). The Trend Analysis module can compute Mann-Kendall (MK) and Theil-Sen (T-S) trend statistics to determine upward or downward trends potentially present in analyte concentrations. ProUCL 4.1 also has the Ordinary Least Squares (OLS) Regression module. In ProUCL 4.1, some modifications were made in decision tables used to make recommendations regarding the use of UCL95 to estimate EPC terms. Specifically, based upon the recent experience, developers of ProUCL re-iterated that the use of a lognormal distribution to estimate EPC terms and BTVs should be avoided, as the use of lognormal distribution tends to yield unrealistic and unstable values of the decision making statistics including UCL, UPL, and UTL; this is especially true when the sample size is <20-30 and the data set is moderately skewed to highly skewed. During March 2011, a couple of webinars were presented describing the capabilities and use of the methods available in ProUCL 4.1.

ProUCL version 5.0.00 represents an upgrade of ProUCL 4.1.01 (EPA, June 2011) which represents an upgrade of ProUCL 4.1.00 (EPA 2010). For uncensored and left-censored data sets, ProUCL 5.0 consists of all statistical and graphical methods that are available in the previous versions of the ProUCL software package except for a couple of poor performing and restricted (e.g., can be used only when a single detection limit is present) estimation methods such as the MLE and winsorization methods for left-censored data sets. ProUCL has GOF tests for normal, lognormal, and gamma distributions for uncensored and left-censored data sets with NDs. ProUCL 5.0 has the extended version of the Shapiro-Wilk (S-W) test to perform normal and lognormal GOF tests for data sets of sizes up to 2000 (Royston [1982, 1982a]). In addition to normal and lognormal distribution based decision statistics, ProUCL software computes UCLs, UPLs, and UTLs based upon the gamma distribution.

Several enhancements have been made in the UCLs and BTVs modules of the ProUCL 5.0 software. A new statistic, an upper simultaneous limit (Singh and Nocerino, 2002; Wilks, 1963) has been incorporated in the Upper limits/BTVs module of ProUCL 5.0.00 for data sets consisting of NDs with multiple DLs, a two-sample hypothesis test, the Tarone-Ware (T-W; Tarone and Ware, 1978) test has been incorporated in ProUCL 5.0. Nonparametric tolerance limits have been enhanced, and for specific values of confidence coefficients, coverage probability, and sample size, ProUCL 5.0 outputs the confidence coefficient actually achieved by a UTL. The Trend Analysis and OLS Regression modules can handle missing events to compute trend test statistics and generate trend graphs. Some new methods using KM estimates in gamma (and lognormal) distribution based UCL, UPL, and UTL equations have been incorporated to compute the decision statistics for data sets consisting of nondetect observations. To facilitate the computation of UCLs from ISM based samples (ITRC, 2012); the minimum sample size requirement has been lowered to 3, so that one can compute the UCL95 based upon ISM data sets of sizes ≥3. To select an appropriate UCL95 of mean for ISM data set, the user should consult the ITRC (2012) Tech Reg Guide on Incremental Sampling Methodology.

All known bugs, typographical errors, and discrepancies found by the developers and the various users of the ProUCL software package have been addressed in the ProUCL version 5.0.00. Specifically, a discrepancy found in the estimate of mean based upon the KM method has been fixed in ProUCL 5.0.

Some changes have been made in the decision logic used in GOF and UCL modules. In practice, based upon a given data set, it is well known that the two statistical tests (e.g., Theil-Sen and OLS trend tests) can lead to different conclusions. To streamline the decision logic associated with the computation of the various UCLs, the decision tables in ProUCL 5.0 have been updated. Specifically, for each distribution if at least one of the two GOF tests (e.g., Shapiro-Wilk or Lilliefors test for normality) determines that the hypothesized distribution holds, then ProUCL concludes that the data set follows the hypothesized distribution, and decision statistics are computed accordingly. Additionally, for gamma distributed data sets, ProUCL 5.0 suggests the use of the: adjusted gamma UCL for samples of sizes ≤ 50 (instead of 40 suggested in previous versions); and approximate gamma UCL for samples of sizes >50.

Also, for samples of larger sizes (e.g., with $n > 100$) and small values of the gamma shape parameter, k (e.g., $k \leq 0.1$), significant discrepancies were found in the critical values of the two gamma GOF test statistics (Anderson-Darling and Kolmogorov Smirnov tests) obtained using the two gamma deviate generation algorithms: Whitaker (1974) and Marsaglia and Tsang (2000). For values of $k \leq 0.2$, the critical values of the two gamma GOF tests: Anderson-Darling (A-D) and Kolmogorov-Smirnov (K-S) tests have been updated using the currently available more accurate gamma deviate generation algorithm due to Marsaglia and Tsang's (2000); more details about the implementation of their algorithm can be found in Kroese, Taimre, and Botev (2011). For values of the shape parameter, $k=0.025$, 0.05, 0.1, and 0.2, the critical value tables for these two tests have been updated by incorporating the newly generated critical values for the three significance levels: 0.05, 0.1, and 0.01. The updated tables are provided in Appendix A. It should be noted that for $k=0.2$, the older and the newly generated critical values are in general agreement.

ProUCL 5.0 also has a new Background Incremental Sample Simulator (BISS) module (temporarily not available for general use) which can be used on a large existing discrete background data set to simulate background incremental samples (BIS). The availability of a large discrete data set collected from areas with geological formations and conditions comparable to the DUs (background or onsite) of interest is a requirement for successful application of this module. The simulated BISS data can be compared with the actual field ISM (ITRC, 2012) data collected from the various DUs using other modules of ProUCL 5.0. The values of the BISS data are not directly available to users; however, the simulated BISS data can be accessed by the various modules of ProUCL 5.0 to perform desired statistical evaluations. For example, the simulated background BISS data can be merged with the actual field ISM data after comparing the two data sets using a two-sample t-test; the simulated BISS or the merged data can be used to compute a UCL of the mean or a UTL.

Note: The ISM methodology used to develop the BISS module is a relatively new approach; methods incorporated in this BISS module require further investigation. The BISS module has been temporarily blocked for use in ProUCL 5.0 as this module is awaiting adequate guidance for its intended use on discrete background data sets.

Software ProUCL version 5.0, its earlier versions: ProUCL version 3.00.02, 4.00.02, 4.00.04, 4.1.00, and 4.1.01, associated Facts Sheet, User Guides and Technical Guides (e.g., EPA [2004, 2007, 2009a, 2009b, 2010a, 2010b]) can be downloaded from the EPA website:

http://www.epa.gov/osp/hstl/tsc/software.htm
http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

ProUCL 5.0 is a user-friendly freeware package providing statistical and graphical tools needed to address statistical issues described in several EPA guidance documents. Considerable effort has been made to provide a detailed technical guide to help practitioners understand statistical methods needed to address statistical needs of their environmental projects. ProUCL generates detailed output sheets and graphical displays for each method which can be used to educate students learning environmental statistical methods. Like previous versions, ProUCL 5.0 can process many variables simultaneously to compute various tests (e.g., ANOVA and trend test statistics) and decision statistics including UCL of mean, UPLs, and UTLs, a capability not available in other software packages such as Minitab 16 and NADA for R (Helsel, 2013). Without the availability of this option, the user has to compute decision and test statistics for one variable at a time which becomes cumbersome when dealing with a large number of variables. ProUCL 5.0 also has the capability of processing data by groups. ProUCL 5.0 is easy to use; it does not require any programming skills as needed when using programs written in R Script.

The Need for ProUCL Software

EPA guidance documents (e.g., EPA [1989a, 1989b, 1992a, 1992b, 1994, 1996, 2000, 2002a, 2002b, 2002c, 2006a, 2006b, 2009a, and 2009b]) describe statistical methods including: DQOs based sample size determination procedures, methods to compute decision statistics: UCL95, UPL, and UTLs, parametric and nonparametric hypotheses testing approaches, Oneway ANOVA, OLS regression, and trend determination approaches. Specifically, EPA guidance documents (e.g., EPA [2002c, 2006a, 2006b; and MARSSIM, 2000]) describe DQOs based parametric and nonparametric minimum sample size determination procedures needed: to compute decision statistics (e.g., UCL95); to perform site versus background comparisons (e.g., t-test, proportion test, WMW test); and to determine the number of discrete items (e.g., drums filled with hazardous material) that need to be sampled to meet the DQOs (e.g., specified proportion, p_0 of defective items, allowable error margin in an estimate of mean). Statistical methods are used to compute test statistics (e.g., S-W test, t-test, WMW test, T-S trend statistic) and decision statistics (e.g., 95% UCL, 95% UPL, UTL95-95) needed to address statistical issues associated with CERCLA and RCRA site projects. For example, exposure and risk management and cleanup decisions in support of EPA projects are often made based upon the mean concentrations of the contaminants/constituents of potential concern (COPCs). Site-specific BTVs are used in site versus background evaluation studies. A UCL95 is used to estimate the EPC terms (EPA1992a, 2002a); and upper limits such as upper percentiles, UPLs, or UTLs are used to estimate BTVs or not-to-exceed values (EPA 1992b, 2002b, and 2009). The estimated BTVs are also used: to identify the COPCs; to identify the site areas of concern (AOCs); to perform intra-well comparisons to identify MWs not meeting specified standards; and to compare onsite constituent concentrations with site-specific background level constituent concentrations. Oneway ANOVA is used to perform inter-well comparisons, OLS regression and trend tests are often used to determine potential trends present in constituent concentrations identified in groundwater monitoring wells (MWs). Most of the methods described in this paragraph are available in the ProUCL 5.0 software package.

It is noted that not much guidance is available in the guidance documents cited above to compute rigorous UCLs, UPLs, and UTLs for moderately skewed to highly skewed uncensored and left-censored data sets consisting of NDs with multiple DLs, a common occurrence in environmental data sets. Several parametric and nonparametric methods are available in the statistical literature (Singh, Singh, and Engelhardt, 1997; Singh, Singh, and Iaci, 2002; Krishnamoorthy et al. 2008; Singh, Maichle, and Lee, 2006) to compute UCLs and other upper limits which adjust for data skewness. During the years, as new methods became available to address statistical issues related to the environmental projects, those methods were incorporated in ProUCL software so that environmental scientists and decision makers can make more accurate and informative decisions based upon those rigorous statistical methods. Until 2006, not much guidance was provided on how to compute UCL95 of mean and other upper limits (e.g., UPLs and UTLs) based upon data sets consisting of NDs with multiple DLs. For data sets with NDs, Singh, Maichle, and Lee (EPA 2006) conducted an extensive simulation study to compare the performances of the various estimation methods (in terms of bias in the mean estimate) and UCL computation methods (in terms of coverage provided by a UCL). They demonstrated that the nonparametric KM method performs well in terms of bias in estimates of mean. They also concluded that UCLs computed using the Student's t-statistic and percentile bootstrap method using the KM estimates do not provide the desired coverage to the population mean of skewed data sets. They demonstrated that the depending upon sample size and data skewness, UCLs computed using KM estimates and: the BCA bootstrap method (mildly skewed data sets); the bootstrap-t method, and the Chebyshev inequality (moderately to highly skewed data sets) provide better coverage (closer to the specified 95% coverage) to the population mean than the various other UCL computation methods. Based upon their findings, during 2006-2007, several UCL and other upper limits computation methods based upon KM and ROS estimates were incorporated in the ProUCL 4.0 software. It is noted that since the inclusion of the KM method in ProUCL 4.0 (2007), the use of the KM method based upper limits has become popular in many environmental applications to estimate EPC terms and background threshold values (BTVs). The KM method is also described in the latest version of the unified RCRA guidance document (EPA 2009).

It is not easy to justify distributional assumptions of data sets consisting of both detects and NDs with multiple DLs. Therefore, based upon the published literature and recent experience, parametric UCL computation methods such as the MLE methods for normal and lognormal distributions are excluded from ProUCL 5.0. Additionally, the winsorization method (Gilbert, 1987) has also been excluded from ProUCL 5.0 due to its poor performance. ProUCL software is also used for teaching environmental statistics courses therefore, in addition to statistical and graphical methods routinely used to address statistical needs of environmental projects, due to their popularity some poor performing methods such as the substitution DL/2 method and Land's (1975) H-statistic based UCL computation method have been retained in ProUCL version 5.0.00 for research and comparison purposes.

Methods incorporated in ProUCL 5.0 and in its earlier versions have been tested and verified extensively by the developers and various researchers, scientists, and users. Specifically, the results obtained by ProUCL 5.0 are in agreement with the results obtained by using other software packages including Minitab, SAS, and programs available in R-Script (not all methods are available in these software packages). Additionally, ProUCL 5.0 outputs several intermediate results (e.g., khat and biased corrected kstar estimates of the gamma shape parameter, k) and critical values (e.g., K factor used to compute UTLs, d2max needed to compute USL) needed to compute the various decision statistics of interest, which may help the interested users to verify statistical results computed by the ProUCL software. ProUCL is a user friendly software which can be used to: process multiple variables (analytes) simultaneously (e.g., perform ANOVA on many variables); process grouped data; to generate and display multiple plots (Q-Q plots) on the same graphical display. No programming skills are needed to use ProUCL software. ProUCL provides warning messages and makes suggestions to help a typical user in selecting the most appropriate decision statistic (e.g., UCL).

Note: The availability of intermediate results and critical values can be used to compute lower limits and two-sided intervals which are not as yet available in the ProUCL software.

For left-censored data sets, ProUCL 5.0 computes decision statistics (e.g., UCL, UPL, and UTL) based upon KM estimates computed in a straight forward manner without flipping the data and re-flipping the decision statistics; these operations are not easy for a typical user to understand and perform and can become quite tedious when multiple analytes need to be processed. Moreover, in environmental applications it is important to compute accurate estimates of standard deviations which are needed to compute the decision making statistics including UPLs and UTLs. Decision statistics (UPL, UTL) based upon a KM estimate of the of standard deviation computed using indirect methods can be different from the statistics computed using an estimate of *sd* obtained using the KM method directly, especially when one is dealing with skewed data set or using a log-transformation. These issues are elaborated by examples discussed in this Guide and the accompanying ProUCL 5.0 Tech Guide.

For uncensored data sets, researchers (e.g., Johnson (1978), Chen (1995), Efron and Tibshirani (1993), Hall [1988, 1992], more references in Chapters 2 and 3) had developed parametric (e.g., gamma distribution based) and nonparametric (bootstrap-t and Hall's bootstrap method, modified-t) methods to compute decision statistics which adjust for data skewness. For uncensored positively skewed data sets, Singh, Singh, and Iaci (2002) and Singh and Singh (2003) performed simulation experiments to compare the performances (in terms of coverage probabilities) of the various UCL computation methods described in the literature. They demonstrated that for skewed data sets, UCLs based upon Student's t statistic, central limit theorem (CLT), and percentile bootstrap method tend to underestimate the population mean (EPC term). It is reasonable to state and assume the findings of the simulation studies performed on uncensored skewed data sets to compare the performances of the various UCL computation methods can be extended to skewed left-censored data sets. Based upon the findings of those studies performed on uncensored data sets and also using the findings summarized in Singh, Maichle, and Lee (2006), it is concluded that t-statistic, CLT, and the percentile bootstrap method based UCLs computed using KM estimates (and also ROS estimates) underestimate the population mean of moderately skewed to highly skewed data sets. Interested users may want to verify these statements via simulation experiments or otherwise. Like uncensored skewed data sets, for left-censored data sets, ProUCL 5.0 offers several parametric and nonparametric methods to compute UCLs and other limits which adjust for data skewness.

In earlier versions of the ProUCL software (e.g., ProUCL 4.00.02), for left-censored data sets, KM estimates were used in the normal distribution based equations to compute the various upper limits. However, normal distribution based upper limits (e.g., t-UCL) using KM estimates (or any other estimates such as ROS estimates) fail to provide the specified coverage to the parameters (e.g., mean, percentiles) of populations with skewed distributions (Singh, Singh, and Iaci, 2002, Johnson, 1978, Chen 1995). Also, the nonparametric UCL computation methods (e.g., percentile bootstrap) do not provide the desired coverage to the population means of skewed distributions (e.g., Hall [1988, 1992], Efron and Tibshirani, 1993). For an example, the use of t-UCL or the percentile bootstrap UCL method on robust ROS estimates or on KM estimates underestimates the population mean for moderately skewed to highly skewed data sets. Chapters 3 and 5 of the ProUCL 5.0 Tech Guide describe parametric and nonparametric KM method based upper limits computation methods (and available in ProUCL 5.0) which adjust for data skewness.

The KM method yields good estimates of the population mean and standard deviation (Singh, Maichle, and Lee, 2006); however upper limits computed using the KM or ROS estimates in normal equations or in the percentile bootstrap method do not account for skewness present in the data set. Appropriate UCL computation methods which account for data skewness should be used on KM or ROS estimates. For left-censored data sets, ProUCL 5.0 computes upper limits using KM estimates in gamma (lognormal) UCL, UPL, and UTL equations (e.g., also suggested in EPA 2009) provided the detected observations in the left-censored data set follow a gamma (lognormal) distribution.

Recently, the use of the ISM methodology has been recommended (ISM ITRC, 2012) to collect soil samples needed to estimate mean concentrations of the DUs requiring characterization and remediation activities. ProUCL can be used to compute UCLs based upon ISM data as described and recommended in the ITRC ISM Tech Reg Guide (2012). At many sites, a large amount of discrete background data is already available which are not directly comparable to the actual field ISM data (onsite or background). To compare the existing discrete background data with field ISM data, the BISS module of ProUCL 5.0 (blocked for general use in ProUCL version 5.0 and is awaiting instructions and guidance for its intended use) can be used on a large (e.g., consisting of at least 30 observations) existing discrete background data set. The BISS module simulates incremental sampling methodology based equivalent incremental background samples; and each simulated BISS sample represents an estimate of the mean of the population represented by the discrete background data set. The availability of a large discrete background data set collected from areas with geological conditions comparable to the DU(s) of interest (onsite DUs) is a requirement for successful application of this module. The user cannot see the simulated BISS data; however the simulated BISS data can be accessed by the various other modules of ProUCL 5.0 to perform desired statistical evaluations. For example, the simulated BISS data can be merged with the actual field ISM data (e.g., field background ISM data) after comparing the two data sets using a two-sample t-test. The actual field ISM or the merged ISM and BISS data can be accessed by the various modules of ProUCL to compute a UCL of mean or a UTL.

ProUCL 5.0 Capabilities

A summary of statistical methods available in the ProUCL software is provided as follows.

Assumptions: Like most statistical methods, statistical methods to compute upper limits (e.g., UCLs, UPLs, UTLs) are also based upon certain assumptions including the availability of a randomly collected data set consisting of independently and identically distributed (*i.i.d*) observations representing the population (e.g., site area, reference area) under investigation. A UCL of the mean (of a population) and BTV estimates (UPL, UTL) should be computed using a randomly collected (simple random or systematic random) data set representing a single statistical population (e.g., site population or background population). If multiple populations (e.g., background and site data mixed together) are present in a data set, it is recommended to separate them out first by using the population partitioning techniques (e.g., Singh, Singh, and Flatman 1994), and then compute appropriate decision statistics (e.g., 95% UCLs) separately for each identified population. The topic of population partitioning and the extraction of a valid site-specific background data set from a broader mixture data set potentially consisting of both onsite and offsite data are beyond the scope of ProUCL 5.0. Parametric estimation and hypotheses testing methods (e.g., t-test, UCLs, UTLs) are based upon distributional (e.g., normal distribution, gamma) assumptions. ProUCL has GOF tests for normal, gamma, and lognormal distributions.

Multiple Constituents/Variables: Environmental scientists need to evaluate many constituents in their decision making processes (exposure and risk assessment). ProUCL can process multiple constituents/variables simultaneously in a user friendly manner, an option not available in other freeware or commercial software packages such as NADA for R (Helsel, 2013). This option is very useful when one has to process many variables/analytes and compute decision statistics (e.g., UCLs, UPLs, and UTLs) and test statistics (e.g., ANOVA test, trend test) for those variables/analytes.

Analysis by a Group Variable: ProUCL also has the capability of processing data by groups. A valid group column should be included in the data file. The analyses of data categorized by a group ID variable such as: 1) Surface vs. Subsurface; 2) AOC1 vs. AOC2; 3) Site vs. Background; and 4) Upgradient vs. Downgradient MWs are common in many environmental applications. ProUCL offers this option for data sets with and without nondetects. The Group Option provides a useful option to perform various statistical tests and methods including graphical displays separately for each of the group (samples from different populations) that may be present in a data set. For an example, the same data set may consist of analytical data from the various groups or populations representing site, background, two or more AOCs, surface, subsurface, monitoring wells. By using this option, the graphical displays (e.g., box plots, Q-Q plots, histograms) and statistics including computation of background statistics, UCLs, ANOVA test, trend test and OLS regression statistics can be easily computed separately for each group in the data set.

Exploratory Graphical Displays for Uncensored and Left-Censored Data Sets: Graphical methods included in the Graph module of ProUCL include: Q-Q plots (data in same column), multiple Q-Q plots (data in different columns), box plots, multiple box plots, and histograms. These graphs can also be generated for data sets consisting of ND observations. Additionally, the OLS Regression and Trend Analysis module can be used to generate graphs displaying parametric OLS regression lines with confidence intervals and prediction intervals around the regression lines and nonparametric Theil-Sen trend lines. The Trend Analysis module can generate trend graphs for data sets without a sampling event variable, and also generate time series graphs for data sets with a sampling event (time) variable. ProUCL 5.0 accepts only numerical values for the event variable. Graphical displays of a data set are useful to gain added insight contained in a data set that may not otherwise be clear by looking at test statistics such as t-test, Dixon test or T-S test. Unlike test statistics (e.g., t-test, MK test, AD test) and decision statistics (e.g., UCL, UTL), graphical displays do not get influenced by outliers and nondetect observations. It is suggested that the final decisions be made based upon statistical results as well as graphical displays.

Side-by-side box plots or multiple Q-Q plots are useful to graphically compare concentrations of two or more groups (e.g., several monitoring wells). The GOF module of ProUCL generates Q-Q plots for normal, gamma, and lognormal distributions based upon uncensored as well as left-censored data sets with NDs. All relevant information such as the test statistics, critical values and p-values (when available) are also displayed on the GOF Q-Q plots. In addition to providing information about the data distribution, a *normal* Q-Q plot in the original raw scale also helps to identify outliers and multiple populations that may be present in a data set. On a Q-Q plot, observations well-separated from the majority of the data may represent potential outliers coming from a population different from the main dominant population (e.g., background population). In a Q-Q plot, jumps and breaks of significant magnitude suggest the presence of observations coming from multiple populations (onsite and offsite areas). ProUCL can also be used to display box plots with horizontal lines displayed at pre-specified compliance limits or computed upper limits (e.g., UPL, UTL) superimposed on the same graph. This kind of graph provides a visual comparison of site data with compliance limits and/or BTV estimates.

Outlier Tests: ProUCL also has a couple of classical outlier test procedures (EPA 2006b, 2009), such as the Dixon test and the Rosner test. The details of these outlier tests are described in Chapter 7. These outlier tests often suffer from masking effects in the presence of multiple outliers. It is suggested that the classical outlier procedures should always be accompanied by graphical displays including box plots and Q-Q plots. Description and use of the robust and resistant (to masking) outlier procedures (Rousseeuw and Leroy, 1987; Singh and Nocerino, 1995) are beyond the scope of ProUCL 5.0. Interested users are encouraged to try the Scout 2008 software package (EPA 2009) to use the robust outlier identification methods especially when dealing with multivariate data sets consisting of data for several variables/analytes.

Outliers represent observations coming from populations different from the main dominant population represented by the majority of the data set. Outliers distort most statistics (e.g., mean, UCLs, UPLs, test statistics) of interest. Therefore, it is desirable to compute decisions statistics based upon data sets representing the main dominant population and not to compute distorted statistics by accommodating a few low probability outliers (e.g., by using a lognormal distribution). Moreover, it should be noted that even though outliers might have minimal influence on hypotheses testing statistics based upon ranks (e.g., WMW test), outliers do distort several nonparametric statistics including bootstrap methods such as bootstrap-t and Hall's bootstrap UCLs and other nonparametric UPLs and UTLs computed using the higher order statistics.

Goodness-of-Fit Tests: In addition to computing simple summary statistics for data sets with and without NDs, ProUCL 5.0 has GOF tests for normal, lognormal and gamma distributions. To test for normality (lognormality) of a data set, ProUCL has the Lilliefors test and the extended S-W test for samples of sizes up to 2000 (Royston, 1982, 1982a). For the gamma distribution, two GOF tests: the Anderson-Darling test (1954) and Kolmogorov Smirnov test (Schneider, 1978) are available in ProUCL. For samples of larger sizes (e.g., with $n > 100$) and small values of the gamma shape parameter, k (e.g., $k \leq 0.1$), significant discrepancies were found in the critical values of the two gamma GOF test statistics (Anderson-Darling and Kolmogorov Smirnov tests) obtained using the two gamma deviate generation algorithms: Whitaker (1974) and Marsaglia and Tsang (2000). For values of $k \leq 0.2$, the critical values of the two gamma GOF tests: Anderson-Darling (A-D) and Kolmogorov-Smirnov (K-S) tests have been updated using the currently available more efficient gamma deviate generation algorithm due to Marsaglia and Tsang's (2000); more details about the implementation of their algorithm can be found in Kroese, Taimre, and Botev (2011). For values of the shape parameter, $k=0.025$, 0.05, 0.1, and 0.2, the critical value tables for these two GOF tests have been updated by incorporating the newly generated critical values for three levels of significance: 0.05, 0.1, and 0.01. The updated tables are provided in Appendix A. It should be noted that for $k=0.2$, the older (generated in 2002) and the newly generated critical values are in general agreement.

ProUCL also generates GOF Q-Q plots for normal, lognormal, and gamma distribution displaying all relevant statistics including GOF test statistics. GOF tests for data sets with and without NDs are described in Chapter 8 of this User Guide. For data sets consisting of NDs, it is not easy to verify the distributional assumptions correctly, especially when the data set consists of a large percentage of NDs with multiple DLs and NDs exceeding the detected values. Typically, decisions about distributions of data sets with NDs are based upon GOF test statistics computed using the data obtained: without NDs; replacing NDs by 0, DL, or DL/2; using imputed NDs based upon a ROS (e.g., lognormal ROS) method. For data sets with NDs, ProUCL can perform GOF tests using methods listed above. Using the "Imputed NDs using ROS Methods" option of the "Stats/Sample Sizes" module of ProUCL 5.0, additional columns can be generated to store imputed (estimated) values for NDs based upon normal ROS, gamma ROS, and lognormal ROS (also known as robust ROS) methods.

Sample Size Determination and Power Evaluation: Sample Sizes module in ProUCL can be used to develop DQOs based sampling designs needed to address statistical issues associated with the various polluted sites projects. ProUCL 5.0 provides user friendly options to enter the desired/pre-specified values for decision parameters (e.g., Type I and Type II error rates) and other DQOs used to determine minimum sample sizes for the selected statistical applications including: estimation of mean, single and two-sample hypothesis testing approaches, and acceptance sampling. Both parametric (e.g., for t-tests) and nonparametric (e.g., Sign test, WRS test) sample size determination methods as described in EPA (2002c, 2006a, 2006b) and MARSSIM (2000) guidance documents are available in ProUCL version 5.0. ProUCL also has the sample size determination option for acceptance sampling of lots of discrete objects such as a lot (batch, set) of drums containing of hazardous waste (e.g., RCRA applications, EPA 2002c). When the sample size for an application (e.g., verification of cleanup level) is not computed using the DQOs based sampling design process, the Sample Size module can be used to assess the power of the test statistic used in retrospect. The Sample Sizes module with examples is considered in Chapter 12 of this document.

Bootstrap Methods: Bootstrap methods are computer intensive nonparametric methods which can be used to compute decision statistics of interest when a data set does not follow a known distribution, or when it is difficult to analytically derive the distributions of statistics of interest. It is well-known that for moderately skewed to highly skewed data sets, UCLs based upon standard bootstrap and the percentile bootstrap methods do not perform well (e.g., Efron [1981, 1982]; Efron and Tibshirani,1993; Hall

[1988,1992]; Singh, Singh, and Iaci 2002; Singh and Singh, 2003, Singh, Maichle and Lee 2006) as the interval estimates based upon these bootstrap methods fail to provide the specified coverage (e.g., UCL 95 does not provide adequate 95% coverage to population mean) to the population mean. For skewed data sets, Efron and Tibshirani (1993) and Hall (1988, 1992) considered other bootstrap methods such as the BCA, bootstrap-t and Hall's bootstrap methods. For skewed data sets, bootstrap-t and Hall's bootstrap (meant to adjust for skewness) methods perform better (e.g., in terms of coverage for the population mean) than the other bootstrap methods. However, it has been noted (e.g., Efron and Tibshirani ,1993, Singh, Singh, and Iaci,2002) that these two bootstrap methods tend to yield erratic and inflated UCL values (orders of magnitude higher than other UCLs) in the presence of outliers. Similar behavior of the bootstrap- t UCL and Hall's bootstrap UCL methods is observed on data sets consisting of NDs and outliers. Due to the reasons described above, whenever applicable, ProUCL 5.0 provides cautionary notes and warning messages regarding the use of bootstrap-t and Halls bootstrap UCL methods.

- For nonparametric uncensored and left-censored data sets with NDs, depending upon data variability and skewness, ProUCL recommends the use of BCA bootstrap, bootstrap-t, or Chebyshev inequality based methods to compute decision statistics.

<u>Hypotheses Testing Approaches:</u> ProUCL software has both Single- Sample (e.g., Student's t-test, sign test, proportion test, WSR test) and Two-Sample (Student's t-test, WMW test, Gehan test, and T-W test) parametric and nonparametric hypotheses testing approaches. Hypotheses testing approaches in ProUCL can handle both full-uncensored data sets without NDs, and left-censored data sets with NDs. Most of the hypotheses tests also report associated p-values. For some hypotheses tests (e.g., WMW test, WSR test, proportion test), large sample p-values based upon normal approximation are computed using the continuity correction factors. The various Single-sample and Two-Sample hypotheses testing approaches are considered in Chapter 9.

- <u>Single-sample:</u> parametric (Student's t-test) and nonparametric (Sign test, WSR test, tests for proportions and percentiles) hypotheses testing approaches are available in ProUCL. The single-sample hypotheses tests are used when the environmental parameters such as the cleanup standard, action level, or compliance limits are known, and the objective is to compare site concentrations with those known threshold values. Specifically, a t-test (or a sign test) may be used to verify the attainment of cleanup levels at an AOC) after a remediation activity has taken place; 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 observations collected from an AOC (or a MW) exceeds a certain specified proportion (e.g., 1%, 5%, 10%).

- The differences between these tests should be noted and understood. Specifically, a t-test or a Wilcoxon Signed Rank (WSR) test are used to compare the measures of location and central tendencies (e.g., mean, median) of a site area (e.g., AOC) to a cleanup standard, C_s or action level also representing a measure of central tendency (e.g., mean, median); whereas, a proportion test compares if the proportion of site observations from an AOC exceeding a compliance limit (CL) exceeds a specified proportion, P_0 (e.g., 5%, 10%). The percentile test compares a specified percentile (e.g., 95^{th}) of the site data to a pre-specified upper threshold (e.g., action level).

- <u>Two-sample:</u> Hypotheses tests (Student's t-test, WMW test, Gehan test, T-W test) are used to perform site versus background comparisons, compare concentrations of two or more AOCs, compare concentrations of GW monitoring wells (MWs). It should be noted that as cited in the literature, some of the hypotheses testing approaches (e.g., nonparametric two-sample WMW) deal with the single detection limit scenario. When using the WMW test on a data set with multiple detection limits, all observations (detects and NDs) below the largest detection limit need to be considered as NDs (Gilbert, 1987). This in turn tends to reduce the power and increase uncertainty associated with test. As mentioned before, it is always desirable to supplement the test statistics and conclusions with graphical displays such as the multiple Q-Q plots and side-by-side box plots. Gehan test or Tarone-Ware (new in ProUCL 5.0) should be used in cases where multiple detection limits are present.

Computation of Upper Limits including UCLs, UPLs, UTLs, and USLs: ProUCL software has parametric and nonparametric methods including bootstrap and Chebyshev inequality based methods to compute the various decision making statistics such as UCLs of mean (EPA 2002a), percentiles, UPLs for future k (≥1) observations, UTLs (e.g., EPA 1992b, EPA 2009) and upper simultaneous limits (USLs) (Singh and Nocerino, [1995, 2002]) based upon uncensored full data sets and left-censored data sets consisting of NDs with multiple DLs. Methods incorporated in ProUCL cover a wide range of skewed data distributions with and without NDs. In addition to normal and lognormal distributions based upper limits, ProUCL 5.0 can compute parametric UCLs, percentiles, UPLs for future k (≥1) observations, UTLs, and USLs based upon gamma distributed data sets. For data sets with NDs, ProUCL has several estimation methods including the KM method (1958), ROS methods (Helsel, 2005) and substitution methods such as replacing NDs by DL or DL/2 (Gilbert, 1987, EPA 2006b). Substitution DL/2 method has been incorporated in ProUCL for research and comparison purposes as requested by EPA scientists.

Computation of UCLs Based Upon Uncensored Data Sets without NDs: Parametric UCL computation methods in ProUCL for uncensored data sets include: Student's t-UCL, Approximate gamma UCL (using chi-square approximation), Adjusted gamma UCL (adjusted for level significance), Land's H-UCL, and Chebyshev inequality-based UCL (using MVUEs of parameters of a lognormal distribution). Nonparametric UCL computation methods for data sets without NDs include: CLT-based UCL, Modified-t-statistic (adjusted for skewness)-based UCL, Adjusted-CLT (adjusted for Skewness)-based UCL, Chebyshev inequality based-UCL (using sample mean and standard deviation), Jackknife method-based UCL, UCL based upon standard bootstrap, UCL based upon percentile bootstrap, UCL based upon BCA bootstrap, UCL based upon bootstrap-t, and UCL based upon Hall's bootstrap method. The details of UCL computation methods for uncensored data sets are summarized in Chapter 2 of the associated ProUCL 5.0 Technical Guide; and computations of the various parametric and nonparametric UCLs using ProUCL 5.0 are described in Chapter 11 of this document.

Computations of UPLs, UTLs, and USLs Based Upon Uncensored Data Sets without NDs: For uncensored data sets without NDs, ProUCL can compute parametric percentiles, UPLs for k (k≥1) future observations, UPLs for mean of k (≥1) future observations, UTLs, and USLs based upon normal, gamma, and lognormal distributions. Nonparametric upper limits are typically based upon order statistics of a data set such as a background or a reference area data set. Depending upon the size of the data set, the higher order statistics (maximum, second largest, third largest, and so on) are used to compute these upper limits (e.g., UTLs). Depending upon the sample size, specified confidence coefficient and coverage probability, ProUCL 5.0 outputs the actual confidence coefficient achieved by a nonparametric UTL. The mathematical details of the various parametric and nonparametric computation methods for UPLs, UTLs, and USLs are described in Chapter 3 of the ProUCL 5.0 Technical Guide; and computations of the these intervals using ProUCL 5.0 are described in Chapter 10 of this User Guide.

Computation of UCLs, UPLs, UTLs, and USLs Based Upon Left-Censored Data Sets with NDs: For data sets with NDs, ProUCL computes UCLs, UPLs, UTLs, and USLs based upon mean and *sd* computed using logROS (LROS, robust ROS), Gamma ROS (GROS), KM, and DL/2 methods. For nonparametric data sets, to adjust for skewness, ProUCL uses bootstrap methods and Chebyshev inequality to compute UCLs and other limits using estimates of mean and standard deviation obtained using methods listed above. ProUCL also uses parametric methods on KM (and ROS) estimates provided detected observations in the left-censored data set follow a parametric distribution. For example, if the detected data follow a gamma distribution, ProUCL uses KM estimates in gamma distribution based equations to compute UCLs, UTLs, and other upper limits. Based upon a Monte Carlo study performed by Singh, Maichle, and Lee (EPA, 2006), ProUCL recommends the use of the Kaplan-Meier (1958) estimates in bootstrap and Chebyshev inequality to compute the various decision statistics (e.g.,UCL95, UPL, UTL) of interest. ProUCL 5.0 suggests the use of KM-Gamma upper limits when the detected data follow a gamma distribution. ProUCL computes KM estimates directly using left-censored data sets without flipping data and re-flipping decision statistics. The KM method incorporated in ProUCL computes both *sd* and standard error (*SE*) of the mean. For historical reasons and for comparison and research purposes, the DL/2 substitution method and H-UCL based upon LROS method have been retained in ProUCL 5.0. The inclusion of the substitution method in ProUCL should not be inferred as an endorsement of those methods by ProUCL software and its developers. The mathematical details of the UCL computation methods for data sets with NDs are given in Chapter 4 and the description of the various other upper limits: UPLs, UTLs, and USLs for data sets with NDs are given in Chapter 5 of the ProUCL 5.0 Technical Guide. The computations of these limits for data sets consisting of NDs using ProUCL 5.0 are considered in chapters 10 and 11 of this User Guide.

<u>One-Way ANOVA, OLS Regression and Trend Analysis:</u> The Oneway ANOVA module has both classical and nonparametric K-W ANOVA tests as described in EPA guidance documents (e.g., EPA [2006b, 2009]). Oneway ANOVA is used to compare means (or medians) of multiple groups such as comparing mean concentrations of several areas of concern; and performing inter-well comparisons comparing concentrations of several MWs. The OLS Regression option computes the classical OLS regression line, and generates graphs displaying the OLS line, confidence bands and prediction bands around the regression line. All statistics of interest including slope, intercept, and correlation coefficient are displayed on the OLS line graph. The Trend Analysis module has two nonparametric trend tests: M-K trend test and Theil-Sen trend test. Using this option, one can generate trend graphs and time-series graphs displaying Theil-Sen trend line and all other statistics of interest with associated p-values.

In GW monitoring applications, OLS regression, trend tests, and time series plots are often used to identify trends (e.g., upwards, downwards) in constituent concentrations of the various GW monitoring wells over a certain period of time (EPA 2009). The details of Oneway ANOVA are given in Chapter 9, and OLS regression line and Trend tests methods are described in Chapter 10 of the ProUCL 5.0 Technical Guide. Chapters 13 and 14 of this User Guide respectively, illustrate the use of Oneway ANOVA module and OLS Regression and Trend Analysis module.

<u>BISS Module:</u> At many sites, a large amount of discrete onsite and background data are already available which are not directly comparable to actual field ISM data. In order to provide a tool to compare the existing discrete data with ISM data, the BISS module of ProUCL 5.0 may be used on a large existing discrete data set. The ISM methodology used to develop the BISS module is a relatively new approach; methods incorporated in this BISS module require further investigation. The BISS module has been temporarily blocked for use in ProUCL 5.0 as this module is awaiting adequate guidance for its intended use on discrete background data sets.

<u>Recommendations and Suggestions in ProUCL:</u> Not much guidance is available in the environmental literature including the available guidance documents to compute rigorous UCLs, UPLs, and UTLs for moderately skewed to highly skewed uncensored and left-censored data sets consisting of NDs with multiple DLs, a common occurrence in environmental data sets. For uncensored positively skewed data sets, Singh, Singh, and Iaci (2002) and Singh and Singh (2003) performed extensive simulation

experiments to compare the performances (in terms of coverage probabilities) of several UCL computation methods described in statistical and environmental literature. They noted that the optimal choice of a decision statistic (e.g., UCL 95) depends upon the sample size, data distribution and data skewness. Until 2006, not much guidance was available on how to compute UCL95 of mean and other upper limits (e.g., UPLs and UTLs) based upon skewed data sets consisting of NDs with multiple DLs. For data sets with NDs, Singh, Maichle, and Lee (EPA 2006) conducted a similar simulation study to compare the performances of the various estimation methods (in terms of bias in the mean estimate); and of some the UCL computation methods (in terms of coverage provided by a UCL). They concluded that the nonparametric KM estimation method performs well in terms of bias in estimate of the mean; and for skewed data sets, t-statistic, CLT, and the percentile bootstrap method based UCLs computed using KM estimates (and ROS estimates) underestimate the population mean. Based upon the findings summarized in Singh, Singh, and Iaci (2002) and Singh, Maichle, and Lee (2006), it is reasonable to state and assume that the findings of the simulation studies performed on uncensored skewed data sets to compare the performances of the various UCL computation methods can be extended to skewed left-censored data sets.

For data sets with and without NDs, ProUCL computes decision statistics including UCLs, UPLs, and UTLs using several parametric and nonparametric methods covering a wide-range of sample size, data variability and skewness. Using the results and findings summarized in the literature cited above, based upon the sample size, data distribution, and data skewness, some modules of ProUCL make suggestions about using a decision statistic to estimate population parameters of interest (e.g., EPC). The recommendations made in ProUCL are based upon the extensive experience of the developers in environmental statistical methods, published literature (e.g., Efron and Tibshirani, 1993; Hall, 1988; Singh, Singh, and Engelhardt 1997; Singh, Singh, and Iaci 2002; and Singh, Maichle, and Lee 2006) and procedures described in the various EPA guidance documents (EPA [1992a, 1992b 2002a, 2002b, 2006b, 2009, 2009a, 2009b]). Based upon the conceptual site model (CSM), expert site and regional knowledge, the project team should make the final decision regarding using or not using the suggestions made by ProUCL. If deemed necessary, the project team may want to consult a statistician.

Even though, ProUCL 5.0 has been developed using limited government funding, for data sets with and without NDs, ProUCL 5.0 provides many statistical and graphical methods described in the EPA documents cited above. However, one may not compare the availability of methods in ProUCL 5.0 with methods available in the commercial software packages such as SAS and Minitab 16. For example, trend tests correcting for seasonal/spatial variations are not available in the ProUCL software. For those methods the user is referred to the commercial software packages. As mentioned earlier, it is recommended to supplement test results (e.g., two-sample test) with graphical displays (e.g., Q-Q plots, side-by-side box plots); especially when data sets consist of NDs and outliers. With the inclusion of BISS module, Oneway ANOVA, Regression and Trend tests, and the user-friendly DQOs based Sample Size determination modules, ProUCL represents a comprehensive statistical software package equipped with statistical methods and graphical tools needed to address many environmental sampling and statistical issues as described in the various CERCLA (EPA 1989a, 1992a, 2002a, 2002b, 2006a, 2006b), MARSSIM (EPA 2000), and RCRA (EPA 1989b, 1992b, 2002c, 2009) guidance documents.

Finally, the users of ProUCL are cautioned about the use of methods and suggestions described in some recent environmental literature. For example, many decision statistics (e.g., UCLs, UPLs, UTLs,) computed using the methods (e.g., percentile bootstrap, statistics using KM estimates and t-critical values) described in Helsel (2012) will fail to provide desired coverage to the environmental parameters of interest (mean, upper percentile) of moderately skewed to highly skewed populations; and conclusions derived based upon those decisions statistics may lead to incorrect conclusions which may not be cost-effective or protective of human health and the environment.

ProUCL 5.0 Technical Guide

In addition to this User Guide, a Technical document also accompanies ProUCL 5.0.00, providing technical details of the graphical and statistical methods incorporated in ProUCL 5.0.00. Most of the mathematical algorithms and formulae (with references) used in the development of ProUCL 5.0 are described in the associated Technical Guide.

Chapter 1

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

Decisions based upon statistics computed using discrete data sets of small sizes (e.g., < 6) cannot be considered reliable enough to make remediation decisions that affect human health and the environment. For example, a background data set of size less than 6 is not large enough to characterize background population, to compute background threshold values (BTV) estimates, or to perform background versus site comparisons. Several EPA guidance documents (e.g., MARSSIM 2000; EPA [2006a, 2006b]) describe data quality objectives (DQOs) and minimum sample size computations needed to address statistical issues associated with the various environmental applications. In order to obtain reliable results using statistical methods, an adequate amount of data should be collected using desired DQOs (confidence coefficient, decision error rates). The Sample Sizes module of ProUCL computes DQOs based minimum sample sizes needed to use the statistical methods described in the various guidance documents. In some cases, it may not be possible (e.g., due to resource constraints) to collect DQOs based number of samples; under these circumstances one can use the Sample Sizes module to assess the power of the test statistic used in retrospect. Some suggestions about the minimum sample size requirements needed to use statistical methods to estimate environmental parameters of interest such as exposure point concentration (EPC) terms and BTVs, to compare site data with background data or with some pre-established screening levels (e.g., action levels [ALs], compliance limits [CLs]), are provided in this chapter. It is noted that similar minimum sample size suggestions made by ProUCL (EPA 2007, 2009a, 2009b) have been made in some other guidance documents including the RCRA Guidance Document (EPA 2009).

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 with. 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 onsite observations are compared with those threshold values. It should be noted that 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 (known) or estimated (unknown) standards and BTVs. Several upper limits, and single and two sample hypotheses testing approaches, for both full-uncensored and left-censored data sets are available in the ProUCL software package to perform the comparisons described above.

1.1 Background Data Sets

Based upon the conceptual site model (CSM), the project team familiar with the site 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. An appropriate random sample of independent observations (e.g., *i.i.d*) should be collected from the background area. A defensible background data set represents a "single" population possibly without any outliers. In a background data set, in addition to reporting and/or laboratory errors, statistical outliers may also be present. A few elevated statistical outliers present in a background data set may actually represent potentially contaminated locations belonging to impacted site areas and/or possibly from other polluted site(s); those elevated outliers may not be coming from the main dominant background population under evaluation. Since the presence of outliers in a data set tends to yield distorted (incorrect and misleading) values of the decision making statistics (e.g., UCLs, UPLs and UTLs), elevated outliers should not be included in background data sets and estimation of BTVs. The objective here is to compute background statistics based upon the majority of the data set representing the main dominant background population, and not to accommodate a few low probability high outliers (e.g., coming from extreme tails of the data distribution) that may also be present in the background data set. The occurrence of elevated outliers is common 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. Since both of these classical tests suffer from masking effects (e.g., some extreme outliers may mask the occurrence of other intermediate outliers), it is suggested that these classical outlier tests be supplemented with graphical displays such as a box plot and a Q-Q plot. The use of exploratory graphical displays helps in determining the number of outliers potentially present in a data set. The use of graphical displays also helps in identifying extreme high outliers as well as intermediate and mild outliers. The use of robust and resistant outlier identification procedures (Singh and Nocerino, 1995, Rousseeuw and Leroy, 1987) is recommended when multiple outliers are present in a data set. Those methods are beyond the scope of ProUCL 5.0. However, several robust outlier identification methods are available in the Scout 2008 version 1.0 software package (EPA 2009).

An appropriate background data set of a reasonable size (preferably computed using DQOs processes) is needed to represent a background area and to compute upper limits (e.g., estimates of BTVs) based upon background data sets and also to compare site and background data sets using hypotheses testing approaches. At the minimum, a background data set should have at least 10 (more observations are preferable) observations to perform background evaluations.

1.2 Site Data Sets

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

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 action levels 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. If many onsite values need to be compared with a BTV, it is recommended to use UTL or upper simultaneous limit (USL) to control the false pos1itive error rate (Type I Error Rate). Alternatively, one can use hypothesis testing approaches provided enough observations (at least 10, more are preferred) are available.

- single-sample hypotheses tests to compare site data with a pre-established cleanup standards, 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 least 10 (more are desirable) site observations are available; it is preferable to use hypotheses testing approaches to compare site observations with specified threshold values. The use of hypotheses testing approaches can control 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 the BTV (e.g., 95^{th} 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.

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 at 10 observations 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 by graphical displays.

1.3 Discrete Samples or Composite Samples?

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

- When using a single-sample hypothesis testing approach, site data can be obtained by collecting all discrete or all composite samples. The hypothesis testing approach is used when many (e.g., ≥

10) site observations are available. Details of the single-sample hypothesis approaches are widely available in EPA guidance documents (MARSSIM, 2000; EPA [1989a 2006b]). Several single-sample hypotheses testing procedures available in ProUCL are described in Chapter 6 of the ProUCL 5.0 Tech Guide.

- If a two-sample hypothesis testing approach is used to perform site versus background comparisons, then samples from both of the populations should be either all discrete samples, or all composite samples. The two-sample hypothesis testing approaches are used when many (e.g., at least 10) site, as well as background, observations are available. For better results with higher statistical power, the availability of more observations perhaps based upon an appropriate DQOs process (EPA 2006a) is desirable. Several two-sample hypotheses tests available in ProUCL 5.0 are described in Chapter 6 of the ProUCL 5.0 Tech Guide.

1.4 Upper Limits and Their Use

The computation and use of statistical limits depend upon their applications and the parameters (e.g., EPC term, BTVs) they are supposed to be estimating. Depending upon the objective of the study, a pre-specified cleanup standard, $C_{s,}$ can be viewed as to represent: 1) an average (or median) constituent concentration, μ_0; or 2) a not-to-exceed upper threshold concentration value, A_0. These two threshold values, an average value, μ_0, and a not-to-exceed value, A_0, represent two significantly different parameters, and different statistical methods and limits are used to compare the site data with these two very different threshold values. Statistical limits, such as an UCL of the population mean, an UPL for an independently obtained "single" observation, or independently obtained "k" observations (also called future k observations, next k observations, or k different observations), upper percentiles, and UTLs are often used to estimate the environmental parameters: an EPC term (μ_0) and a BTV (A_0). A new upper limit, USL has been included in ProUCL 5.0 which may be used to estimate a BTV based upon a well-established background data set without any outliers.

It is important to understand and note the differences between the uses and numerical values of these statistical limits so that they can be properly used. Specifically, the differences between UCLs and UPLs (or upper percentiles), and UCLs and UTLs should be clearly understood and acknowledged. A UCL with a 95% confidence limit (UCL95) of the mean represents an estimate of the population mean (measure of the central tendency), whereas a UPL95, a UTL95%-95% (UTL95-95), and an upper 95th percentile represent estimates of a threshold from the upper tail of the population distribution such as the 95th percentile. Here, UPL95 represents a 95% upper prediction limit, and UTL95-95 represents a 95% confidence limit of the 95th 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 $\leq$ UCL95 of Mean $\leq$ Upper 95th Percentile $\leq$ UPL95 of a Single Observation $\leq$ UTL95-95

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). The differences among the various upper limits discussed above are illustrated by the following example.

Example 1.1. Consider a background real data set collected from a Superfund site (EPA 2002b). The data set has several inorganic COPC, including aluminum, arsenic, chromium, iron, and lead. Iron concentrations follow a normal distribution. Some upper limits for the iron data set are summarized as follows. However, the various upper limits do follow the order as described above.

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

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

A brief discussion about the differences between the applications and uses of the various statistical limits is provided below.

- A UCL represents an average value that should be compared with a threshold value also representing an average value (pre-established or estimated), such as a mean $C_{s.}$ For example, a site 95% UCL exceeding a $C_{s,}$ may lead to the conclusion that the C_s has not been attained by the average site area concentration. It should also be noted that UCLs of means are typically computed based upon 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 (single-sample hypothesis), or with the background population average or median (two-sample hypothesis).

- A UPL, an upper percentile, or an UTL represents an upper limit to be used for point-by-point individual site observation comparisons. UPLs and UTLs are computed based upon background data sets, and 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.

- When enough (e.g., at least10) site observations are available, it is preferable to use hypotheses testing approaches. Specifically, single-sample hypotheses testing (comparing site to a specified threshold) approaches should be used to perform site versus a known threshold comparison; and two-sample hypotheses testing (provided enough background data are also available) approaches should be used to perform site versus background comparison. Several parametric and nonparametric single and two-sample hypotheses testing approaches are available in ProUCL 5.0.

It is re-emphasized that only averages should be compared with averages or UCLs, 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.

1.5 Point-by-Point Comparison of Site Observations with BTVs, Compliance Limits, 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 an onsite (or a monitoring well) observation may be considered as 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 (taking a few more collocated samples) that site location (or a monitoring well) exhibiting constituent concentration in excess of the BTV. If all collocated (or collected during the same time period) sample observations collected from the same site location (or well) exceed the BTV or PRG, then it may be concluded that the location (well) requires further investigation (e.g., continuing treatment and monitoring) and 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 representing 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, enough background observations (minimum of 10) should be collected, perhaps using an appropriate DQOs process as described in EPA (2006a) and MARSSIM (2000). Several statistical limits listed above are used to estimate the BTVs based upon a defensible (free of outliers, representing the background population) background data set of an adequate size.

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 or a 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 contaminants/constituents of potential concern (COPCs), 2) identify the polluted site AOCs, or 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 various onsite locations representing the site area under investigation, then the use of hypotheses testing approaches (both single-sample and a two-sample) is preferred. The use of hypothesis testing approaches control the error rates more tightly and efficiently than the individual point-by-point site comparisons.

1.6 Hypothesis Testing Approaches and Their Use

Both single-sample and two-sample hypotheses testing approaches are used to make cleanup decisions at polluted sites, and also to compare constituent concentrations of two (e.g., site versus background) or more populations (e.g., MWs).

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

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 (provided enough site data are available) with known and pre-established threshold values. It is suggested that the project team determine (e.g., using DQOs) or decide (depending upon resources) about the number of site observations that should be collected and compared with the "pre-established" standards before coming to a conclusion about the status (clean or polluted) of the site AOCs. As mentioned earlier, when the number of available site samples is less than 6, one might perform point-by-point site observation comparisons with a BTV; and when enough site observations (at least 10) are available, it is desirable to use single-sample hypothesis testing approaches. Depending upon the parameter (e.g., the average value, μ_0, or a not-to-exceed value, A_0), represented by the known threshold value, one can use single-sample 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 this 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 mean) is used (assuming normality of site data set or when sample size is large such as larger than 30, 50) 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 all NDs (e.g., associated detection limits) 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 (or equivalently a 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.

1.6.2 Two-Sample Hypotheses (BTVs and Not-to-Exceed Values are Unknown)

When BTVs, not-to-exceed values, and other cleanup standards are not available, then site data are compared directly with the background data. In such cases, 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. Site and background data requirements (e.g., based upon DQOs) to perform two-sample hypothesis test approaches are described in EPA (2002b, 2006a, 2006b), MARSSIM (2000) and also in Chapter 6 of the ProUCL 5.0 Technical Guide. 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 (e.g., > 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 (or more) hypotheses approaches are used when the site parameters (e.g., mean, shape, distribution) are being compared with the background parameters (e.g., mean, shape, distribution). The two-sample hypotheses testing approach is also used when the cleanup standards or screening levels are not known *a priori*. Specifically, in environmental applications, two-sample hypotheses testing approaches are used to compare average or median constituent concentrations of two or more populations. To derive reliable conclusions with higher statistical power based upon hypothesis testing approaches, an adequate amount of data (e.g., minimum of 10 samples) should be collected from all of the populations under investigation.

The two-sample hypotheses testing approaches incorporated in ProUCL 5.0 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 Tarone-Ware tests are meant to be used on left-censored data sets with multiple detection limits (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 be all discrete or all composite (obtained using the same design and pattern), and be collected from the same medium (soil) at similar depths (e.g., all surface samples or all subsurface samples) and time (e.g., during the same quarter in groundwater applications) using comparable (preferably same) analytical methods. Good sample collection methods and sampling strategies are given in EPA (1996, 2003) guidance documents.

Notes: ProUCL 5.0 (and previous versions) has been developed using limited government funding. ProUCL 5.0 is equipped with statistical and graphical methods needed to address many environmental sampling and statistical issues as described in the various CERCLA, MARSSIM, and RCRA documents cited earlier. However, one may not compare the availability of methods in ProUCL 5.0 with methods incorporated in commercial software packages such as SAS and Minitab 16. Not all methods available in the statistical literature are available in ProUCL.

1.7 Minimum Sample Size Requirements and Power Assessment

Due to resource limitations, it may not be possible (nor needed) to sample the entire population (e.g., background area, site area, AOCs, EAs) under study. Statistics is used to draw inference(s) about the populations (clean, dirty) and their known or unknown parameters (e.g., mean, variance, upper threshold values) based upon much smaller data sets (samples) collected from those populations. To determine and establish BTVs and site specific screening levels, defensible data set(s) of appropriate size(s) need to be collected from background areas (e.g., site-specific, general reference area, or historical data). The project team and site experts should decide what represents a site population and what represents a background population. The project team should determine the population area and boundaries based upon all current and 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, one needs to determine the appropriate sample sizes that need to be collected from the populations (e.g., site and background) under investigation using appropriate DQOs processes (EPA [2006a, 2006b]; MARSSIM, 2000). ProUCL has the Sample Sizes module which can be used to develop DQOs based sampling designs needed to address statistical issues associated with the various 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 5.0. ProUCL 5.0 also has sample size determination methods for acceptance sampling of lots of discrete objects such as a lot of drums containing hazardous waste (e.g., RCRA applications, EPA 2002c).

However, due to budget constraints, it may not be possible to collect the same number of samples as determined by using a DQOs process. For example, the data might have already been collected (often is the case) without using a DQOs process, or due to resource constraints, it may not be possible to collect as many samples as determined by using a DQOs based sample size formula. In practice, the project team and the decision makers may decide not to collect enough background samples. It is suggested to collect at least10 background observations before using statistical methods to perform background evaluations based upon data collected using discrete samples. The minimum sample size recommendations described here are useful when resources are limited, though it may not be possible to collect as many background and site samples as computed using DQOs based sample size determination formulae. In case data are collected without using a DQOs 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 (e.g., using Sample Size module of ProUCL) 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.

<u>Notes:</u> From a mathematical point of view, the statistical methods incorporated in ProUCL and described in this guidance document to estimate EPC terms and BTVs, and compare 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. It is recommended not to use those statistics to draw cleanup and remediation decisions potentially impacting human health and the environment. The minimum sample size recommendation (at least 10 observations) may be used only when data sets of size determined by a DQOs process (EPA, 2006) cannot be collected. Some of the recent guidance documents (e.g., EPA 2009) are also suggesting collecting a minimum of about 10 samples in the circumstance that data cannot be collected using a DQOs based process.

- To allow the users to compute decision statistics based upon composite data collected using the Incremental Sampling Methodology (ITRC, 2012), ProUCL 5.0 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 Tech Reg Guide (2012) to determine which UCL (e.g., Student's t-UCL or Chebyshev UCL) should be used to estimate the EPC term.

1.7.1 *Sample Sizes for Bootstrap Methods*

Several nonparametric methods including bootstrap methods to compute UCL, UTL, and other limits for both full-uncensored data sets and left-censored data sets with NDs are available in ProUCL 5.0. 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 (with replacement) 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. Moreover, bootstrapping a large data set of size greater than 500 or 1000 will be time consuming.

1.8 Statistical Analyses by a Group ID

The analyses of data categorized by a group ID variable such as: 1) Surface vs. Subsurface; 2) AOC1 vs. AOC2; 3) Site vs. Background; and 4) Upgradient vs. Downgradient monitoring wells are common in environmental and various other applications. ProUCL 5.0 offers this option for data sets with and without NDs. The Group Option provides a useful tool to perform various statistical tests and methods (including graphical displays) separately for each of the group (samples from different populations) that may be present in a data set. The graphical displays (e.g., box plots, (quantile-quantile) Q-Q plots) and statistics (e.g., background statistics, UCLs, hypotheses testing approaches) 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 users' responsibility to provide adequate amount of data to perform the group operations. For an 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, standard deviation [*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. ProUCL generates a warning message (colored orange) in the lower Log Panel of the ProUCL 5.0 screen.

1.9 Statistical Analyses for Many Constituents/Variables

ProUCL software can process multiple analytes/variables simultaneously in a user friendly manner – an option not available in other software packages such as Minitab 16 (2012), NADA for R (Helsel, 2013). This option is very useful when one has to process multiple variables and compute decision statistics (e.g., UCLs, UPLs, and UTLs) and test statistics (e.g., ANOVA test, trend test) for those variables. It is the user's responsibility to make sure that each selected variable has an adequate amount of data so that ProUCL can perform the selected statistical method correctly. ProUCL displays warning messages when a selected variable does not have enough data needed to perform the selected statistical method.

1.10 Use of Maximum Detected Value as Estimates of Upper Limits

Some practitioners tend to use the maximum detected value as an estimate of the EPC term. This is especially true when the sample size is small such as < 5 or when a UCL95 exceeds the maximum detected values (EPA, 1992a). Also, many times in practice, the BTVs and not-to-exceed values are estimated by the maximum detected value (e.g., nonparametric UTLs, USLs).

1.10.1 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). The use of higher order statistics to estimate the UTLs depends upon the sample size. For an example, 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.

1.10.2 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 values (EPA, 1992a). Specifically, the EPA (1992a) document suggests the use of the maximum detected value as a default value to estimate the EPC term when a 95% UCL (e.g., the H-UCL) exceeds the maximum value. ProUCL computes 95% UCLs of mean using several methods based upon normal, gamma, lognormal, and non-discernible distributions. In the past (e.g., EPA 1992), a lognormal distribution was used as the default distribution to model positively skewed environmental data sets; and 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 (1971, 1975). The use of the H-statistic often yields unstable and impractically large UCL95 of the mean (Singh, Singh, and Engelhardt, 1997; Singh, Singh, and Iaci, 2002). For skewed data sets of smaller sizes (e.g., < 30, < 50,...), H-UCL often exceeds the maximum detected value. Since the use of a lognormal distribution has been quite common (e.g., suggested as a default model in a risk assessment guidance for Superfund [RAGS] document [EPA, 1992a]), the exceedance of the maximum value by an H-UCL95 is frequent for many skewed data sets of smaller sizes (e.g., < 30, < 50). These occurrences result in the possibility of using the maximum detected value as an estimate of the EPC term.

It should be pointed out that in some cases, the maximum observed value actually might represent an impacted location. Obviously, it is not desirable to use a potential outlier representing an impacted location to estimate the EPC term for an AOC. The EPC term represents the average exposure contracted by an individual over an EA during a long period of time; therefore, the EPC term should be estimated by using an average value (such as an appropriate 95% UCL of the mean) and not by the maximum observed concentration. One needs to compute an average exposure and not the maximum exposure. Singh and Singh (2003) studied the performance of the max test (using the maximum observed value as an estimate of the EPC term) via Monte Carlo simulation experiments. They noted that for skewed data sets of small sizes (e.g., < 10-20), 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.

Today, several methods, some of which are described in EPA (2002a), are available in the various versions of ProUCL (e.g., ProUCL 3.00.02 [EPA 2004], ProUCL 4.0 [EPA 2007], ProUCL 4.00.05[EPA 2009, 2010]) to estimate the EPC terms. For data sets with NDs, ProUCL 5.0 has some new UCL (and other limits) computation methods which were not available in earlier versions of ProUCL. 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.

1.10.2.1 Chebyshev Inequality Based UCL95

ProUCL 5.0 (and its earlier versions) 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 recommends the use of an alternative UCL computation method based upon the Chebyshev inequality. One may use a 97.5% or 99% Chebyshev UCL to estimate the mean of a highly skewed population. The use of the Chebyshev inequality to compute UCLs tends to yield more conservative (but stable) UCLs than other methods available in ProUCL software. In such cases, when the sample size is large (and other UCL methods such as the bootstrap-t method yield unrealistically high values due to presence of outlier(s)), one may want to use a 95% Chebyshev UCL or a Chebyshev UCL with lower confidence coefficient such as 90% as an estimate of the population mean, especially when the sample size is large (e.g., >100, 150). The detailed recommendations (as functions of sample size and skewness) for the use of those UCLs are summarized in various versions of ProUCL Technical Guides (EPA, 2004, 2007, 2009, and 2010d).

<u>Notes:</u> It is recommended not to use the maximum observed value to estimate the EPC term representing the average exposure contracted by an individual over an EA. For the sake of interested users, ProUCL displays a warning message when the recommended 95% UCL (e.g., Hall's bootstrap UCL) of the mean exceeds the observed maximum concentration. For such scenarios (when a 95% UCL does exceed the maximum observed value), an alternative 95% UCL computation method based upon Chebyshev inequality is recommended by the ProUCL software.

1.11 Samples with Nondetect Observations

ND observations are inevitable in most environmental data sets. Singh, Maichle, and Lee (EPA, 2006) studied the performances (in terms of coverages) of the various UCL95 computation methods including the simple substitution methods (such as the DL/2 and DL methods) for data sets with ND observations. They concluded that the UCLs obtained using the substitution methods, including the replacement of NDs by respective 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 to compute UCL95 based upon data sets with ND observations.

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

Based upon the results of the report by Singh, Maichle, and Lee (EPA, 2006), it is recommended to avoid the use of the DL/2 method to perform a GOF test, and to compute the summary statistics and various other limits (e.g., UCL, UPL, UTLs) often used to estimate the EPC terms and BTVs. Until recently, the DL/2 method has been the most commonly used method to compute the various statistics of interest for data sets with 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 Kaplan-Meier [KM] estimates) including

Chebyshev inequality and bootstrap methods with better performance are available that can be used to compute the various upper limits of interest. Several of those parametric and nonparametric methods are available in ProUCL 4.0 and higher versions. It should be noted that 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 DL/2 method in ProUCL for comparison and research purposes.

Notes: Even though the DL/2 method (to compute UCLs, UPLs, and for goodness-of-fit [GOF] tests) has been incorporated in ProUCL, its use is not recommended due to its poor performance. The DL/2 method has been retained in ProUCL 5.0 for historical and comparison purposes. NERL-EPA, Las Vegas strongly recommends avoiding the use of DL/2 method even when the % of NDs is as low as 5% to 10%.

1.12 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 reporting limit (RL) or the maximum RL/2. Statistics (e.g., UCL95) computed based upon only a few detected values (e.g., < 4) cannot be considered reliable enough to estimate the EPC terms having 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 the EPC terms and other upper limits. Specifically, it is suggested that for data sets consisting of less than 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 together should decide on a site-specific basis on how to estimate the average exposure (EPC term) for the constituent and area under consideration. For such data sets with low detection frequencies, other measures such as the median or mode represents better estimates (with lesser uncertainty) of the population measure of central tendency.

Additionally, it is also suggested that when most (e.g., > 95%) of the observations for a constituent lie below the DLs, the sample median or the sample mode (rather than the sample average) may be used as an estimate the EPC term. 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.

1.13 Some Other Applications of Methods in ProUCL 5.0

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

1.13.1 Identification of COPCs

Risk assessors and remedial project managers (RPMs) often use screening levels or BTVs to identify the COPCs during the screening phase of a cleanup project to be conducted at a contaminated site. The screening for the COPCs is performed prior to any characterization and remediation activities that may have to be conducted at the site. 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 to obtain reliable 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.

1.13.2 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 by a MW concentration may be considered an indication of contamination in that MW. In such individual concentration comparisons, the presence of contamination (determined by an exceedance) may have to be confirmed by re-sampling from that MW. If concentrations of constituents in the original sample and re-sample(s) exceed the MCL or BTV, then that MW may require further scrutiny, perhaps triggering remediation remedies as determined by the project team. If the concentration data from a MW for about 4 to 5 continuous quarters (or some other designated time period determined by the project team) are below the MCL or BTV level, then that MW may be considered as complying with (achieving) the pre-established or estimated standards.

1.13.3 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 polluted 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 DQOs process (or a minimum of 10 observations should be collected) 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 site 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*.

1.13.4 Using BTVs (Upper Limits) to Identify Hot Spots

The use of upper limits (e.g., UTLs) to identify hot spot(s) has also been mentioned in the *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (EPA, 2002b). Point-by-point site observations are compared with a pre-established or estimated BTV. Exceedances of the BTV by site observations may be considered as representing impacted locations with elevated concentrations (hot spots).

1.14 Some General Issues and Recommendations made by ProUCL

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

1.14.1 Multiple Detection Limits

ProUCL 5.0 does not make distinctions between method detection limits (MDLs), adjusted MDLs, sample quantitation limits (SQLs), or DLs. Multiple DLs in ProUCL mean different values of the DL. An indicator variable with of 0 (=nondetect) and 1(= detect) is assigned to each variable consisting of NDs. All ND observations in ProUCL are indentified by the value '0' of the indicator variable used in ProUCL to distinguish between detected (=1) and nondetected (=0) observations. It is the users' responsibility to supply correct numerical values for NDs (should be entered as the reported detection limit or RL values) and not as qualifiers (e.g., J, U, B, UJ, ...) for ND observations in the data set.

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

For data sets with NDs, ProUCL 5.0 can compute point estimates of population mean and standard deviation using the KM and ROS methods (and also using DL/2 method). The DL/2 method has been retained in ProUCL for historical and research purposes. ProUCL uses Chebyshev inequality, bootstrap methods, and normal, gamma, and lognormal distribution based equations on KM (or ROS) estimates to compute the various 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 and BCA bootstrap (mild skewness), KM estimates and Chebyshev inequality (moderate to high skewness), and KM estimates and bootstrap-t method (moderate to high skewness) yield better (in terms of coverage probability) estimates of EPC terms than other UCL methods based upon Student's t-statistic on KM estimates, percentile bootstrap method on KM or ROS estimates.

1.15 The Unofficial User Guide to ProUCL4 (Helsel and Gilroy, 2012)

Several ProUCL users sent inquiries about the validity of the comments made about the ProUCL software in the Unofficial User Guide to ProUCL4 (Helsel and Gilroy, 2012) and in the Practical Stats webinar, "ProUCL v4: The Unofficial User Guide," presented by Dr. Helsel on October 15, 2012 (Helsel 2012a). Their inquiries led us to review comments made about the ProUCL v4 software and its associated guidance documents (EPA 2007, 2009a, 2009b, 2010c, and 2010d) in the Unofficial ProUCL v4 User Guide and in the webinar, "ProUCL v4: The Unofficial User Guide". These two documents collectively are referred to as the Unofficial ProUCLv4 User Guide in this ProUCL document. The pdf document describing the material presented in the Practical Stats Webinar (Helsel, 2012a) was downloaded from the http://www.practicalstats.com website.

In the "ProUCL v4: The Unofficial User Guide", comments have been made about the software and its guidance documents, therefore, it is appropriate to address those comments in the present ProUCL guidance document. It is necessary to provide the detailed response to comments made in the Unofficial ProUCL v4 User Guide to assure that: 1) rigorous statistical methods are used to compute the decision making statistics; and 2) the methods incorporated in ProUCL software are not misrepresented and misinterpreted. Some general responses and comments about the material presented in the Practical Stats webinar and in the Unofficial User Guide to ProUCLv4 are described as follows. Specific comments and responses are also considered in the respective chapters of ProUCL 5.0 Technical and User Guides. The detailed responses to the comments made about the ProUCL software in the Unofficial ProUCL v4 User Guide are provided elsewhere.

ProUCL is a freeware software package which has been developed under limited government funding to address statistical issues associated with various environmental site projects. Not all statistical methods (e.g., Levene test) described in the statistical literature have been incorporated in ProUCL. One may not compare ProUCL with the commercial software packages which are expensive and not as easy to use as the ProUCL software to address environmental statistical issues. The existing and some new statistical methods based upon the research conducted by ORD-NERL, EPA Las Vegas during the last couple of decades have been incorporated in ProUCL to address the statistical needs of the various environmental site projects and research studies. Some of those new methods may not be available in text books, in the library of programs written in R-script, and in commercial software packages. However, those methods are described in detail in the cited published literature and also in the ProUCL Technical Guides (e.g., EPA [2007, 2009a, 2009b, 2010c and 2010d]). Even though for uncensored data sets, programs to compute gamma distribution based UCLs and UPLs are available in R Script, programs to compute a 95% UCL of mean based upon a gamma distribution on KM estimates are not easily available in commercial software packages and in R script.

- In the Unofficial ProUCL v4 User Guide, several statements have been made about percentiles. There are several ways to compute percentiles. Percentiles computed by ProUCL may or may not be identical (don't have to be) to percentiles computed by NADA for R (Helsel, 2013) or described in Helsel and Gilroy (2012). To address users' requests, ProUCL 4.1 (2010) and its higher versions compute percentiles that are comparable to the percentiles computed by Excel 2003 and higher versions.

- The literature search suggests that there are a total of nine (9) known types of percentiles, i.e., 9 different methods of calculating percentiles in statistics literature (Hyndman and Fan, 1996). The R programming language (R Core Team, 2012) has all of these 9 types which can be computed using the following statement in R

  quantile(x, p, type=k) where p = percentile, k = integer between 1 - 9

  ProUCL computes percentiles using Type 7; Minitab 16 and SPSS compute percentiles using Type 6. It is simply a matter of choice, as there is no 'best' type to use. Many software packages use one type for calculating a percentile, and another for a box plot (Hyndman and Fan, 1996).

- An incorrect statement "*By definition, the sample mean has a 50% chance of being below the true population mean"* has been made in Helsel and Gilroy (2012) and also in Helsel (2012a). The above statement is not correct for means of skewed distributions (e.g., lognormal or gamma) commonly occurring in environmental applications. Since Helsel (2012) prefers to use a lognormal distribution, the incorrectness of the above statement has been illustrated using a lognormal distribution. The mean and median of a lognormal distribution (details in Section 2.3.2 of Chapter 2 of ProUCL Tech Guide) are given by:

  $$\text{mean} = \mu_1 = \exp(\mu + 0.5\sigma^2); \text{ and median} = M = \exp(\mu)$$

  From the above equations, it is clear that the mean of a lognormal distribution is always greater than the median for all positive values of σ (*sd* of log-transformed variable). Actually the mean is greater than the p^{th} percentile when $\sigma > 2z_p$. For example, when $p = 0.80$, $z_p = 0.845$, and mean of a lognormal distribution, μ_1 exceeds $x_{0.80}$, the 80th percentile when $\sigma > 1.69$. In other words, when $\sigma > 1.69$ the lognormal mean will exceed the 80th percentile of a lognormal distribution.

To demonstrate the incorrectness of the above statement, a small simulation study was conducted. The distribution of sample means based upon samples of size 100 were generated from lognormal distributions with $\mu = 4$, and varying skewness. The experiment was performed 10,000 times to generate the distributions of sample means. The probabilities of sample means less than the population means were computed. The following results are noted.

Table 1-2. Probabilities $p(\bar{x} < \mu_1)$ Computed for Lognormal Distributions with μ=4 and Varying Values of σ

Results are based upon 10000 Simulation Runs for Each Lognormal Distribution Considered

Parameter	**$\mu=4$, $\sigma=0.5$ $\mu_1=61.86$ $\sigma_1=32.97$**	**$\mu=4$, $\sigma=1$ $\mu_1=90.017$ $\sigma_1=117.997$**	**$\mu=4$, $\sigma=1.5$ $\mu_1=168.17$ $\sigma_1=489.95$**	**$\mu=4$, $\sigma=2$ $\mu_1=403.43$ $\sigma_1=2953.53$**	**$\mu=4$, $\sigma=2.5$ $\mu_1=1242.65$, $\sigma_1=28255.23$**
$p(\bar{x} < \mu_1)$	**0.519**	**0.537**	**0.571**	**0.651**	**0.729**
Mean	61.835	89.847	168.70	405.657	1193.67
Median	61.723	89.003	160.81	344.44	832.189

The probabilities summarized in the above table demonstrate that the statement about the mean made in Helsel and Gilroy (2012) is incorrect.

- Graphical Methods: Graphical methods are available in ProUCL as exploratory tools which can be generated for both uncensored and left-censored data sets. The Unofficial ProUCL Guide makes several comments about Box plots and Q-Q plots incorporated in ProUCL. The Unofficial ProUCL Guide states that all graphs with NDs are incorrect. These statements are misleading and incorrect. The intent of the graphical methods in ProUCL is exploratory to gain information (e.g., outliers, multiple populations, data distribution, patterns, and skewness) present in a data set. Based upon the data displayed (ProUCL displays a message [e.g., as a sub-title] in this regard) on those graphs, all statistics shown on those graphs generated by ProUCL are correct.

- Box Plots: 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. All box plot methods including the one in ProUCL convey the same information about the data set (outliers, mean, median, symmetry, skewness). ProUCL uses a couple of development tools such as FarPoint spread (for Excel type input and output operations) and ChartFx (for graphical displays); and ProUCL generates box plots using the built-in box plot feature in ChartFx. For all practical and exploratory purposes, box plots in ProUCL are equally good (if not better) as available in the various commercial software packages to get an idea about the data distribution (skewed or symmetric), to identify outliers, and to compare multiple groups (main objectives of box plots in ProUCL).

  - As mentioned earlier, it is a matter of choice of using percentiles/quartiles to construct a box plot. There is no 'best' method to construct a box plot. Many software packages use one method (e.g., out of 9 described above) for calculating a percentile, and another for constructing a box plot (Hyndman and Fan, 1996).

- Q-Q plots: All Q-Q plots incorporated in ProUCL are correct and of high quality. In addition to identify outliers, Q-Q plots are also used to assess data distributions. Multiple Q-Q plots are useful to perform point-by-point comparisons of grouped data sets unlike box plots based upon the five point summary statistics. ProUCL has Q-Q plots for normal, lognormal, and gamma distributions - not all of these graphical capabilities are directly available in other software packages such as NADA for R (Helsel, 2013). ProUCL offers several exploratory options to generate Q-Q plots for data sets with NDs. Only detected outlying observations may require additional investigation; therefore, from an exploratory point of view, ProUCL can generate Q-Q plots excluding all NDs (and other options). Under this scenario there is no need to retain place holders (computing plotting positions used to impute NDs) as the objective is not to impute NDs. To impute NDs, ProUCL uses ROS methods (Gamma ROS and log ROS) requiring place holders; and ProUCL computes plotting positions for all detects and NDs to generate a proper regression model which is used to impute NDs. Also for comparison purposes, ProUCL can be used to generate Q-Q plots on data sets obtained by replacing NDs by their respective DLs or DL/2s. In these cases also, no NDs are imputed, and there is no need to retain placeholders for NDs. On these Q-Q plots, ProUCL displays some relevant statistics which are computed based upon the data displayed on those graphs.

- Helsel (2012a) states that the Summary Statistics module does not display KM estimates and that statistics based upon logged data are useless. Typically, estimates computed after processing the data do not represent summary statistics. Therefore, KM and ROS estimates are not displayed in Summary Statistics module. These statistics are available in several other modules including the UCL and BTV modules. At the request of several users, summary statistics are computed based upon logged data. It is believed that mean, median, or standard deviation of logged data do provide useful information about data skewness and data variability.

- To test for the equality of variances the F-test, as incorporated in ProUCL, performs fairly well and the inclusion of the Levene's (1960) test will not add any new capability to the ProUCL software. Therefore, taking budget constraints into consideration, Levene's test has not been incorporated in the ProUCL software.

  - However, although it makes sense to first determine if the two variances are equal or unequal; this is not a requirement to perform a t-test. The t-distribution based confidence interval or test for μ_1 - μ_2 based on the pooled sample variance does not perform better than the approximate confidence intervals based upon Satterthwaite's test. Hence testing for the equality of variances is not <u>required</u> to perform a two-sample t-test. The use of Welch-Satterthwaite's or Cochran's method is recommended in all situations (see, for example, F. Hayes [2005]).

- Helsel (2012a) suggested that imputed NDs should not be made available to the users. The developers of ProUCL and other researchers like to have access to imputed NDs. As a researcher, for <u>exploratory purposes</u>, one may want to have access to imputed NDs to be used by exploratory advanced methods such as multivariate methods including data mining and principal component analysis. It is noted that one cannot easily perform exploratory methods on multivariate data sets with NDs. The availability of imputed NDs makes it possible for researchers to use data mining exploratory methods on multivariate data sets with NDs.

  - The statements summarized above should not be misinterpreted. One may not use parametric hypothesis tests such as a t-test or a classical ANOVA on data sets consisting of imputed NDs. These methods require further investigation as the decision errors associated with such methods remain unquantified. There are other methods such as Gehan and Tarone-Ware tests in ProUCL5.0 which are better suited for data sets with multiple detection limits.

- Outliers: Helsel (2012a) and Helsel and Gilroy (2012) make several comments about outliers. The philosophy (with input from EPA scientists) of the developers of ProUCL about the outliers in environmental applications is that those outliers (unless they represent typographical errors) may potentially represent impacted (site related or otherwise) locations or monitoring wells; and therefore may require further investigation.

    - The presence of outliers in a data set tends to destroy the normality of the data set. In other words, a data set with outliers can seldom (may be when outliers are mild lying around the border of the central and tail part of a normal distribution) follow a normal distribution. There are modern robust and resistant outlier identification methods (e.g., Rousseeuw and Leroy, 1987; Singh and Nocerino, 1995) which are better suited to identify outliers present in a data set; several of those robust outlier identification methods are available in the Scout 2008 version 1.0 (EPA 2009) software package.

    - For both Rosner and Dixon tests, it is the data set (also called the main body of the data set) obtained after removing the outliers (and not the data set with outliers) that needs to follow a normal distribution. Outliers are not known in advance. ProUCL has normal Q-Q plots which can be used to get an idea about potential outliers (or mixture populations) present in a data set. However, since a lognormal model tends to accommodate outliers, a data set with outliers can follow a lognormal distribution; this does not imply that the outlier potentially representing an impacted/unusual location does not exist! In environmental applications, outlier tests should be performed on raw data sets, as the cleanup decisions need to be made based upon values in the raw scale and not in log-scale or some other transformed space. More discussion about outliers can be found in Chapter 7 of the ProUCL Technical Guide.

- In Helsel (2012a), it is stated, "Mathematically, the lognormal is simpler and easier to interpret than the gamma (opinion)." We do agree with the opinion that the lognormal is simpler and easier to use but the log-transformation is often misunderstood and hence incorrectly used and interpreted. Numerous examples (e.g., Example 2-1 and 2-2, Chapter 2 of ProUCL Technical Guide) are provided in the ProUCL guidance documents illustrating the advantages of the using a gamma distribution.

- It is further stated in Helsel (2012 a) that ProUCL prefers the gamma distribution because it downplays outliers as compared to the lognormal. This argument can be turned around - in other words, one can say that the lognormal is preferred by practitioners who want to inflate the effect of the outlier. Setting this argument aside, we prefer the gamma distribution as it does not transform the variable so the results are in the same scale as the collected data set. As mentioned earlier, log-transformation does appear to be simpler but problems arise when practitioners are not aware of the pitfalls (e.g., Singh and Ananda, 2002; Singh, Singh, and Iaci, 2002).

- Helsel (2012a) and Helsel and Gilroy (2012) state that "lognormal and gamma are similar, so usually if one is considered possible, so is the other." This is an incorrect and misleading statement. There are significant differences in the two distributions and in their mathematical properties. Based upon the extensive experience in environmental statistics and published literature, for skewed data sets that follow both lognormal and gamma distributions, the developers do favor the use of the gamma distribution over the lognormal distribution. The use of the gamma distribution based decision statistics is preferred to estimate the environmental parameters (mean, upper percentile). A lognormal model tends to hide contamination by accommodating outliers and multiple populations whereas a gamma distribution tends not to accommodate contamination as can be seen in Examples 2-1and 2-2 of Chapter 2 of ProUCL Technical Guide. The use of the lognormal distribution on a data set with outliers tends to yield inflated and distorted estimates which may not be protective of human health and the environment; this is especially true for skewed data sets of small of sizes <20-30.

- In the context of computing a UCL95 of mean, Helsel and Gilroy (2012) and Helsel (2012a) state that GROS and LROS are probably never better than KM. It should be noted that these three estimation methods compute estimates of mean and standard deviation and not the upper limits used to estimate EPC terms and BTVs. The use of KM method does yield good estimates of mean and standard deviation as noted by Singh, Maichle, and Lee (2006). Computing good estimates of mean and *sd* based upon left-censored data sets addresses only half of the problem. The main issue is to compute decision statistics (UCL, UPL, UTL) which account for uncertainty and data skewness inherently present in environmental data sets.

- Realizing that for skewed data sets, Student's t-UCL, CLT-UCL, and standard and percentile bootstrap UCLs do not provide the specified coverage to the population mean, for uncensored data sets researchers (e.g., Johnson (1978), Chen (1995), Efron and Tibshirani (1993), Hall [1988, 1992], Grice and Bain (1980), Singh, Singh, and Engelhardt (1997), Singh, Singh, and Iaci (2002)) have developed parametric (e.g., gamma distribution) and nonparametric (e.g., bootstrap-t and Hall's bootstrap method, modified-t, and Chebyshev inequality) methods to compute confidence intervals and upper limits which adjust for data skewness.

- Analytically, it is not feasible to compare the various estimation and UCL computation methods for skewed data sets consisting of nondetect observations. Instead, researchers use simulation experiments to learn about the distributions and performances of the various statistics (e.g., KM-t-UCL, KM-percentile bootstrap UCL, KM-bootstrap-t UCL, KM-Gamma UCL). Based upon the suggestions made in published literature and findings summarized in Singh, Maichle, and Lee (2006), it is reasonable to state and assume that the findings of the simulation studies performed on uncensored skewed data sets to compare the performances of the various UCL computation methods can be extended to skewed left-censored data sets.

- Like uncensored skewed data sets, for left-censored data sets, ProUCL 5.0 has several parametric and nonparametric methods to compute UCLs and other limits which adjust for data skewness. Specifically, ProUCL uses KM estimates in gamma equations; in bootstrap-t method, and in Chebyshev inequality to compute upper limits for left-censored skewed data sets.

- Helsel (2012a) states that ProUCL 4 is based upon presuppositions. It is emphasized that ProUCL does not make any suppositions in advance. Due to the poor performance of a lognormal model (as demonstrated in the literature and illustrated via examples throughout the ProUCL Technical Guide), the use of a gamma distribution is preferred when a data set can be modeled by a lognormal model and a gamma model. To provide the desired coverage (as close as possible) for the population mean, in earlier versions of ProUCL (version 3.0), in lieu of H-UCL, the use of Chebyshev UCL was suggested for moderately and highly skewed data sets. In later (3.00.02 and higher) versions of ProUCL, depending upon data skewness and data distribution, for gamma distributed data sets, the use of Gamma distribution was suggested to compute the UCL of mean.

Upper limits (e.g., UCLs, UPLs, UTLs) computed using the Student's t statistic and percentile bootstrap method (Helsel, 2012, NADA for R, 2013) often fail to provide the desired coverage (e.g., 95% confidence coefficient) to the parameters (mean, percentile) of most of the skewed environmental populations. It is suggested that the practitioners compute the decision making statistics (e.g., UCLs, UTLs) by taking: data distribution; data set size; and data skewness into consideration. For uncensored and left-censored data *sets, several such upper limits computation methods have been incorporated in ProUCL 5.0 and its earlier versions.*

Contrary to the statements made in Helsel and Gilroy (2012), ProUCL software does not favor statistics which yield higher (e.g., nonparametric Chebyshev UCL) or lower (e.g., preferring the use of a gamma distribution to using a lognormal distribution) estimates of the environmental parameters (e.g., EPC and BTVs). The main objectives of the ProUCL software funded by USEPA is to compute rigorous decision statistics to help the decision makers and project teams in making correct decisions which are protective of human health and the environment.

Page 75 (Helsel [2012]): One of the reviewers of the ProUCL 5.0 software drew our attention to the following incorrect statement made on page 75 of Helsel (2012):

"If there is only 1 reporting limit, the result is that the mean is identical to a substitution of the reporting limit for censored observations."

An example left-censored data set consisting of nondetect (NDs) observations with one reporting limit of 20 illustrating this issue is described as follows.

Y	D_y
20	0
20	0
20	0
7	1
58	1
92	1
100	1
72	1
11	1
27	1

The mean and standard deviation based upon the KM and two substitution methods: DL/2 and DL are summarized as follows:

Kaplan-Meier (KM) Statistics

Mean	39.4
SD	35.56

DL Substitution method (replacing censored values by the reporting limit)

Mean	42.7
SD	34.77

DL/2 Substitution method (replacing NDs by the reporting limit)

Mean	39.7
SD	37.19

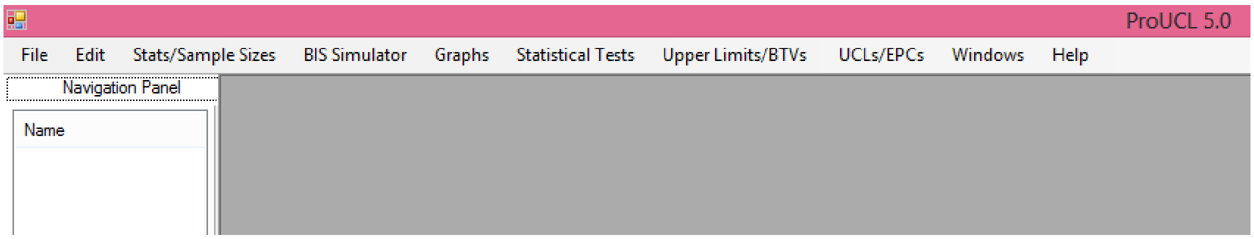
The above example illustrates that the KM mean (when only 1 detection limit is present) is not actually identical to the mean estimate obtained using the substitution, DL method. The statement made in Helsel's text holds when all observations reported as detects are greater than the single reporting limit which is seldom true in environmental data sets consisting of analytical concentrations.

Chapter 2

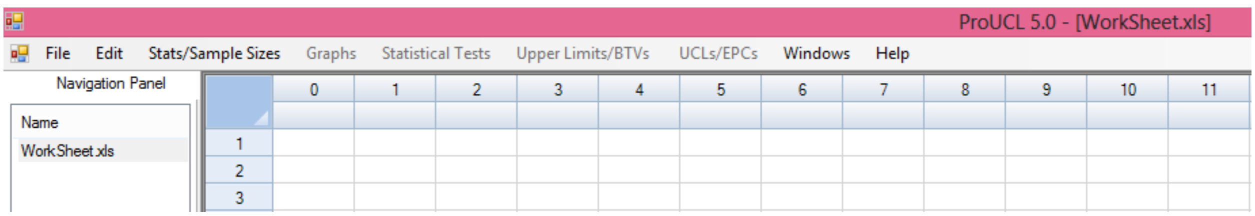
Entering and Manipulating Data

2.1 Creating a New Data Set

By executing ProUCL 5.0, the following file options will appear:

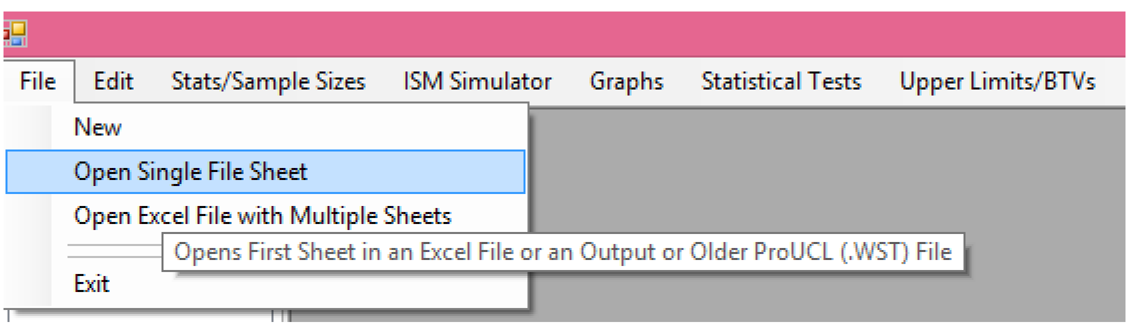


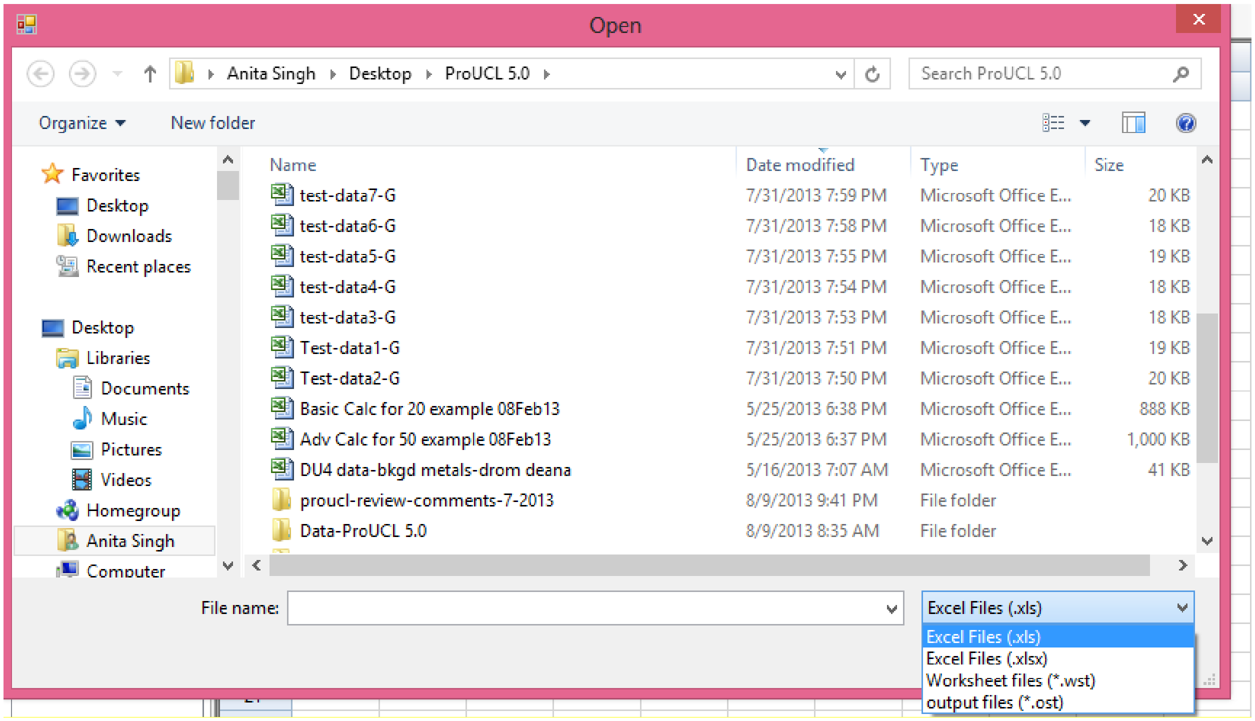
By choosing the **File ► New** option, a new worksheet shown below will appear. The user enters variable names and data following the ProUCL input file format requirements described in Section 2.3.



2.2 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 following drop down menu will appear:





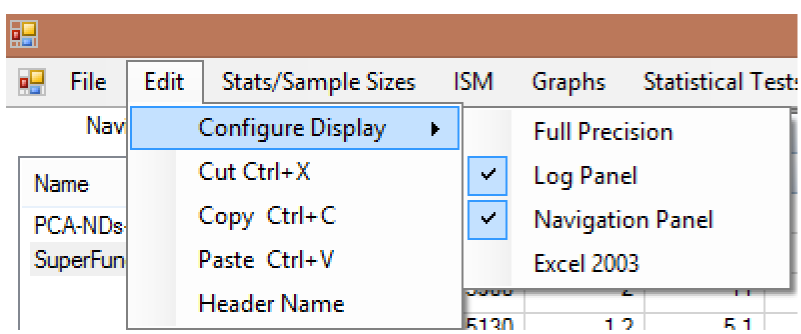
Choose a file by high lighting the type of file such as **.xls** as shown above. This option can also be used to read in *.wst worksheet and *.ost output sheet generated by earlier versions (e.g., ProUCL 4.1 and older) of ProUCL.

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 5.0

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.

2.3 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.



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

    - The Group-ID column holds 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 number of observations. Most of the modules of ProUCL can process data by a group variable.

    - Except for the header row and columns representing the group labels, only numerical values should appear in all other rows.

    - All alphanumeric strings and characters (e.g., blank, other characters, and strings), and all other values (that do not meet the requirements above) in the data file are treated as missing values and are omitted from statistical evaluations.

    - Also, a large value denoted by 1E31 (= $1x10^{31}$) can be used to represent missing data values. All entries with this value are ignored from the computations. These values are counted under the number of missing values.

2.4 Number Precision

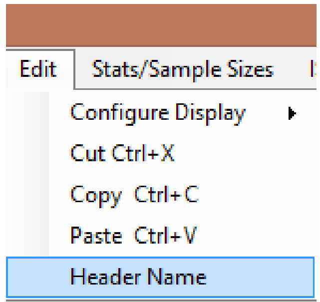
- The user may turn "Full Precision" on or off by choosing **Configure ► Full Precision On/OFF**

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

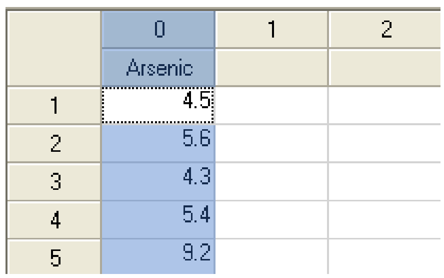
- "Full Precision" **on** option is specifically useful when one is dealing with data sets consisting of small numerical values (e.g., < 1) resulting in small values of the various estimates and test statistics. These values may become so small with several leading zeros (e.g., 0.00007332) after the decimal. In such situations, 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.*

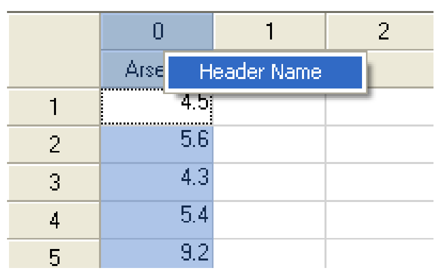
2.5 Entering and Changing a Header Name



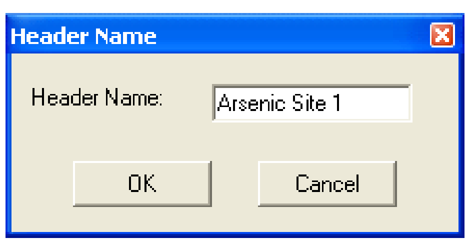
1. The user can change variable names (Header Name) using the following process. Highlight the column whose header name (variable name) you want to change by clicking either the column number or the header as shown below.



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



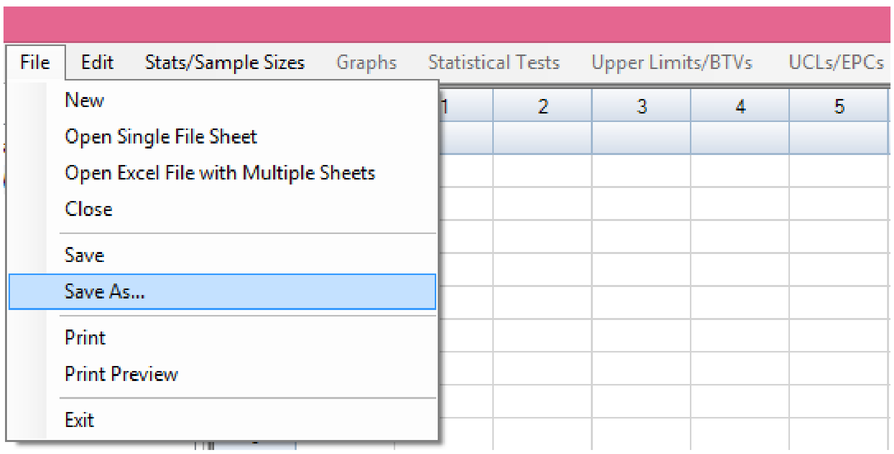
3. Change the Header Name.



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

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

2.6 Saving Files

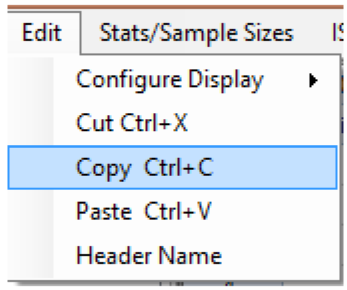


- The **Save** option allows the user to save the active window in Excel 2003 or Excel 2007.

  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.

2.7 Editing

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

2.8 Handling Nondetect Observations and Generating Files with Nondetects

- Several modules of ProUCL (e.g., Statistical Tests, Upper limits/BTVs, UCLs/EPCs) handle data sets consisting of 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 consisting of 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 observations reported as NDs; and the second column represents their detection status consisting of only 0 (for ND values) and 1 (for detected values) values. The name of the corresponding variable representing the detection status should start with d_, or D_ (not case sensitive) and the variable name. The detection status column with variable name starting with a D_ (or a d_) should have only two values: 0 for ND values, and 1 for detected observations.

- For example, the header name, D_Arsenic is used for the variable, Arsenic having ND observations. The variable D_Arsenic contains a 1 if the corresponding Arsenic value represents a detected entry, and contains a 0 if the corresponding entry represents a ND entry.

The user should follow this format otherwise the program will not recognize that your data set has NDs. An example data set illustrating these points is given as follows.

D:\example.wst

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

2.9 Caution

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

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

- It is mandatory that the header name of a nondetect column associated with a variable such as XYZ should be D_XYZ (or d_Xyz). No other characters or blanks are allowed. However, the header (column) names are not case sensitive. If the nondetect column is not labeled properly, methods to handle nondetect data will not be activated and shown.

- **Two-Sample Hypotheses:** It should be noted when using two-sample hypotheses tests (WMW test, Gehan test, and Tarone-Ware test) on data sets with NDs, both samples or variables (e.g., site-As, Back-As) should be specified as having NDs, even though one of the variables may not have any ND observations. This means that a ND column (with 0 = ND, and 1 = detect) should be provided for each variable (here D_site-As, and D_Back-As) to be used in this comparison. *If a variable (e.g., site-As) does not have any NDs, still a column with label D_site-As should be included in the data set with all entries = 1 (detected values).*

- The sample data set given on the previous page illustrates points related to this option and issues listed above. The data set contains some ND measurements for Arsenic and Mercury. It should be noted that mercury concentrations are used to illustrate the points related to ND observations; arsenic and zinc concentrations are used to illustrate the use of the group variable, Group (Surface, Subsurface).

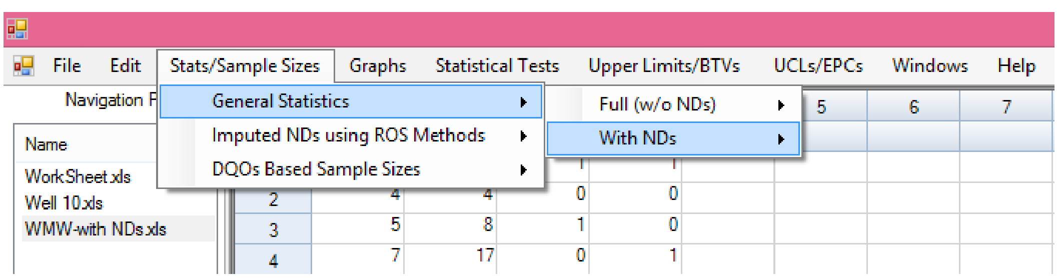
- If for mercury, one computes summary statistics (assuming no ND values) using "Full" data set option, then all ND values (with "0" entries in D_Mercury column) will be treated as detected values, and summary statistics will be computed accordingly.

2.10 Summary Statistics for Data Sets with Nondetect Observations

- To compute various statistics of interest (e.g., background statistics, GOF test, UCLs, WMW test) for variables with ND values, one should choose the ND option, **With NDs** from the various available menu options such as S**tats/Sample Sizes**, **Graphs**, **Statistical Tests**, **Upper Limits/BTVs**, and **UCLs/EPCs**.

- The NDs option of these modules gets activated only when your data set consists of NDs.

- For data sets with NDs, the **Stats/Sample Sizes** module of ProUCL 5.0 computes summary statistics and other general statistics such as the KM mean and KM standard deviation based upon raw as well as log-transformed data.



- The **General Statistics/With NDs** option also provides simple statistics (e.g., % NDs, max detect, Min detect, Mean of detected values) based upon detected values. The statistics computed in log-scale (e.g., *sd* of log-transformed detected values) may help a user to determine the degree of skewness (e.g., mild, moderate, high) of a data set based upon detected values. These statistics may also help the user to choose the most appropriate method (e.g., KM bootstrap-t UCL or KM percentile bootstrap UCL) to compute UCLs, UPLs, and other limits used to compute decision statistics.

- All other parametric and nonparametric statistics and estimates of population mean, variance, percentiles (e.g., KM, and ROS estimates) for variables with ND observations are provided in other menu options such as **Upper Limits/BTVs** and **UCLs/EPCs**.

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

- ProUCL 5.0 provides warning messages and recommendations for datasets with insufficient amount of data to calculate meaningful estimates and statistics of interest. For example, it is not desirable to compute an estimate of the EPC term based upon a discrete 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 5.0 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 (ITRC, 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 discrete data sets, the users should use at least 10 observations to compute UCLs and various other limits.

- Some examples of datasets with insufficient amount of data include datasets with less than 3 distinct observations, datasets with only one detected observation, and datasets consisting of all nondetects.

- Some of the warning messages generated by ProUCL 5.0 are shown as follows.

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			

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!			

2.12 Handling Missing Values

- The various modules (e.g., Stats, GOF, UCLs, BTVs, Regression, Trend tests) of ProUCL 5.0 can handle missing values within a data set. Appropriate messages are displayed when deemed necessary.

- All blanks, alphanumeric strings (except for group variables), or the specific large value 1e31 are considered as missing values.

- A group variable (representing two or more groups, populations, MWs) can have alphanumeric values (e.g., MW01, MW02, AOC1, AOC2, ...).

- ProUCL ignores all missing values in all statistical evaluations it performs. Missing values are therefore not treated as being part of a data set.

- Number of Valid Samples or Number of Valid Observations represents the Total Number of Observations minus the Number of Missing Values. If there are no missing values, then number of valid samples = total number of observations.

  Valid Samples = Total Number of Observations – Missing Values.

- It is important to note, however, that if a missing value not meant (e.g., a blank, or 1e31) to represent a group category is present in a “Group” variable, ProUCL 5.0 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 5.0.

- Number of missing values in Regression: The OLS module also handles number of missing values in the two columns (X and Y) representing independent (X) and dependent (Y) variables. ProUCL provides warning messages for bad data sets (e.g., all identical values) when statistics of interest cannot be computed. However, a bad/extreme data set can occur in numerous different ways, and ProUCL may not cover all of those extreme bad data sets. In such cases, ProUCL may still yield an error message. The user needs to review and fix his data set before performing regression or trend analysis again.

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

Example: The following example illustrates the notion of Valid Samples, Unique or Distinct Samples, and Missing Values. The data set also has ND values. ProUCL 5.0 computes these numbers and prints them on the UCLs and background statistics output.

x	D_x
2	1
4	1
2.3	1
1.2	0
w34	0
1.0E+031	0
	0
anm	0

34	1
23	1
0.5	0
0.5	0
2.3	1
2.3	1
2.3	1
34	1
73	1

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

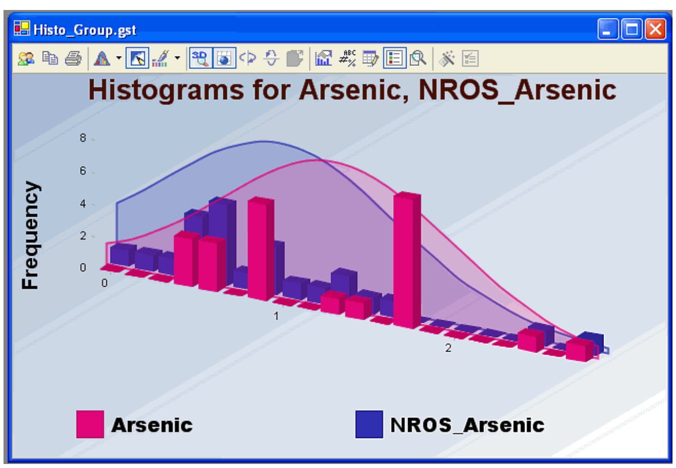
Missing Values: All values not representing a real numerical number are treated as missing values. Specifically, all alphanumeric values including blanks are considered to be missing values. Big numbers such as 1.0e31 are also treated as missing values and are considered as not valid observations.

Unique or Distinct Samples: The number of unique samples or number of distinct samples represents all unique (or distinct) detected values. Number of unique or distinct values is computed for detected values only. This number is especially useful when using bootstrap methods. As well known, it is not desirable and advisable to use bootstrap methods, when the number of unique samples is small.

2.13 User Graphic Display Modification

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

2.13.1 Graphics Tool Bar

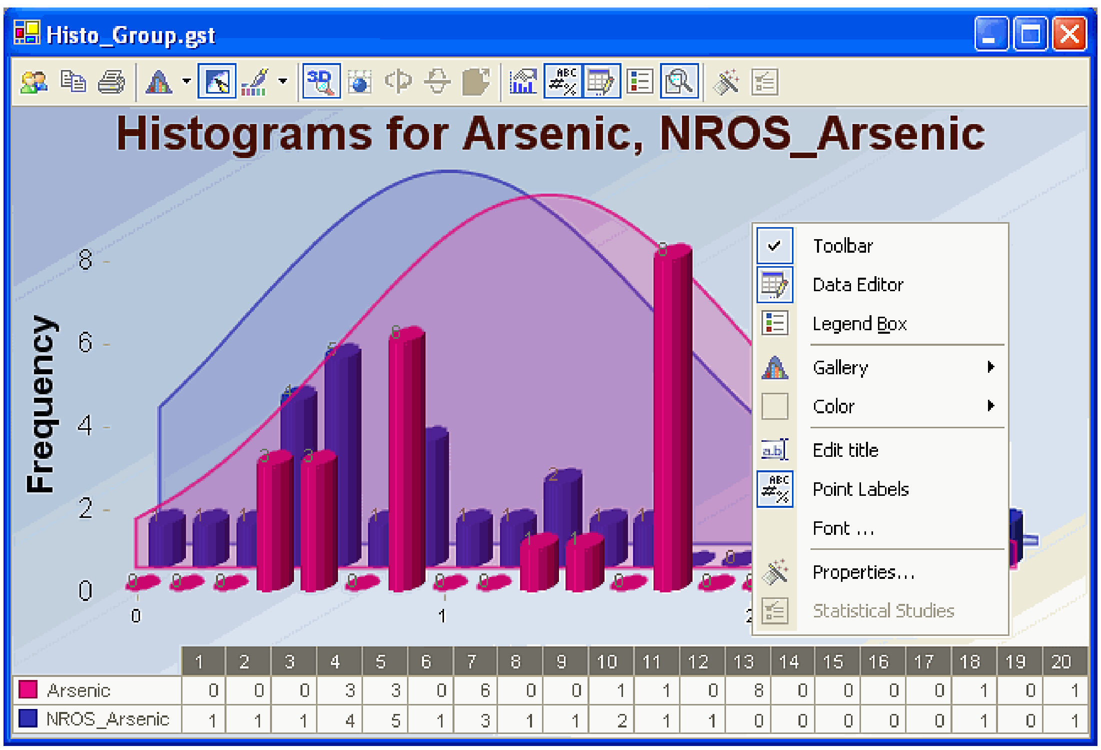


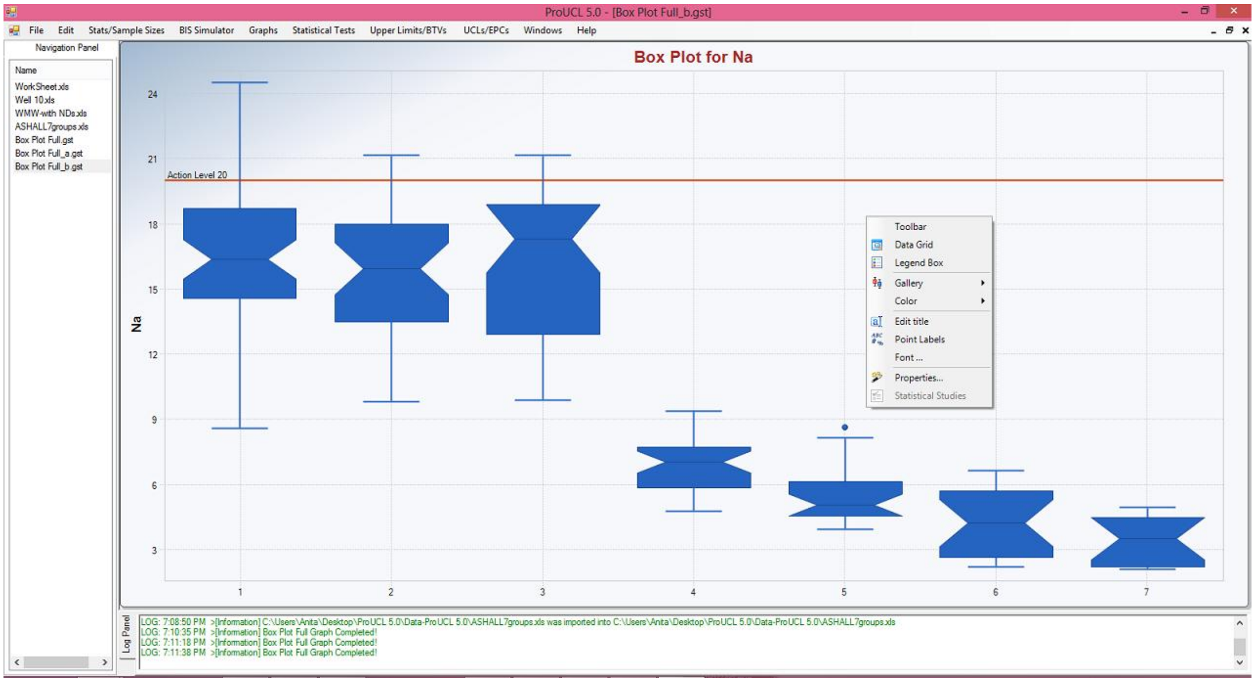
The user can change fonts, font sizes, vertical and horizontal axis's, select new colors for the various features and text. All these actions are generally used to modify the appearance of the graphic display.

The user is cautioned that these tools can be unforgiving and may put the user in a situation where the user cannot go back to the original display. Users are on their own in exploring the robustness of these tools. Therefore, less experienced users may not want to use these drop-down menu graphic tools.

2.13.2 Drop-Down Menu Graphics Tools

Graphs can be modified by using the options shown on the two graphs displayed below. These tools allow the user to move the mouse to a specific graphic item like an axis label or a display feature. The user then right-clicks their mouse and a drop-down menu will appear. This menu presents the user with available options for that particular control or graphic object. For example, the user can change colors, title name, axes labels, font size, and re-size the graphs. There is less chance of making an unrecoverable error but that risk is always present. As a cautionary note, the user can always delete the graphics window and redraw the graphical displays by repeating their operations from the datasheet and menu options available in ProUCL. A couple of examples of a drop-down menu obtained by right-clicking the mouse on the background area of the graphics display are given as follows.





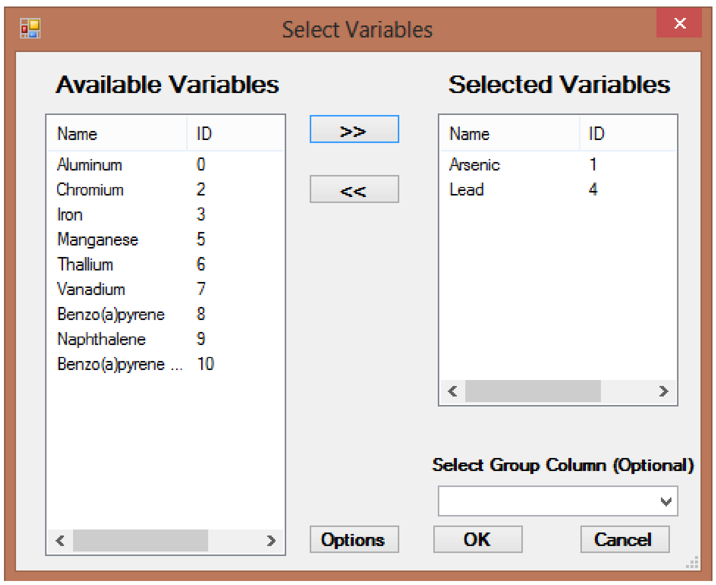
ProUCL 5.0 - [Box Plot Full_b.gst]
File
Edit
Stats/Sample Sizes
BIS Simulator
Graphs
Statistical Tests
Upper Limits/BTVs
UCLs/EPCs
Windows
Help
Navigation Panel
Name
WorkSheet.xls
Well 10.xls
WMW-with NDs.xls
ASHALL7groups.xls
Box Plot Full.gst
Box Plot Full_a.gst
Box Plot Full_b.gst
Box Plot for Na
Na
Action Level 20
Toolbar
Data Grid
Legend Box
Gallery
Color
Edit title
Point Labels
Font ...
Properties...
Statistical Studies
Log Panel

Chapter 3

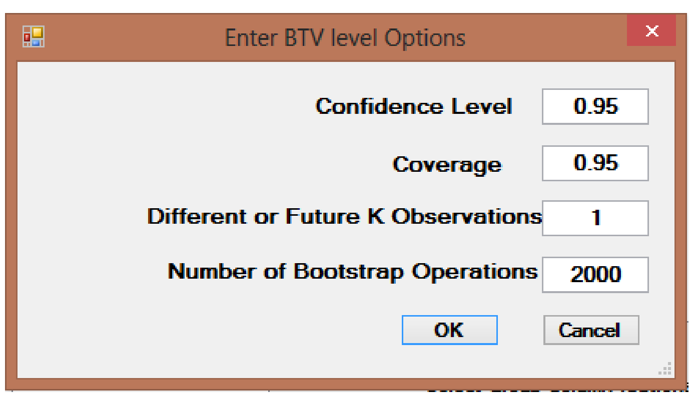
Select Variables Screen

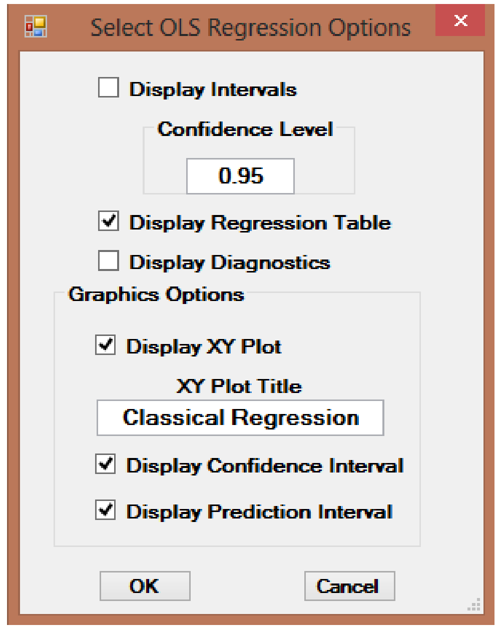
3.1 Select Variables Screen

- 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.



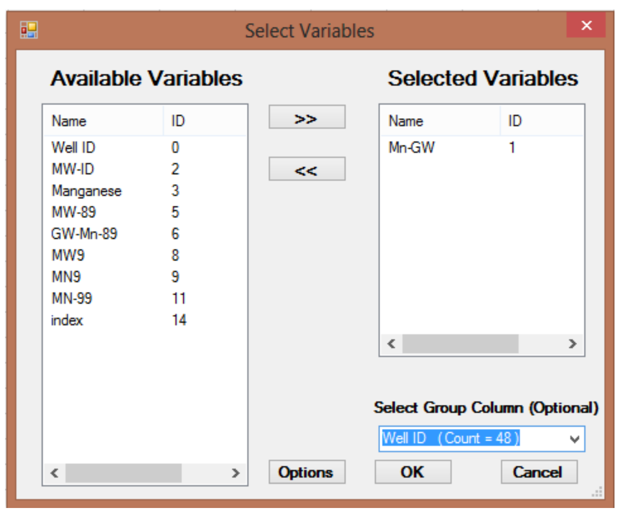
- The **Options** button is available in certain menus. The use of this option leads to another pop-up window such as shown below. This window provides various options associated with the selected statistical method (e.g., BTVs, OLS Regression).





- 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.

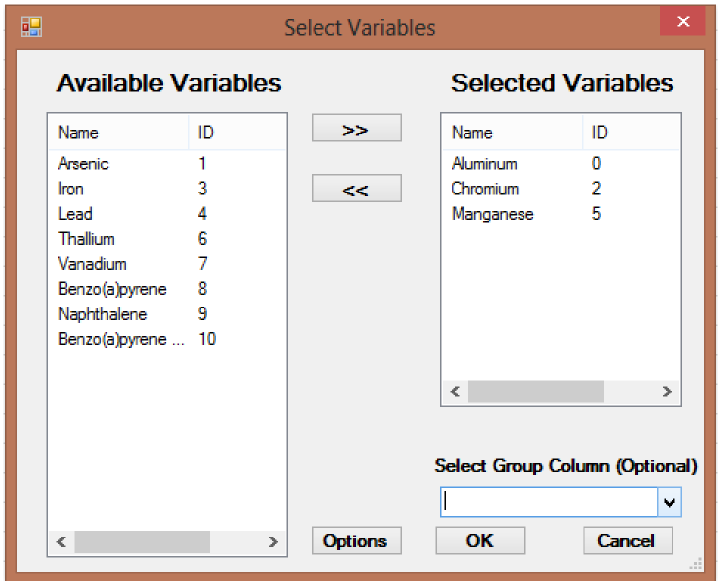
- If the user wants to perform statistical analysis on a variable (e.g., manganese) by a Group variable, click the arrow below the **Select Group Column (Optional)** to get a drop-down list of available variables from which to select an appropriate group variable. For example, a group variable (e.g., Well ID) can have alphanumeric values such as MW8, MW9, and MW1. Thus in this example, the group variable name, Well ID, takes 3 values: MW1, MW8, and MW9. 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., Well ID)



- 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 (categories) should be selected as a group variable.

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



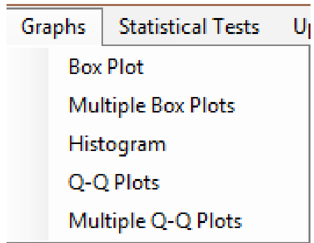
- As mentioned earlier, one should not assign any missing value such as a "Blank" for the group variable. If there is a missing value (represented by blanks, strings or 1E31) for a group variable, ProUCL will treat those missing values as a new group. As such, data values corresponding to the missing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

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

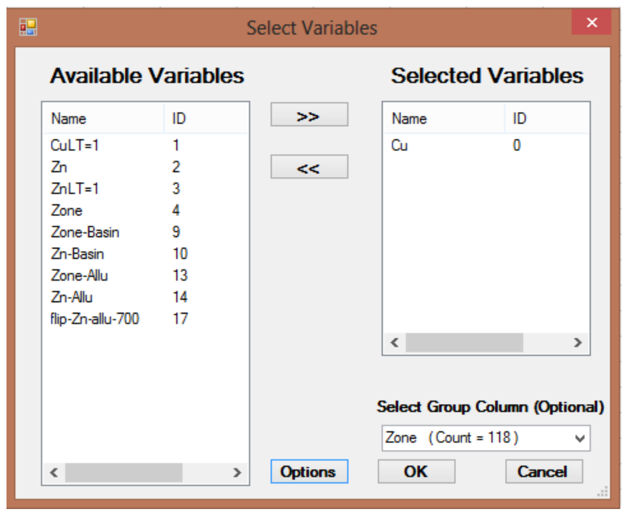
<u>Notes:</u> Once again, care should be taken to avoid misrepresentation and improper use of group variables. It is recommended not to assign any missing value for the group variable.

3.1.1 Graphs by Groups

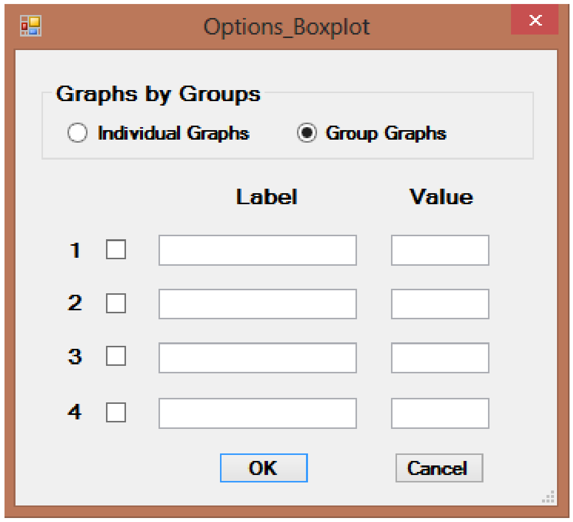
The following options are available to generate graphs by groups.



- Individual or multiple graphs (Q-Q plots, box plots, and histograms) can be displayed on a graph by selecting a "Group Column (Optional)" option shown as follows



- An individual graph for each group (specified by the selected group variable) is produced by selecting the **Individual Graph** option; and multiple graphs (e.g., side-by-side box plots, multiple Q-Q plots on the same graph) are produced by selecting the **Group Graph** option as shown below. Using the **Group Graph** option, multiple graphs are displayed for all sub-groups included in the Group variable. This option is used when data are given in the same column and are classified by a group variable.



- Multiple graphs for selected variables are produced by selecting options: **Multiple Box Plots** or **Multiple Q-Q Plots**. Using the **Group Graph** option, multiple graphs for all selected variables are shown on the same graphical display. This option is useful when data (e.g., site lead and background lead) to be compared are given in different columns.

Notes: It should be noted that it is the users' responsibility to provide adequate amount of detected data to perform the group operations. For example, if the user desires to produce a graphical Q-Q plot (using only detected data) with regression lines displayed, then there should be at least two detected points (to compute slope, intercept, and sd) in the data set. Similarly if graphs are desired for each group specified by a Group ID variable, there should be at least two detected observations in each group specified by the Group ID variable. ProUCL displays a warning message (in orange) in the lower Log Panel of the ProUCL screen when not enough data are available to perform a statistical or graphical operation.

Chapter 4

General Statistics

The "General Statistics" option is available under the Stats/Sample Sizes module of ProUCL 5.0. 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 are computed for full uncensored data sets (w/o NDs), and for data sets with nondetect (with NDs) observations (e.g., estimates based upon the KM method). Two Menu options: Full and With NDs are available.

- **Full (w/o NDs)**: This option computes various general statistics for all selected variables.

- **With NDs**: This option computes general statistics including 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:

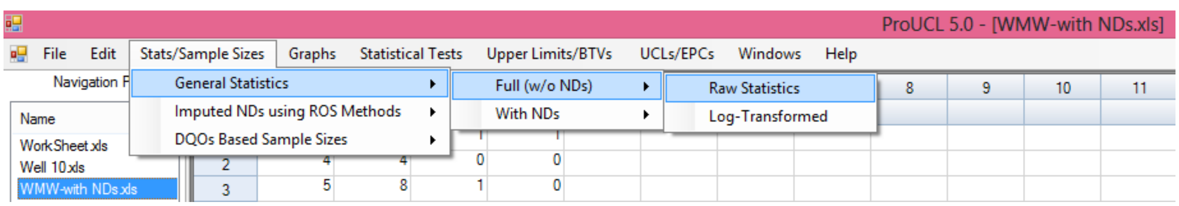
- Raw Statistics

- 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.

4.1 General Statistics for Full Data Sets without NDs

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

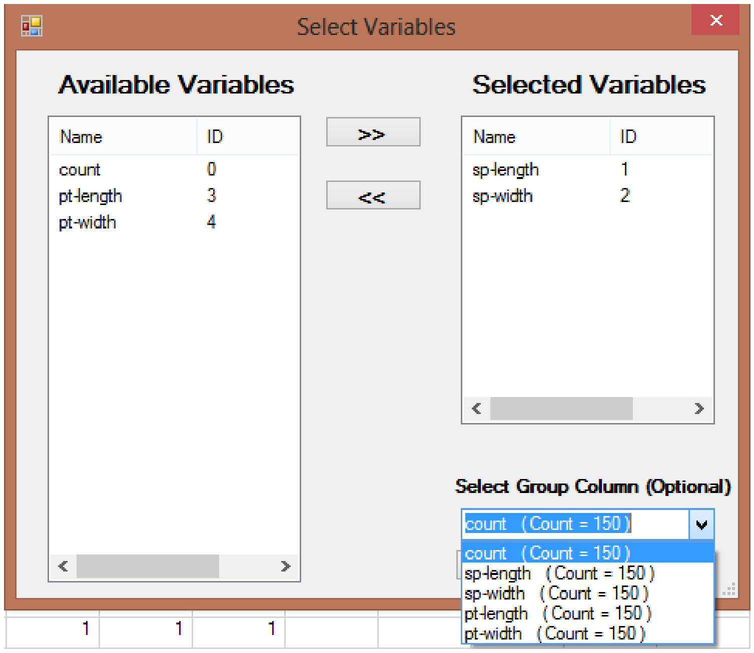


2. Select either **Log-Transformed** or **Raw Statistics** option.

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

- Select one or more variables from the **Select Variables** screen.

- If 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.



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

Raw Statistics

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

Log-Transformed Statistics

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

4. 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.

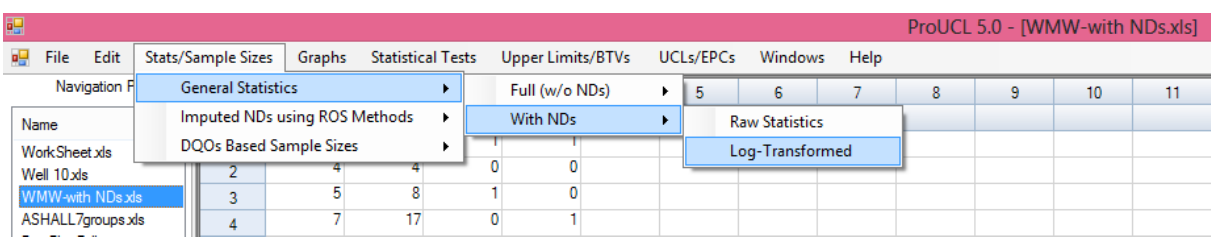
5. On the output screen shown above, most of the statistics are self explanatory and described in the ProUCL Technical Guide (EPA 2013). A couple of simple robust statistics (Hoaglin, Mosteller, and Tukey, 1983) included in the above output are described as follows.

   MAD = Median absolute deviation

   MAD/0.675 = Robust and resistant (to outliers) estimate of variability, population standard deviation, σ

4.2 General Statistics with NDs

1. As above, Click **General Statistics ► With NDs**



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

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

- Select variable(s) from the list of variables.

- Only those variables that have ND values will be shown. The user should make sure that the variables with NDs are defined properly including the column showing the detection status of the various observations.

- 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 a drop-down list of available variables. Select a proper group variable.

- Click on the **OK** button to continue or on the **Cancel** button to cancel the summary statistics operations.

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

- The **Summary Statistics** screen shown above can be saved as an Excel 2003 or 2007 file. Click **Save** from the file menu.

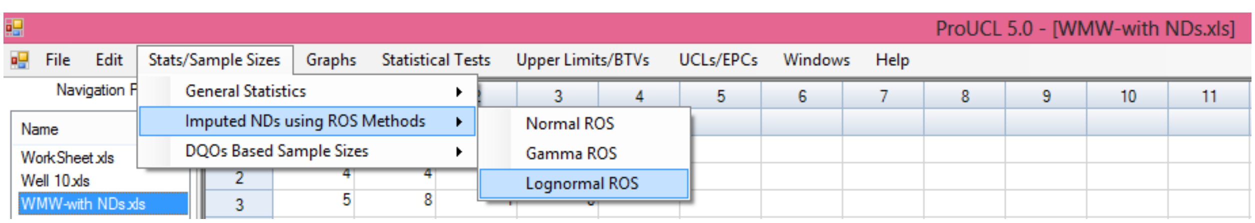
Chapter 5

Imputing Nondetects Using ROS Methods

The imputing of NDs using regression on order statistics (ROS) methods option is available under the **Stats/Sample Sizes** module of ProUCL 5.0. This option is provided for advanced users who want to use the detected and imputed NDs data for exploratory and data mining purposes on multivariate data sets. For exploratory methods such as the principal component analysis (PCA), cluster, and discriminant analysis to gain additional insight into potential structures and patterns present in a multivariate (more than one variable) data set, one may want to use imputed values in graphical displays (line graphs, scatter plots, boxplots etc.) and in exploratory PCA and cluster analysis. To derive conclusions based upon multivariate data sets consisting of nondetects, the developers suggest the use of the KM method based covariance or correlation matrix to perform principal component and regression analysis. These methods are beyond the scope of the ProUCL software which deals only with univariate methods. The details of computing an Orthogonalized Kettenring and Gnanadesikan (OKG) positive definite KM matrix can be found in Maronna, Martin, and Yohai (2006) and in Scout 2008 Version 1.0 guidance documents (2009) which can be downloaded from the EPA NERL Site. One may not use ROS imputed data to perform parametric statistical tests such as t-test and ANOVA test without further investigation. These issues require further research to evaluate decision errors associated with conclusions derived using such methods.

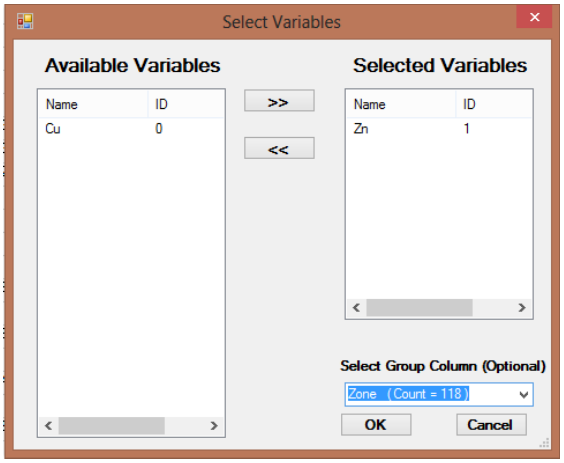
The ROS methods can be used to impute ND observations using a normal, lognormal, or gamma model. ProUCL has three ROS estimation methods that can be used to impute ND observations. The use of this option generates additional columns consisting of all imputed NDs and detected observations. These columns 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 consisting of ND observations. Like other statistical methods in ProUCL, NDs can also be imputed by a group variable. One can impute NDs using the following steps.

1. Click **Imputed NDs using ROS Methods ► Lognormal ROS**



2. The **Select Variables** screen (Chapter 3) 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.



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

Output Screen for ROS Est. NDs (Lognormal ROS) Option

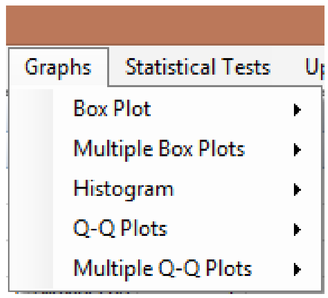
0	1	2	3	4	5	6
Cu	Zn	Zone	D_Cu	D_Zn	LnROS_Zn (alluvial fan)	LnROS_Zn (basin trough)
1	10	Alluvial Fan	0	0	2.12437794466611	20
1	9	Alluvial Fan	0	1	9	10
3		Alluvial Fan	1		1.000000E+031	60
3	5	Alluvial Fan	1	1	5	20
5	18	Alluvial Fan	1	1	18	12
1	10	Alluvial Fan	1	0	2.7045642735474	8
4	12	Alluvial Fan	1	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

<u>Notes:</u> 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.

Chapter 6

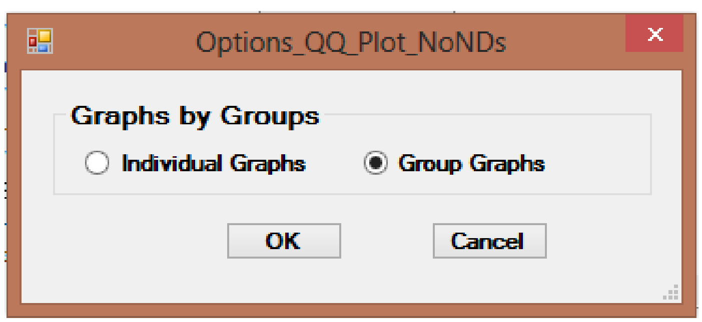
Graphical Methods (Graph)

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 5.0

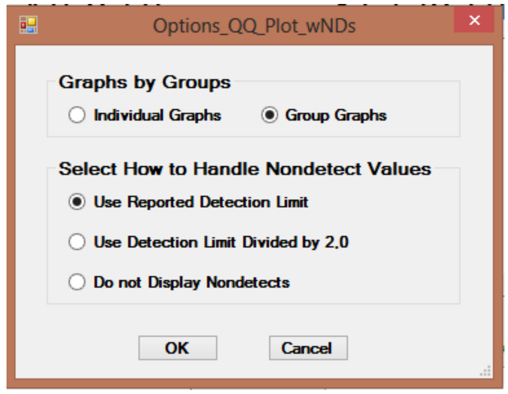


- All graphical displays listed above can be generated using uncensored full data sets (Full w/o NDs) as well as left-censored data sets with nondetect (With NDs) observations. On box plot graphs for data sets with NDs, a horizontal line is also displayed at the largest RL associated with ND observations.

- Q-Q Plots and Histograms: Q-Q plots and histograms can be generated individually as well as by using a Group variable. Graphs generated using the **Group Graphs** option shown below is useful when data for selected variable(s) are given in the same column (stacked data) categorized by a Group ID.



- For data sets with NDs, three options described below are available to draw Q-Q plots and histograms. Specifically, these graphs are displayed only for detected values, or with NDs replaced by ½ DL values, or with NDs replaced by the respective DLs. The statistics displayed on a Q-Q plot (mean, *sd*, slope, intercept) are computed according to the method used. On Q-Q plots, ND values are displayed using a smaller font. The exploratory Q-Q plots described here do not require any placeholders for NDs. These graphs are used only to determine the distribution of detected values and to identify potential outliers and/or multiple populations present in a data set. On histograms, the user can change the number of bins (more bins, less bins) used to generate histograms.



Do not Display Nondetects: 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.

Use Reported Detection Limit: Selection of this option treats DLs 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.

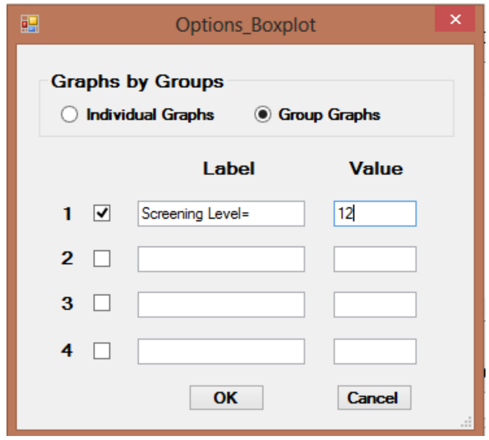
Use Detection Limit Divided by 2.0: Selection of this option replaces the DLs with their half values. All Q-Q plots and histograms are generated using the half detection limits and detected values. The statistics displayed on Q-Q plots are computed accordingly.

- For data sets in different columns, one can use the **Multiple Q-Q Plots** option. By default, this option will display multiple Q-Q plots for all selected variables on the same graph. One can also generate multiple Q-Q plots by using a group variable.

- <u>Box Plot</u>: Like Q-Q plots, box plots can also be generated by a Group variable. This option is useful when all data are given in the same column (stacked data) categorized by a Group ID variable. On box plots with NDs, a horizontal line is displayed at the largest detection limit level. ProUCL 5.0 constructs a box plot using all detected and nondetected (using associated DL values) values. A horizontal line is displayed at the largest detection limit. Box Plots are generated using ChartFx, a software used in the development of ProUCL 5.0

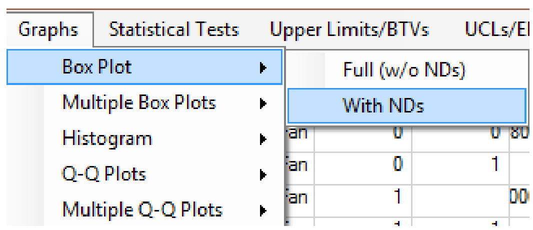
- <u>Multiple Box Plots</u>: For data in different columns, one can use the **Multiple Box Plots** option to display multiple box plots for all selected variables on the same graph. One can also generate multiple box plots by using a group variable.

- Box Plots have an optional feature, which can be used to draw up to four (4) horizontal lines at pre-established screening levels or at statistical limits (e.g., upper limits of a background data set) computed using a background data set. This option can be used when box plots are generated using onsite data and one may be interested in comparing onsite data with background threshold values and/or pre-established screening levels. This type of box plot represents a useful visual comparison of site data with background threshold values and/or other action levels. Up to four (4) values can be displayed on a box plot as shown below. If the user inputs a value in the value column, the check box in that row will get activated. For example, the user may want to display horizontal lines at a background UTL95-95 or some pre-established action level(s) on box plots generated using AOCs data.



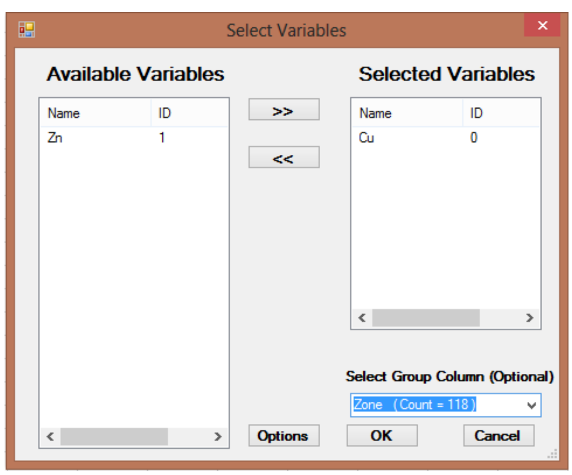
6.1 Box Plot

1. Click **Graphs ► Box Plot**

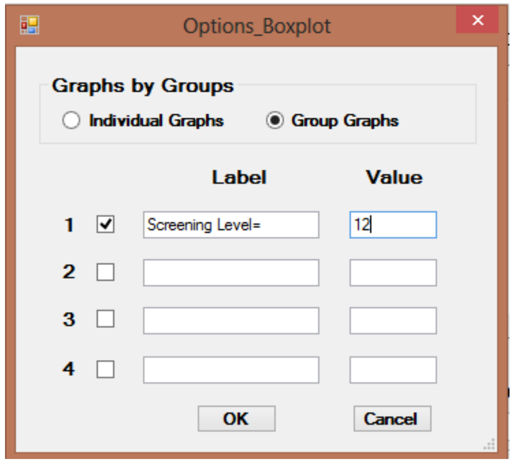


2. The **Select Variables** screen (Chapter 3) will appear.

   - Select one or more variable(s) from the **Select Variables** screen.
   - If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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.



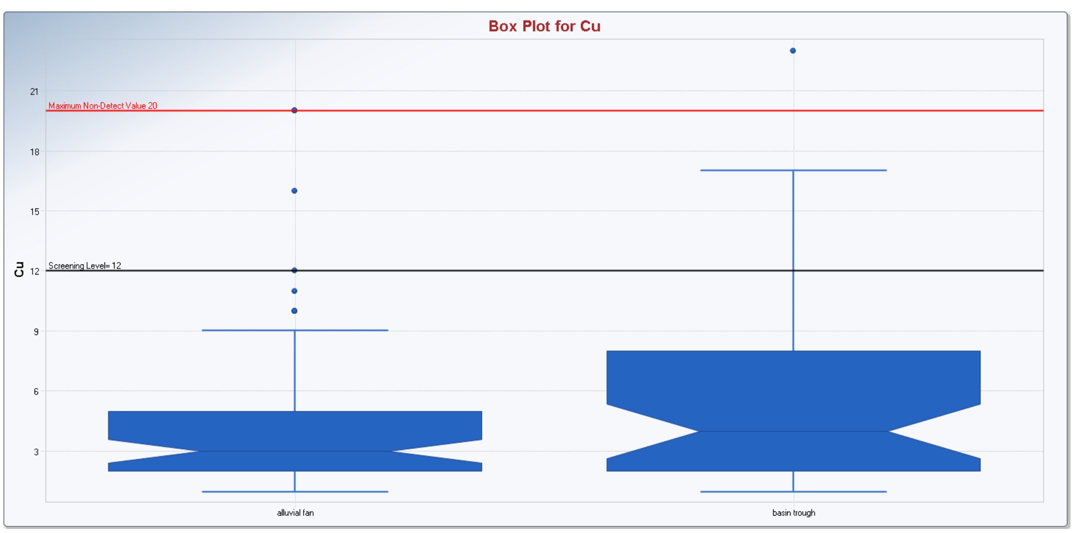
The default option for **Graph by Groups** is **Group Graphs**. This option produces side-by-side box plots for all groups included in the selected Group ID Column (e.g., Zone here). The **Group Graphs** option is used when multiple graphs categorized by a group variable need to be produced on the same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



- 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. 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.

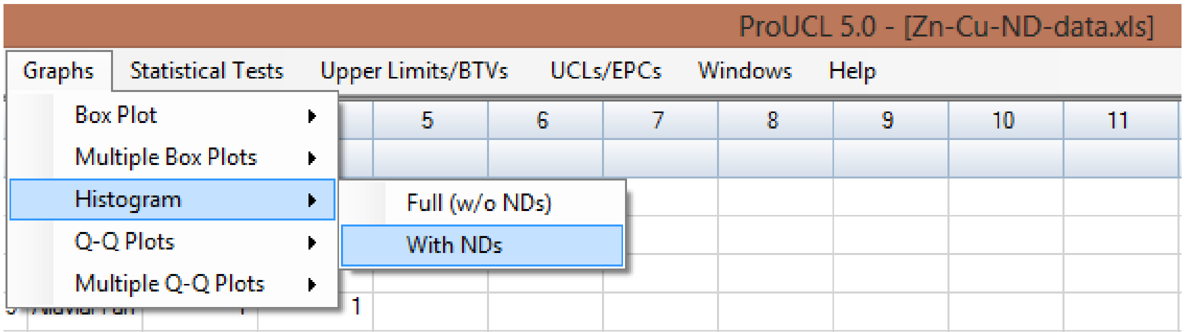
- Click on the **OK** button to continue or on the **Cancel** button to cancel the Box Plot (or other selected graphical) option.

Box Plot Output Screen (Group Graph)
Selected options: Label (Screening Level), Value (12)



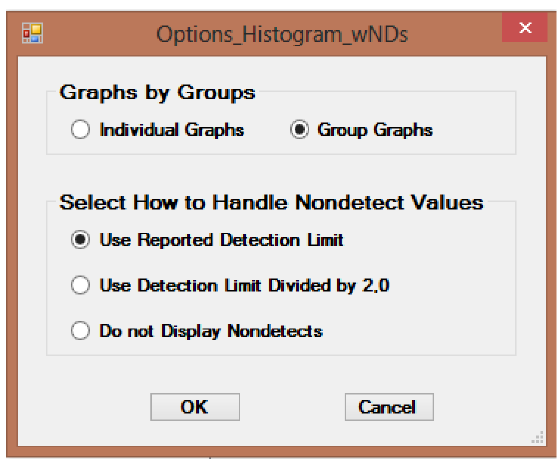
6.2 Histogram

1. Click **Graphs ► Histogram**



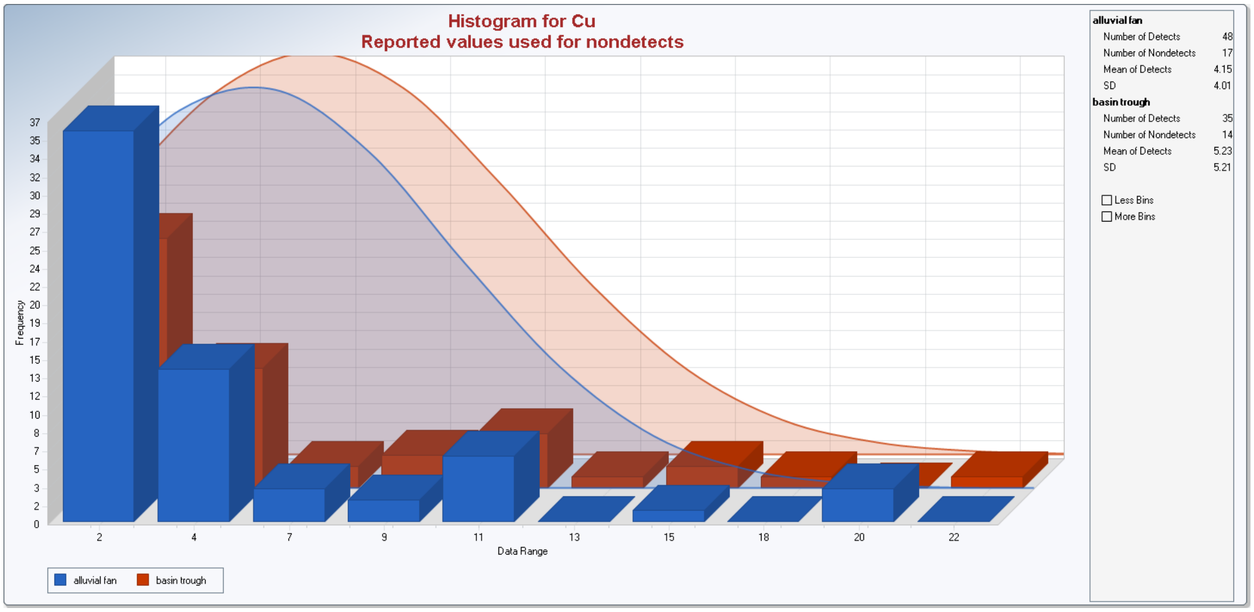
2. The **Select Variables** screen (Chapter 3) will appear.

- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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.



- 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 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.

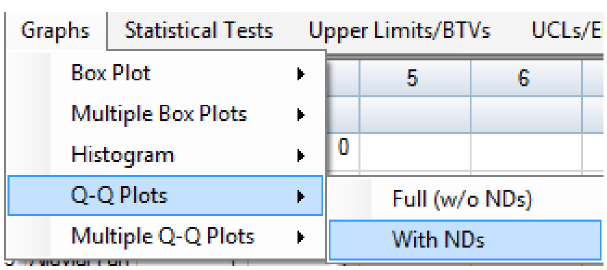
Histogram Output Screen
Selected options: Group Graphs



<u>Notes:</u> ProUCL does not perform any GOF tests when generating histograms. Histograms are generated using the development software ChartFx. Histogram option automatically generates a normal probability density function (pdf) curve irrespective of the data distribution. At this time, ProUCL 5.0 does not display a pdf curve for any other distribution (e.g., gamma) on a historgram. The user can increase or decrease the number of bins to be used in a histogram.

6.3 Q-Q Plots

1. Click **Graphs ► Q-Q Plots**. When that option button is clicked, the following window will be shown.

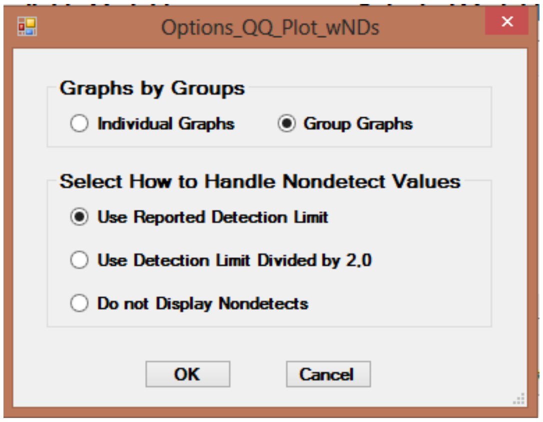


2. Q-Q Plots can be generated for data sets **With NDs** and without NDs [**Full (w/o NDs)**].

   - Select either **Full (w/o NDs)** or **With NDs** option.
   - The **Select Variables** screen (Chapter 3) will appear.
   - Select one or more variable(s) from the **Select Variables** screen.

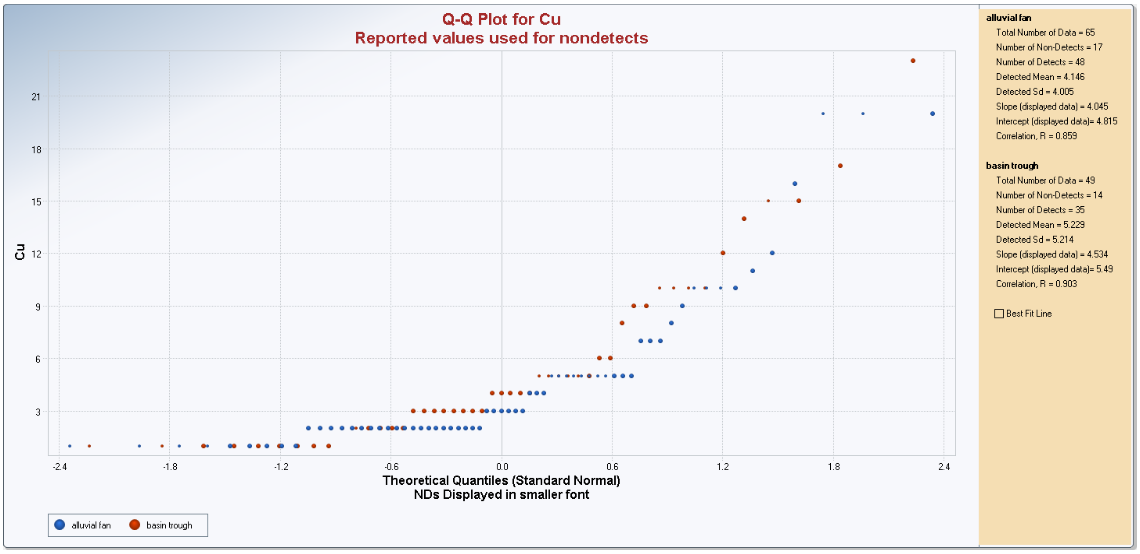
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **Select Group Column (Optional)** button. This will result in a drop-down list of available variables. The user should select and click on an appropriate variable representing a group 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 to treat NDs. The default option is to use the reported values for all NDs.



- 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.

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



Note: The font size of ND values is smaller than that of the detected values.

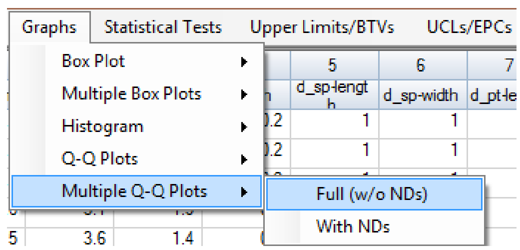
6.4 Multiple Q-Q Plots

6.4.1 Multiple Q-Q plots (Uncensored data sets)

*1. Click **Graphs** ► **Multiple Q-Q Plots***

2. Multiple Q-Q Plots can be generated for data sets **With NDs** and without NDs [**Full (w/o NDs)**].

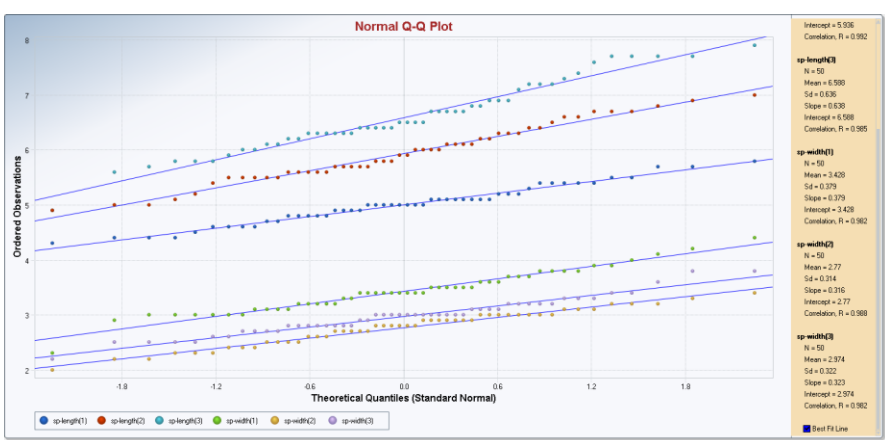
- When that **Option** button is clicked, the following window will be shown.



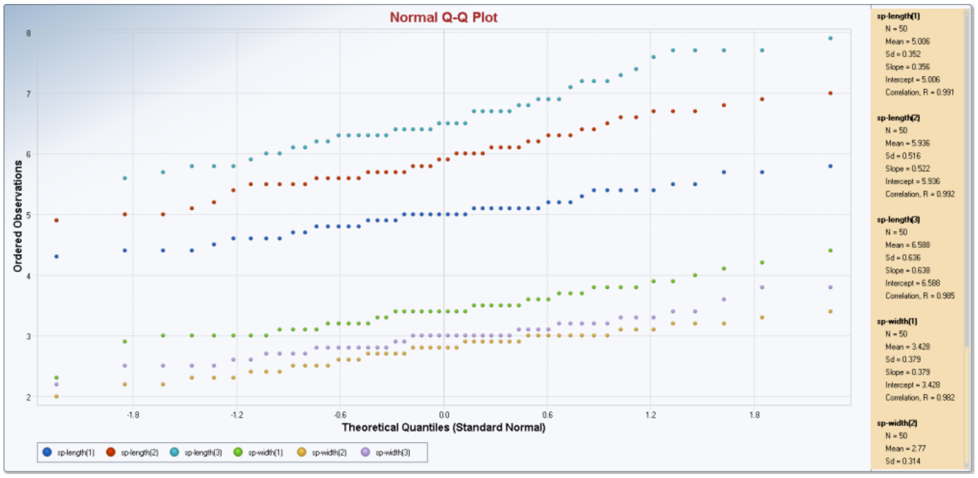
- Select either **Full (w/o NDs)** or **With NDs**.
- The **Select Variables Screen** (Chapter 3) will appear.
- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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 **OK** to continue or **Cancel** button to cancel the selected Multiple Q-Q Plots option.

Example 6-1: The following graph is generated by using Fisher's (1936) data set for 3 Iris species.

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 on the **Best Fit Line** and all regression lines will disappear as shown below.



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 **Goodness-of-Fit (GOF, G.O.F.)** test option.

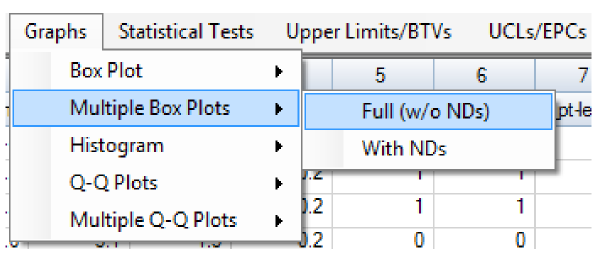
6.5 Multiple Box Plots

6.5.1 Multiple Box plots (Uncensored data sets)

1. *Click **Graphs ► Multiple Box Plots***

2. Multiple Q-Q Plots can be generated for data sets **With NDs** and without NDs [**Full (w/o NDs)]**.

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



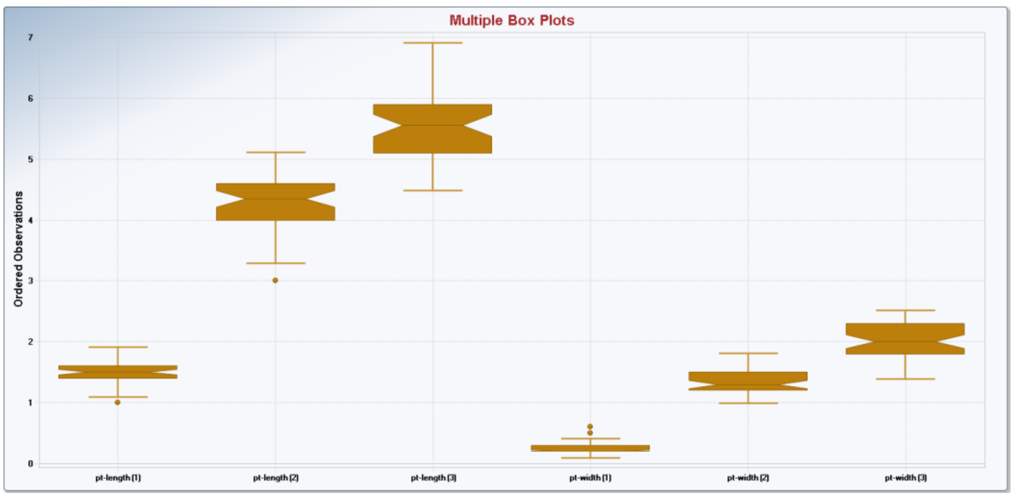
   - Select either Full (w/o NDs) or With NDs.
   - The **Select Variables** screen (Chapter 3) will appear.
   - Select one or more variable(s) from the **Select Variables** screen.

- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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 Multiple Box Plots options. The following graph is generated by using the above options.

Example 6-1 (continued): The following graph is generated by using the above options on Fisher's (1936) Iris data set collected from 3 species of Iris flower.

Output Screen for Multiple Box Plots (Full w/o NDs)
Selected options: Group Graph



Chapter 7

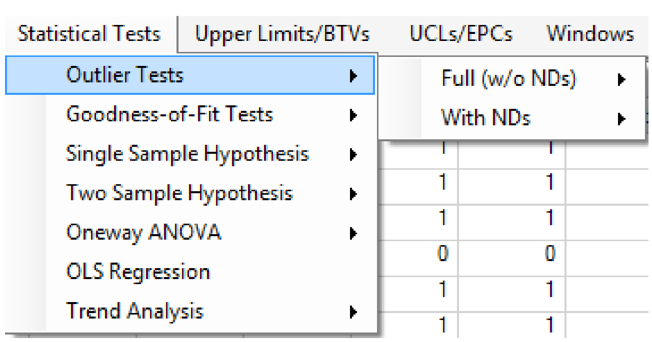
Classical Outlier Tests

Outliers are inevitable in data sets originating from environmental and various other applications. In addition to informal graphical displays (e.g., Q-Q plots and box plots) and classical outlier tests (Dixon test, Rosner test), there exist several robust outlier identification methods (e.g., biweight, Huber, PROP, MCD) to identify any number of multiple outliers potentially present in data sets of various sizes (Scout 2008; EPA 2009). It is well known that the classical outlier tests: Dixon test and Rosner test suffer from masking (e.g., extreme outliers may mask intermediate outliers) effects. The use of robust outlier identification procedures is recommended to identify multiple outliers, especially when dealing with multivariate (having multiple constituents) data sets. However, those preferred and more effective robust outlier identification methods are beyond the scope of ProUCL 5.0. Several robust outlier identification methods (e.g., based upon biweight, Huber, and PROP influence functions, Singh and Nocerino, 1995) are available in the Scout 2008 v1.0 software package (EPA, 2009).

The two classical outlier tests: Dixon and Rosner tests (EPA 2006a; Gilbert, 1987) are available in the ProUCL software. These tests can be used on data sets with and without ND observations. These tests also require the assumption of normality of the data set without the outliers. It should be noted that in environmental applications, one of the objectives is to identify high outlying observations that might be present in the right tail of a data distribution as those observations often represent contaminated locations of a polluted site potentially 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 process. To compute the various statistics of interest, ProUCL uses rigorous 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. On a normal Q-Q plot, observations that are well separated from the bulk (central part) of the data typically represent potential outliers needing further investigation. Also, significant and obvious jumps and breaks in a normal Q-Q plot are indications of the presence of more than one population. Data sets exhibiting such behavior of Q-Q plots should be partitioned out into component sub-populations before estimating EPC terms or BTVs.

Outlier tests in ProUCL 5.0 are available under the Statistical Tests module.

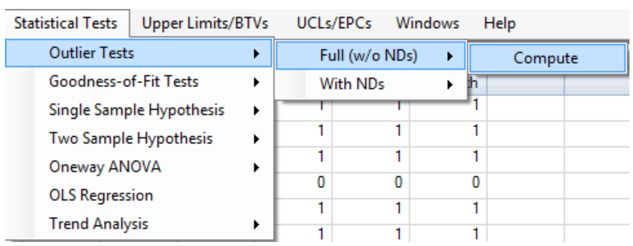


Dixon's Outlier Test (Extreme Value Test): Dixon's test is used to identify statistical outliers when the sample size is ≤ 25. This test identifies outliers or extreme values in the left tail (Case 2) and also in the right tail (Case 1) of a data distribution. In environmental data sets, outliers found 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. If the user wants to perform a normality test on the data set, he should first remove the outliers before performing the normality test. 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. Like the Dixon test, if the user wants to perform a normality test on the data set, he should first remove the outliers (which are not known in advance) before performing the normality test. 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.

7.1 Outlier Test for Full Data Set

1. Click **Outlier Tests ► Full (w/o NDs) ►Compute**

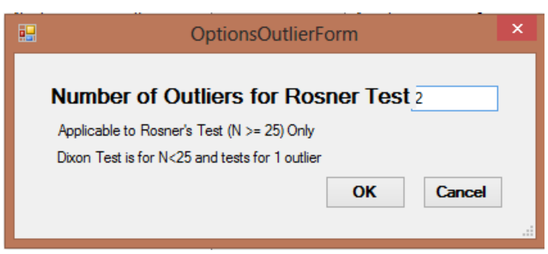


2. The **Select Variables** screen (Chapter 3) will appear.

- 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.



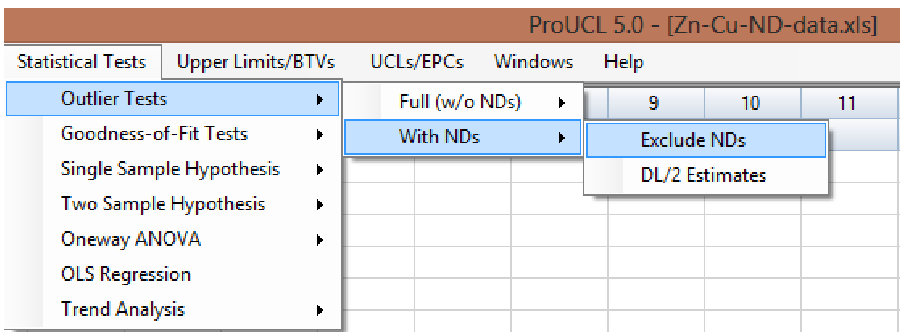
- The default option for the number of suspected outliers is 1. To use the Rosner test, the user has to obtain an initial guess about the number of suspected outliers that may be present in the data set. This can be done by using graphical displays such as a Q-Q plot. On a Q-Q plot, higher observations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

7.2 Outlier Test for Data Sets with NDs

Two options: exclude NDs; or replace NDs by their respective DL/2 are available in ProUCL to perform outlier tests on data sets with NDs.

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

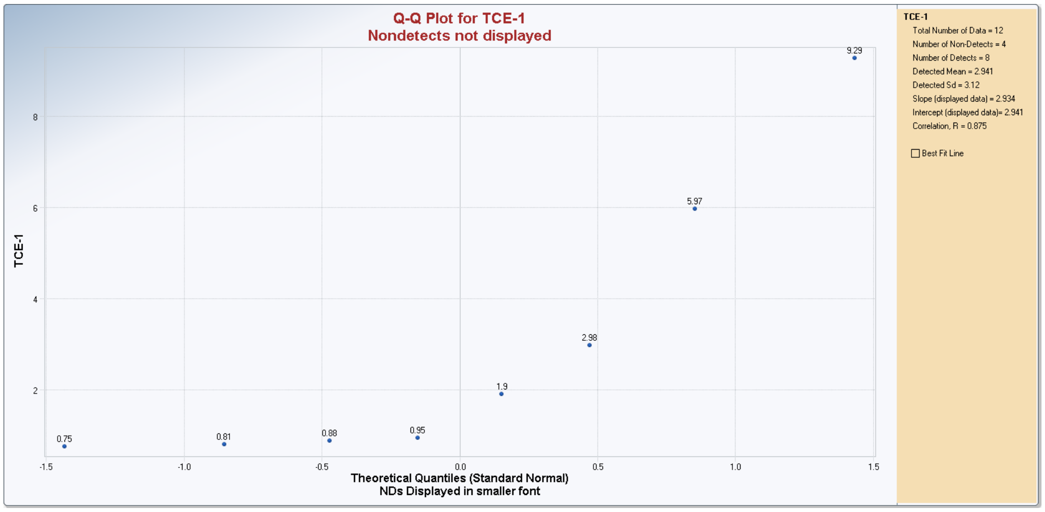


Output Screen for Dixon's Outlier Test

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.

Q-Q plot without Nondetect Observations are Shown as Follows



Example: 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)**

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

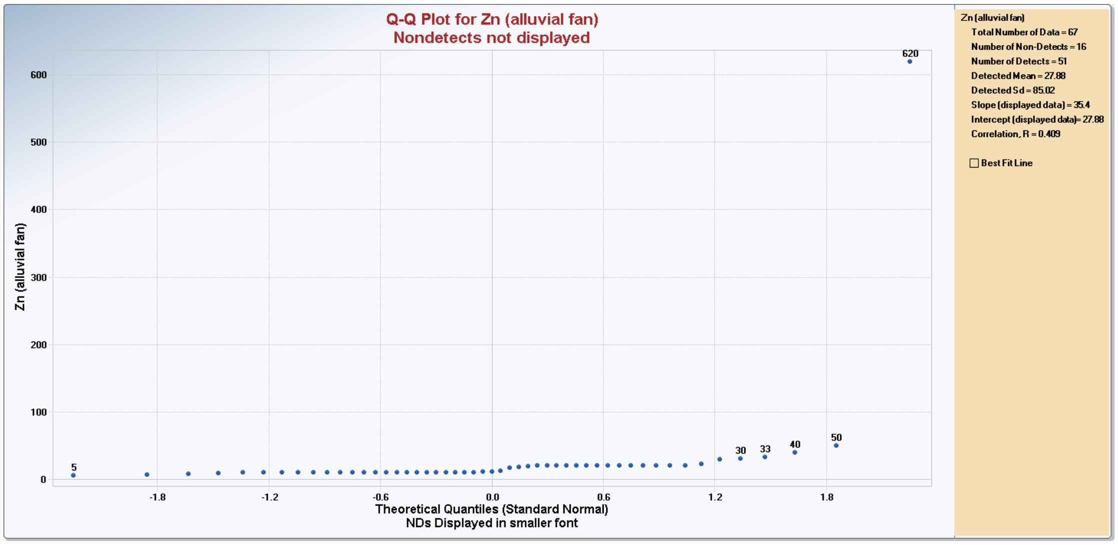
For 5% significance level, there are 3 Potential Outliers

620, 50, 40

For 1% Significance Level, there are 2 Potential Outliers

620, 50

Q-Q plot for Zinc Based upon Detected Data (Alluvial Fan)



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

Chapter 8

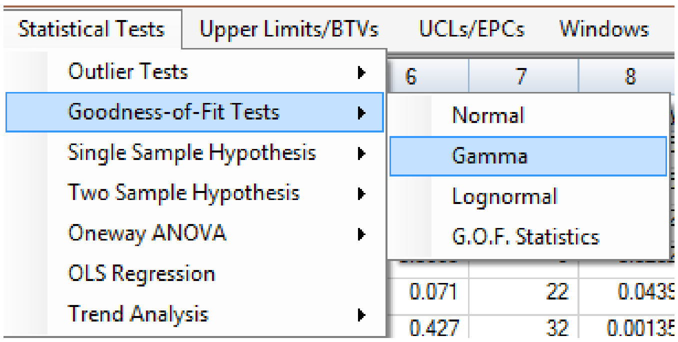
Goodness-of-Fit (GOF) Tests for Uncensored and Left-Censored Data Sets

The GOF tests are available under the **Statistical Test** module of ProUCL 5.0. Throughout this User Guide and in ProUCL 5.0 software, “Full” represents uncensored data sets without ND observations. The details and usage of the various GOF tests are described in the associated ProUCL 5.0 Technical Guide.

8.1 Goodness-of-Fit test in ProUCL

Several GOF tests for uncensored full (Full (w/o NDs)) and left-censored (With NDs) data sets are available in the ProUCL software.

- Full (w/o 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.

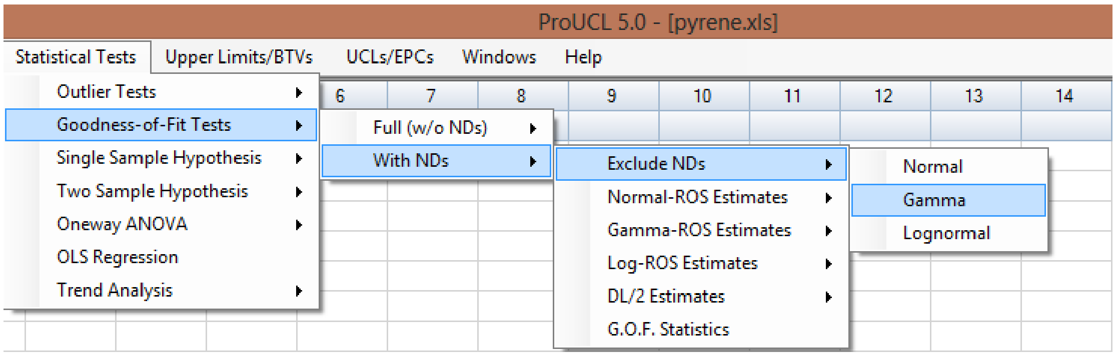
  - Based upon the hypothesized distribution (normal, gamma, lognormal), a Q-Q plot displaying all statistics of interest including the derived conclusion is also generated.

  - The GOF 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.

- **With NDs**

    - This option performs GOF tests on data sets consisting of both nondetected and detected data values.

    - Several sub-menu items shown below are available for this option.



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

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

    - 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 nondetects.

3. **DL/2 Estimates**: tests for normal, gamma, or lognormal distribution of the selected variable(s) using the detected values and the ND values replaced by their respective DL/2 values. This option is included for historical reasons and also for curious users. ProUCL does not make any recommendations based upon this option.

4. **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 nondetects. 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:

    - **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.

- **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 this 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) is an indication of a good fit for that distribution. This high correlation should exhibit a definite linear pattern in the Q-Q plot without abrupt jumps.

  - On a GOF Q-Q plot, observations that are well separated from the majority of the data (central part) 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 should be partitioned out into component sub-populations before estimating EPC terms 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 5.0, the SW test uses the exact SW critical values for samples of size 50 or less. For samples of size, greater than 50, the SW test statistic is displayed along with the *p*-value of the test (Royston, 1982, 1982a).

<u>Notes:</u> As with other statistical tests, sometimes these two tests might lead to different conclusions. The user is advised to exercise caution when interpreting these test results.

- **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.

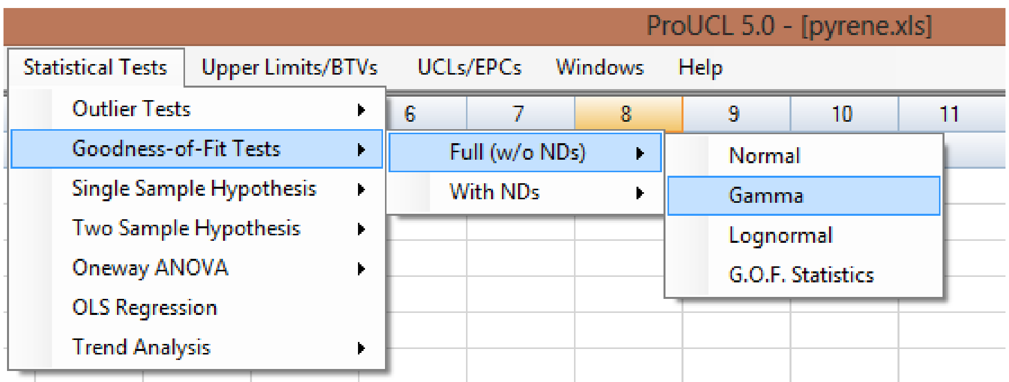
- It is noted that these two tests might lead to different conclusions. Therefore, the user should exercise caution interpreting the results.

- These two tests may be used for samples of sizes in the range of 4-2500. 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 beyond these sample sizes and values of k is not recommended.

Notes: 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.

8.2 Goodness-of-Fit Tests for Uncensored Full Data Sets

1. Click **Goodness-of-Fit Tests ► Full (w/o NDs)**

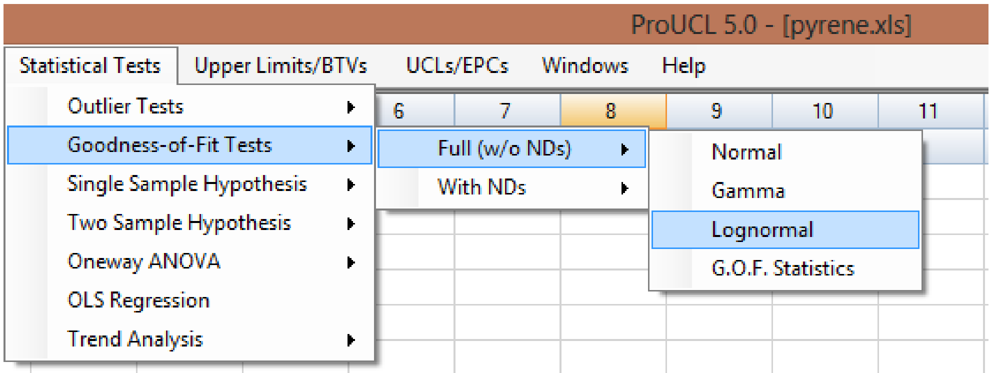


2. Select the distribution to be tested: Normal, Lognormal, or Gamma

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

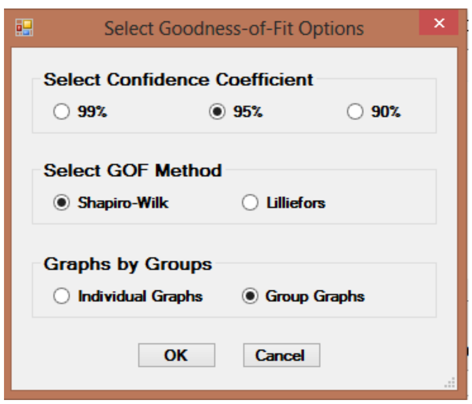
8.2.1 GOF Tests for Normal and Lognormal Distribution

1. Click **Goodness-of-Fit Tests ► Full (w/o NDs) ► Normal or Lognormal**



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

- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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.

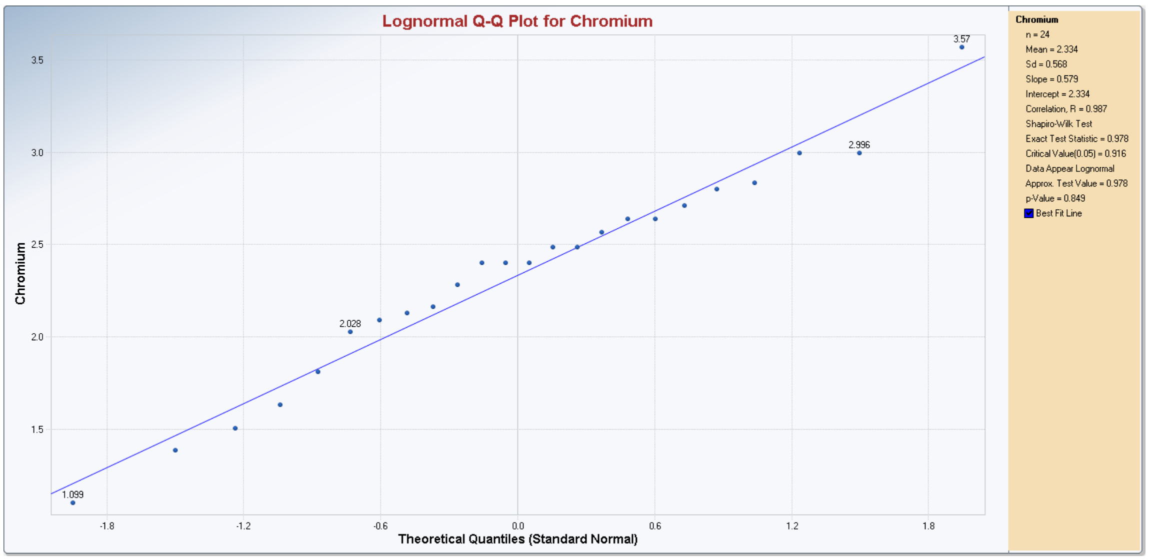


  - The default option for the **Confidence Level** is **95%**.
  - The default GOF Method is **Shapiro-Wilk**.
  - The default option for **Graphs by Group** is **Group Graphs**. If you want to see the plots for all selected variables individually, and then check the button next to **Individual Graphs**.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

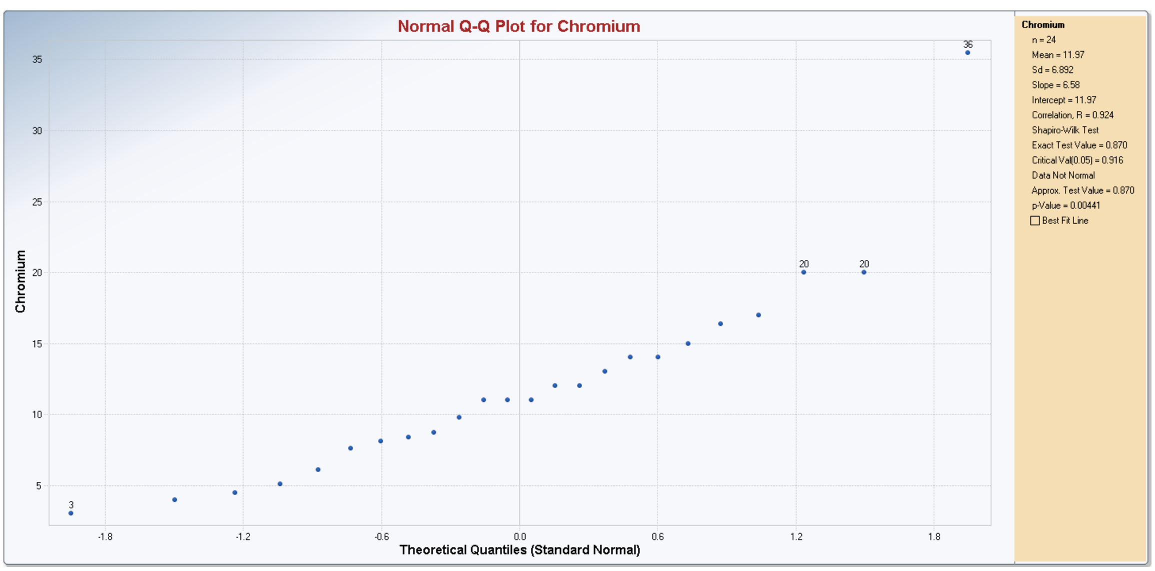
<u>Notes</u>: This option for **Graphs by Group** is specifically provided when the user wants to display multiple graphs for a variable by a group variable (e.g., site AOC1, site AOC2, background). This kind of display represents a useful visual comparison of the values of a variable (e.g., concentrations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

Example 8-1a (Superfund Site Data Continued): The lognormal and normal GOF test results on chromium concentrations are shown in the following figures.

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

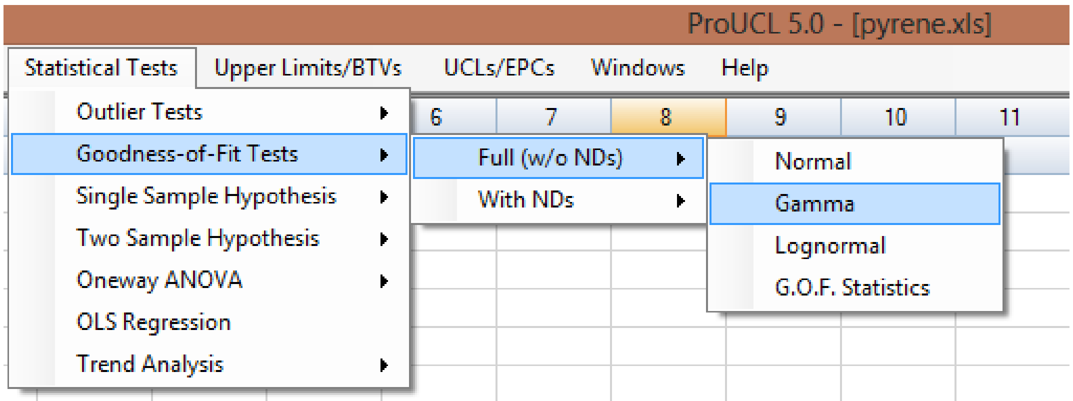


Output Screen for Normal Distribution (Full (w/o NDs))
Selected Options: Shapiro-Wilk, Best Fit Line not Displayed



8.2.2 GOF Tests for Gamma Distribution

1. Click **Goodness-of-Fit Tests ► Full (w/o NDs) ► Gamma**

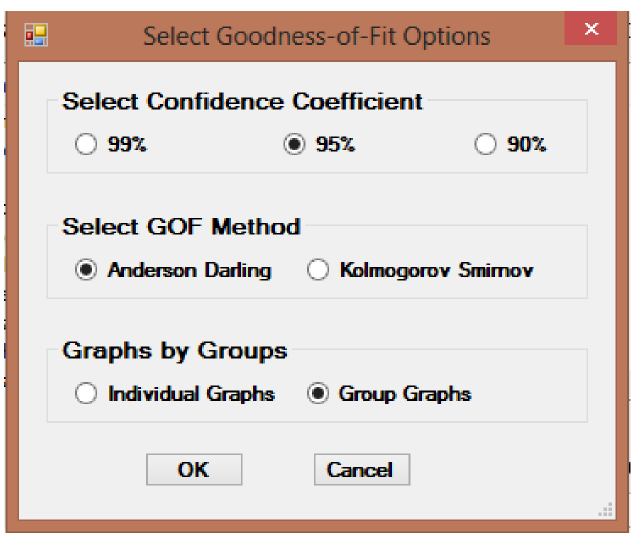


2. The **Select Variables** screen (described in Chapter 3) will appear.

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

- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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.



  - The default option for the **Confidence Coefficient** is **95%**.

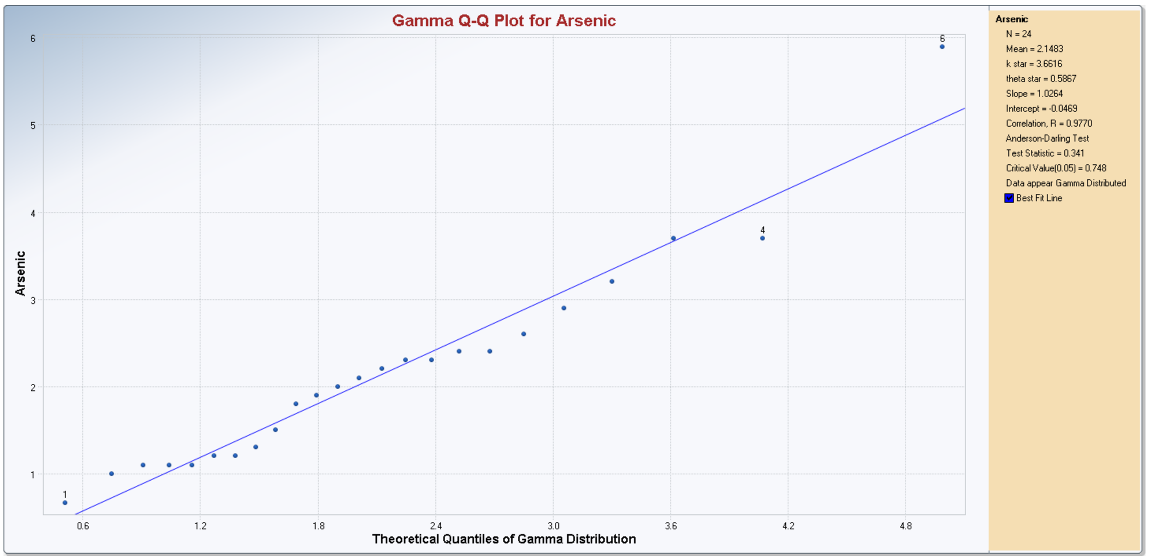
  - The default GOF method is **Anderson Darling**.

- 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 8-1b (Superfund Site Data Continued): The Gamma GOF test results, for the data set of arsenic concentrations, are shown in the following G.O.F. Q-Q plot.

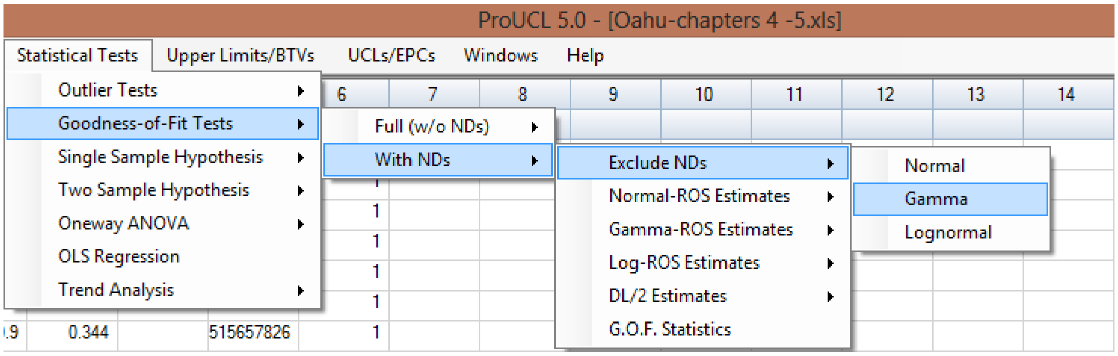
Output Screen for Gamma Distribution (Full (w/o NDs))
Selected Options: Anderson Darling with Best Line Fit



8.3 Goodness-of-Fit Tests Excluding NDs

This option is the most important option for a GOF test based upon data sets with ND observations. Based upon the skewness and distribution of detected data, ProUCL computes 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 terms and BTV estimates.

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

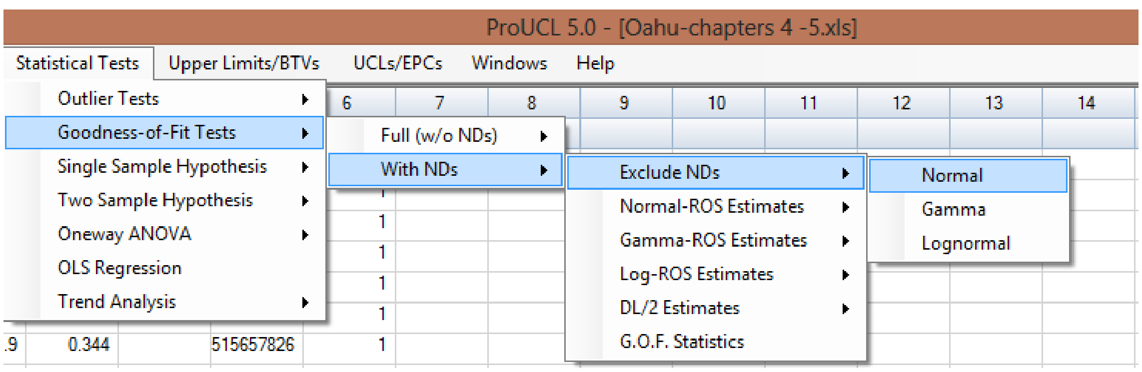


2. Select distribution to be tested: Normal, Gamma, or Lognormal.

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

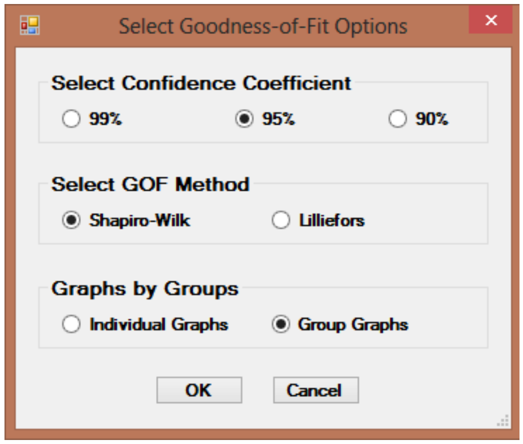
8.3.1 Normal and Lognormal Options

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



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

- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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: **Normal** or **Lognormal** is clicked, the following window is displayed

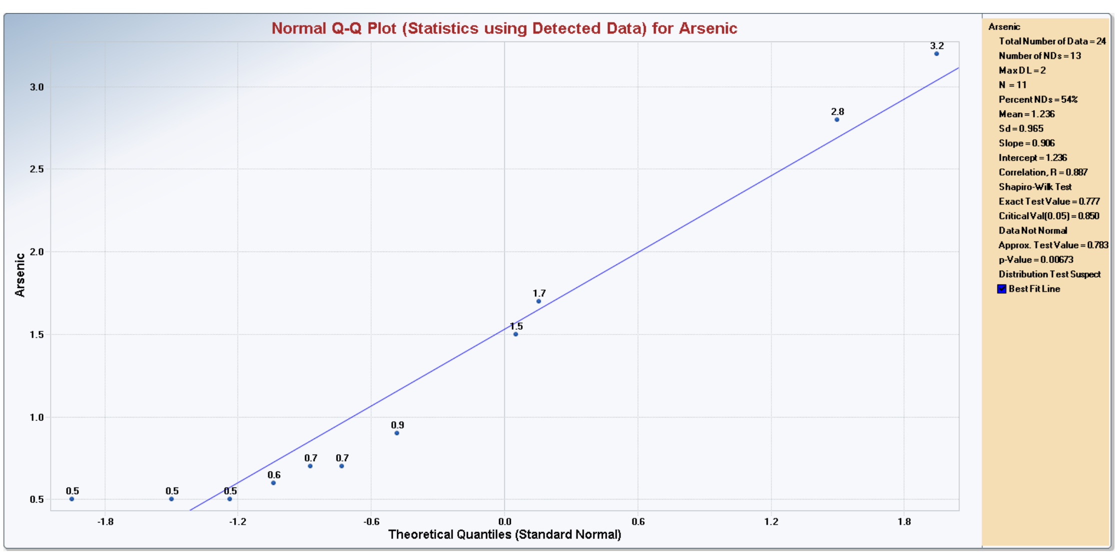


- o The default option for the **Confidence Coefficient** is **95%**.
- o The default GOF Method is **Shapiro-Wilk**.
- o The default option for **Graphs by Group** is **Group Graphs**. If you want to see the plots for all selected variables individually, and then check the button next to **Individual Graphs**.

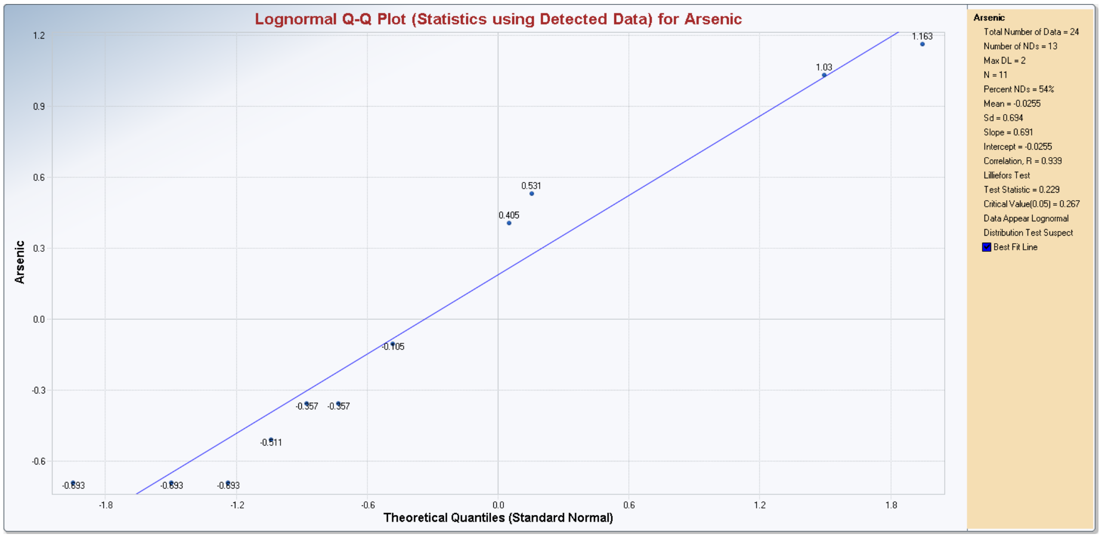
- Click the **OK** button to continue or the **Cancel** button to cancel the option.
- Click the **OK** button to continue or the **Cancel button** to cancel the GOF tests.

Example 8-2a. Consider the arsenic Oahu data set with NDs discussed in the literature (e.g., Helsel, 2012; NADA in R [Helsel, 2013]). The normal and lognormal GOF test results based upon the detected data respectively are shown in the following two figures.

Output Screen for Normal Distribution (Exclude NDs)
Selected Options: Shapiro-Wilk with Best Fit Line

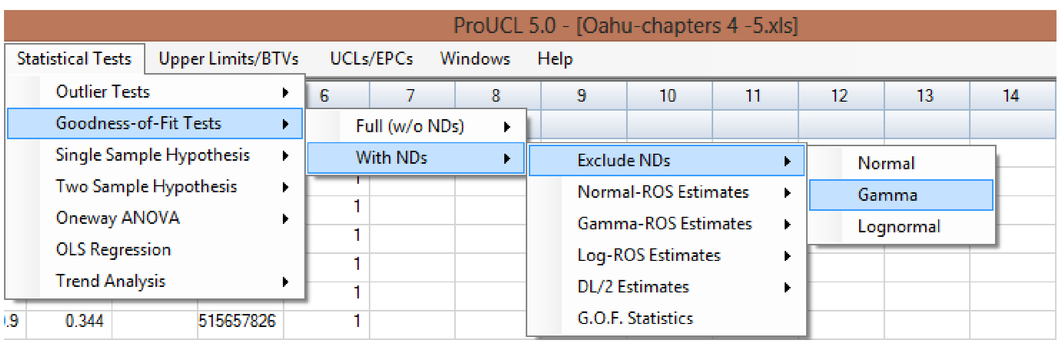


Output Result for Lognormal Distribution (Exclude NDs)
Selected options: Lilliefors Test with Best Fit Line



8.3.2 Gamma Distribution Option

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

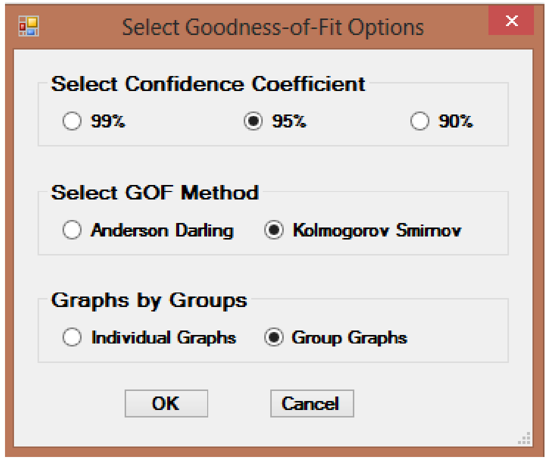


2. The **Select Variables** screen (Chapter 3) will appear.

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

- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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 (Gamma) is clicked, the following window is shown.

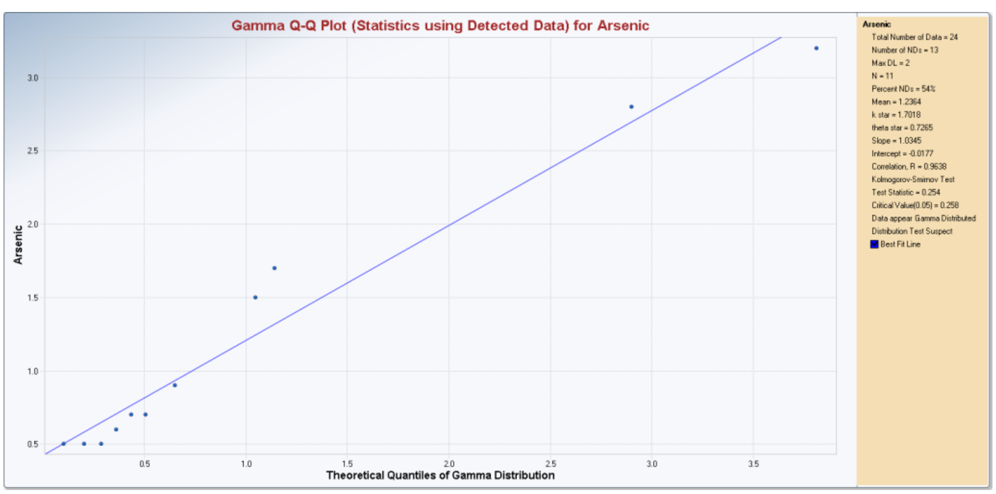


- o The default option for the **Confidence Coefficient** is **95%**.
- o The default GOF test Method is the **Anderson Darling test**.
- o The default option for **Graph by Groups** is **Groups Graphs**. If you want to display all selected variables on separate graphs, check the button next to **Individual Graphs**.

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

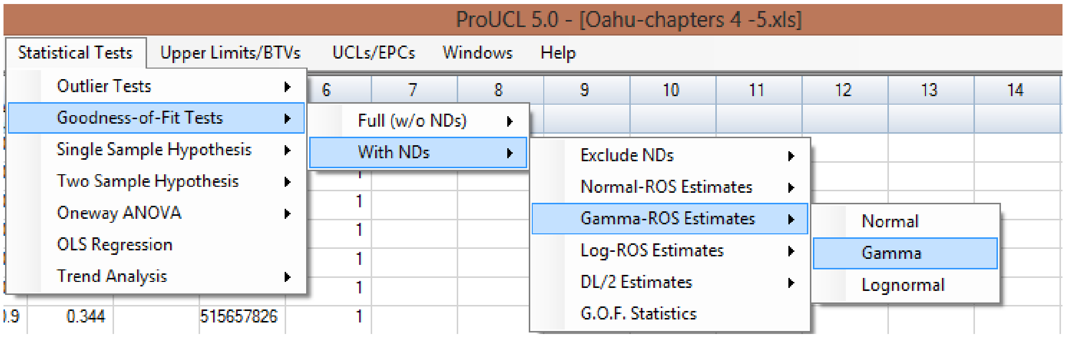
Example 8-2b (continued). Consider the arsenic Oahu data set with NDs as discussed in Example 8-2a above. The gamma GOF test results based upon the detected data are shown in the following GOF Q-Q plot.

Output Screen for Gamma Distribution (Exclude NDs)
Selected Options: Kolmogorov Smirnov Test with Best Fit Line



8.4 Goodness-of-Fit Tests with ROS Methods

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

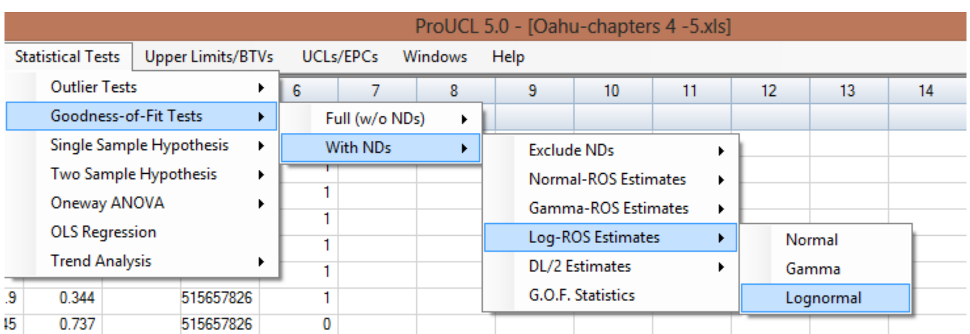


2. Select the distribution to be tested: Normal, Lognormal, or Gamma

- To test for normal distribution, click on **Normal** from the drop-down menu list.
- To test for gamma distribution, click on **Gamma** from the drop-down menu list.
- To test for lognormal distribution, click on **Lognormal** from the drop-down menu.

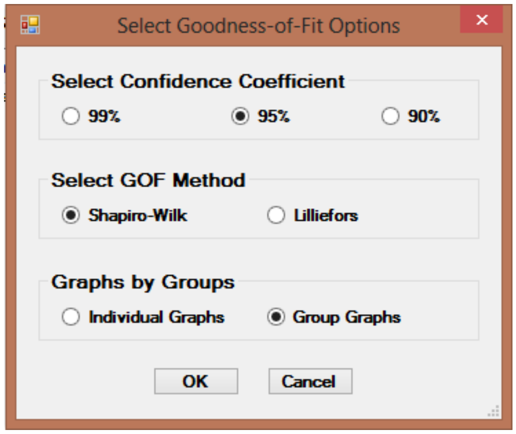
8.4.1 Normal or Lognormal Distribution (Log-ROS Estimates)

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



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

- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be produced by using a Group variable, then select a group variable by clicking the arrow below the **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: Normal or Lognormal is clicked, the following window appears.

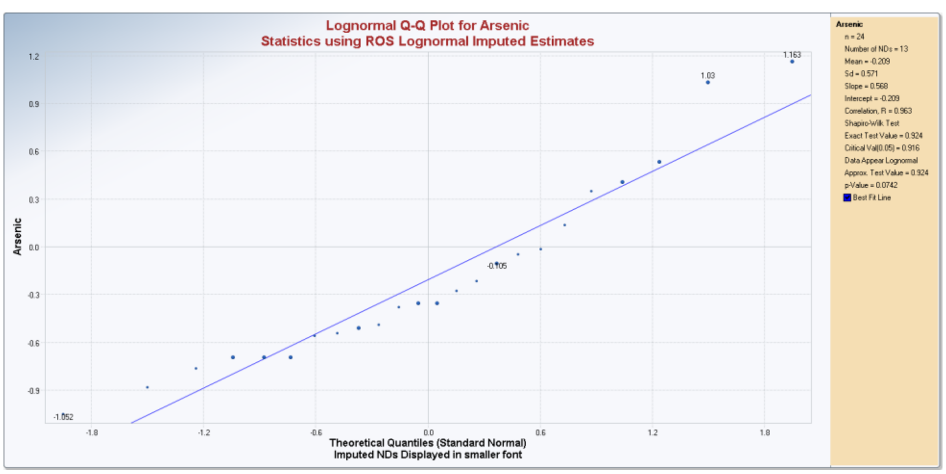


- o The default option for the **Confidence Coefficient** is **95%**.
- o The default GOF test Method is **Shapiro-Wilk**.
- o The default option for **Graphs by Group** is **Group Graphs**. If you want to display graphs for all selected variables individually, check the button next to **Individual Graphs**.

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

Example 8-2c (continued). Consider the arsenic Oahu data set with NDs considered earlier in this chapter. The lognormal GOF test results on LROS data (detected and imputed LROS NDs) is shown in the following GOF Q-Q plot.

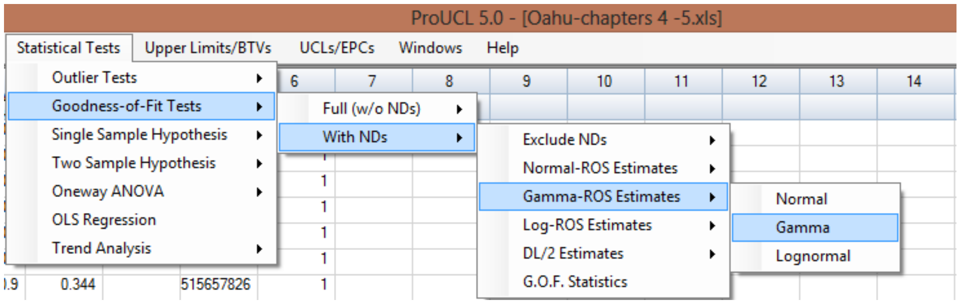
Output Screen for Lognormal Distribution (Log-ROS Estimates)
Selected Options: Shapiro Wilk test with Best Line Fit



Note**:** The font size of ND values is smaller than that of the detected values.

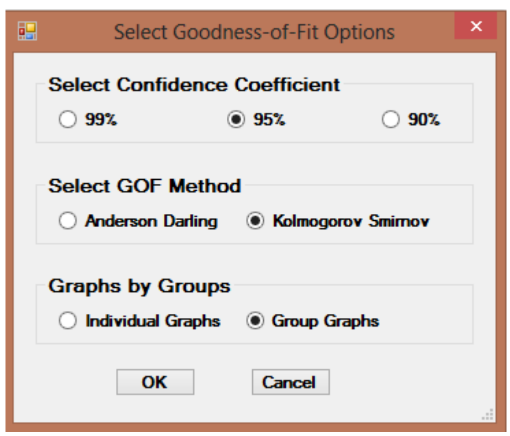
8.4.2 Gamma Distribution (Gamma-ROS Estimates)

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



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

- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be generated 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 (Gamma) is clicked, the following window will be shown.

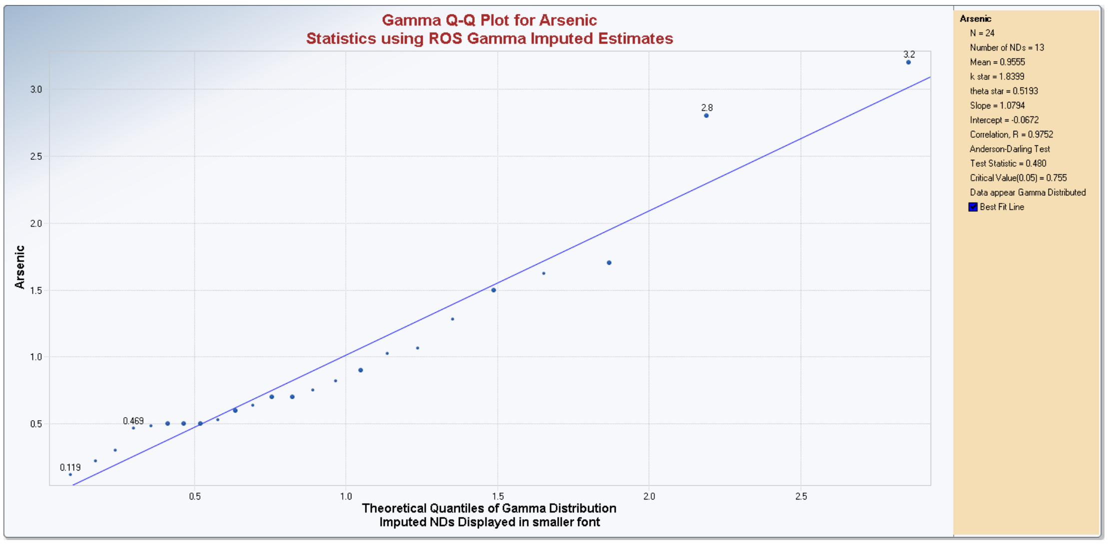


  - The default option for the **Confidence Coefficient** is **95%**.
  - The default GOF test **Method** is **Anderson Darling**.
  - The default option for **Graph by Groups** is Group **Graphs**. If you want to generate separate graphs for all selected variables, the check the button next to **Individual Graphs**.

- Click the **OK** button to continue or the **Cancel** button to cancel the GOF tests.

Example 8-2d (continued). Consider the arsenic Oahu data set with NDs considered earlier. The gamma GOF test results on GROS data (detected and imputed GROS NDs) are shown in the following GOF Q-Q plot.

Output Screen for Gamma Distribution (Gamma-ROS Estimates)
Selected Options: Anderson Darling



<u>Note:</u> The font size of ND values in the above graph (and in all GOF graphs) is smaller than that of detected values.

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

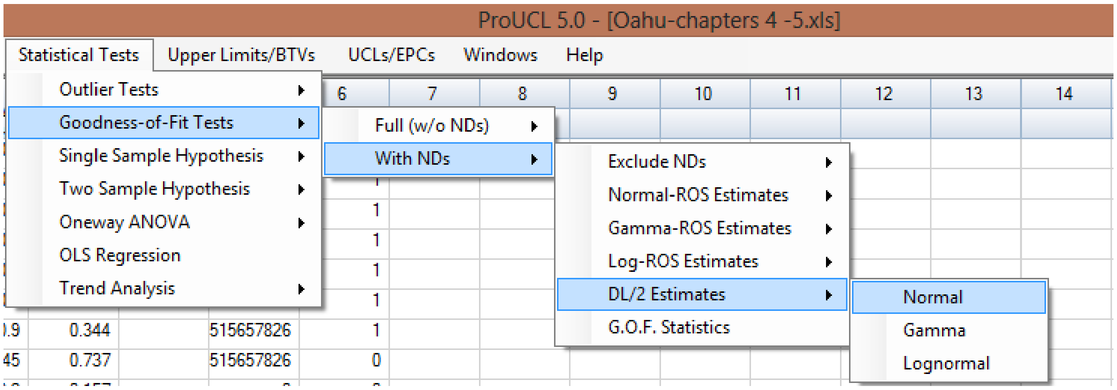
1. Click **Goodness-of-Fit Tests ► With NDs ► DL/2 Estimates**

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

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

8.5.1 *Normal or Lognormal Distribution (DL/2 Estimates)*

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



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

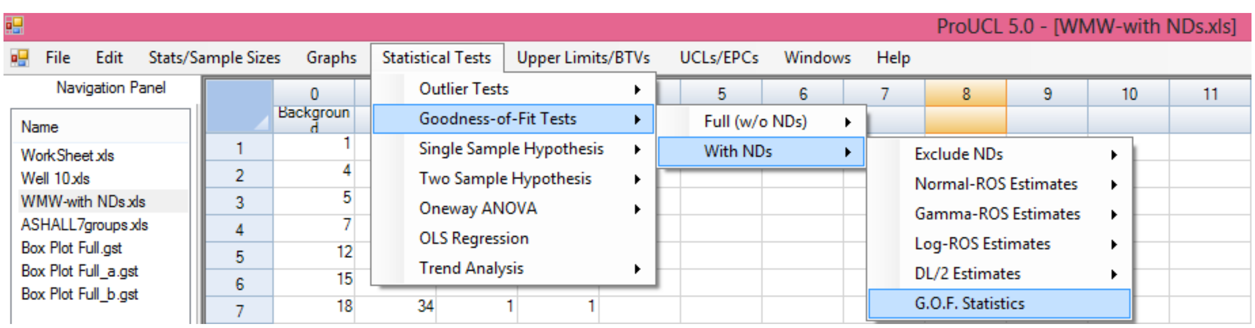
- Select one or more variable(s) from the **Select Variables** screen.
- If graphs have to be generated by using a Group variable, then select a group variable by clicking the arrow below the **Select Group Column (Optional)** button. This will result in a drop-down list of available variables. The user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

The rest of the process to determine the distribution (normal, lognormal, and gamma) of the data set thus obtained is the same as described in earlier sections.

8.6 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.

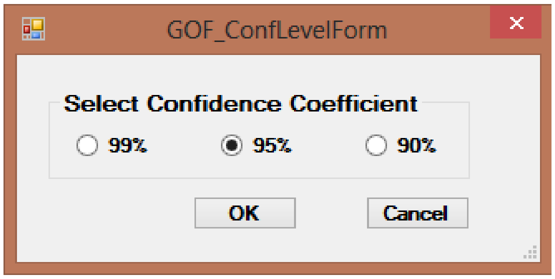
1. Click **Goodness-of-Fit ► With NDs ► G.O.F. Statistics**



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

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

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



- The default confidence level is **95%**.

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

Example 8-2e (continued). Consider the arsenic Oahu data set with NDs discussed earlier. Partial GOF test results, obtained using the **G.O.F. Statistics** option, are summarized in the following table.

Sample Output Screen for G.O.F. Test Statistics on Data Sets with Nondetect 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	0.95

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

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 Distrib
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
Lilliefors (Detects Only)	0.229	0.267	Data Appear Lognormal
Shapiro-Wilk (NDs = DL)	0.906	0.916	Data Not Lognormal
Lilliefors (NDs = DL)	0.214	0.181	Data Not Lognormal
Shapiro-Wilk (NDs = DL/2)	0.865	0.916	Data Not Lognormal
Lilliefors (NDs = DL/2)	0.217	0.181	Data Not Lognormal
Shapiro-Wilk (Lognormal ROS Estimates)	0.924	0.916	Data Appear Lognormal
Lilliefors (Lognormal ROS Estimates)	0.143	0.181	Data Appear Lognormal

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

Chapter 9

Single-Sample and Two-Sample Hypotheses Testing Approaches

This chapter illustrates single-sample and two-sample parametric and nonparametric hypotheses testing approaches as incorporated in the ProUCL software. All hypothesis tests are available under the "Statistical Tests" module of ProUCL 5.0. The 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 5.0 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 all entries equal to '1.' This will allow the user to compare two groups (e.g., arsenic in background vs. site samples) with one of the groups having some NDs and the other group having all detected data.

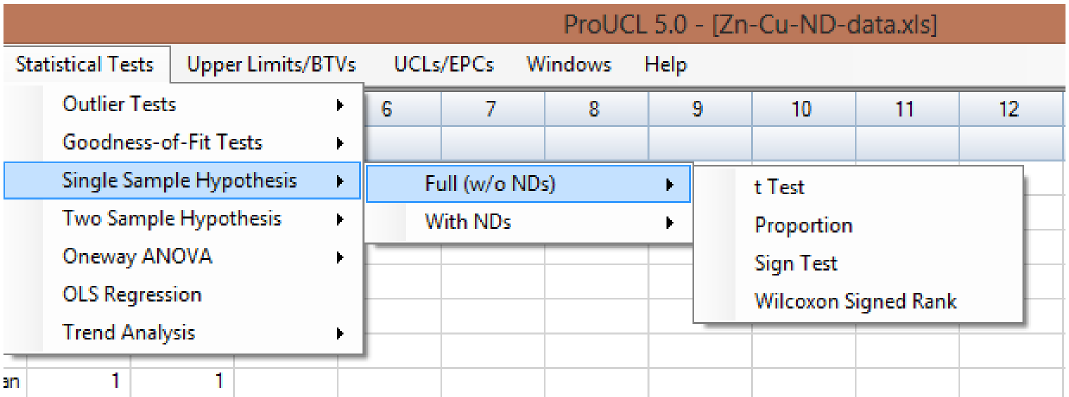
9.1 Single-Sample Hypotheses Tests

In many environmental applications, single-sample hypotheses tests are used to compare site data with pre-specified cleanup standards or compliance limits (CLs). The single-sample hypotheses tests are useful when the environmental parameters such as the cleanup standard (C_s), 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 5.0 can perform these hypotheses tests on data sets with and without ND observations. However, it should be noted that for single-sample hypotheses tests (e.g., sign test, proportion test) used to compare site mean/median concentration level with a C_s or a CL (e.g., proportion test), all NDs (if any) should lie below the cleanup standard, $C_{s.}$ For proper use of these hypotheses testing approaches, the differences between these tests should be noted and understood. Specifically, a t-test or a WSR test 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, C_s, or action level also representing a measure of central tendency (e.g., mean, median); whereas, a proportion test compares if the proportion of site observations from an AOC exceeding a CL exceeds a specified proportion, P_0 (e.g., 5%, 10%). ProUCL 5.0 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.

9.1.1 Single-Sample Hypothesis Testing for Full Data without Nondetects

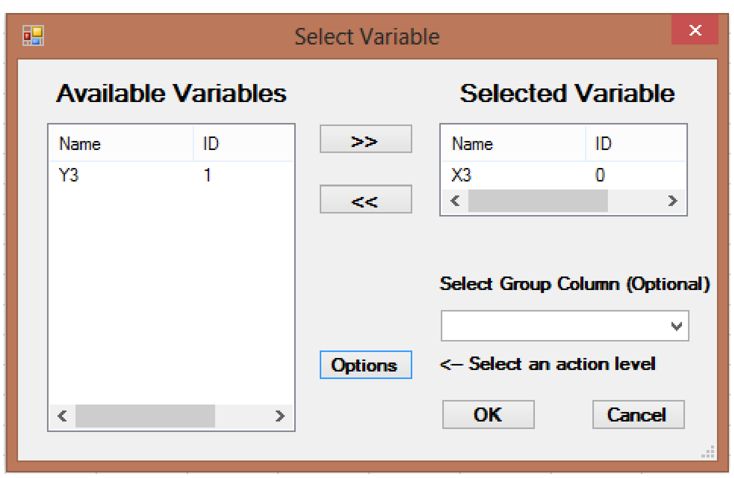
1. Click **Single Sample Hypothesis ► Full (w/o NDs)**



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

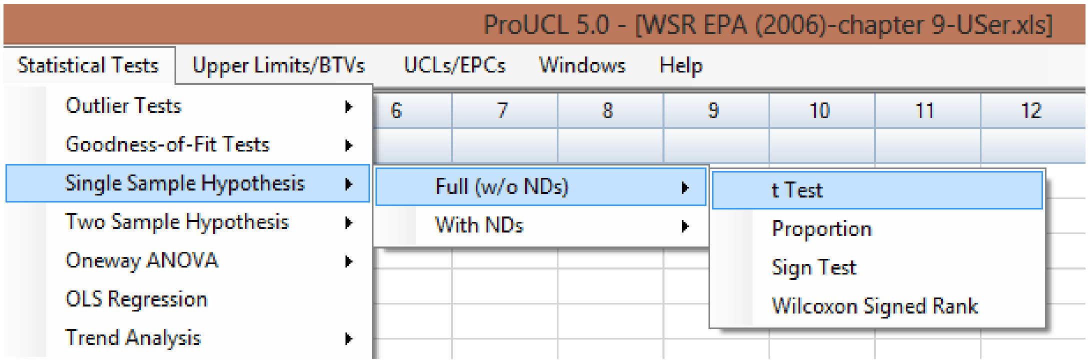
- To perform a t-test, click on **t-Test** from the drop-down menu as shown above.
- To perform a Proportion test, click on **Proportion** from the drop-down menu.
- To run a Sign test, click on **Sign test** from the drop-down menu.
- To run a Wilcoxon Signed Rank (WSR) test, click on **Wilcoxon Signed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



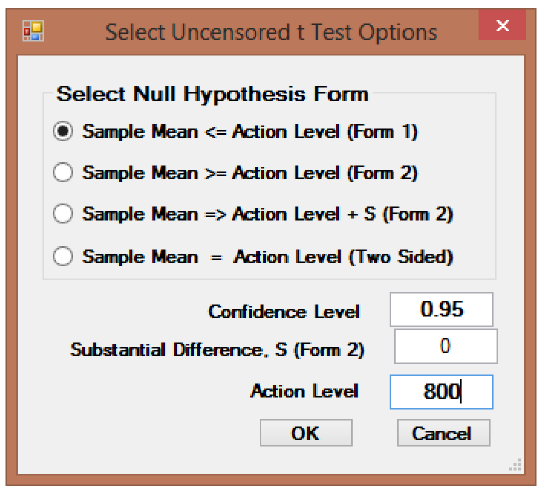
9.1.1.1 Single-Sample t-Test

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



2. 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.



  - 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 9-1a. Consider the WSR data set described in EPA (2006a). One Sample t-test results are summarized as follows.

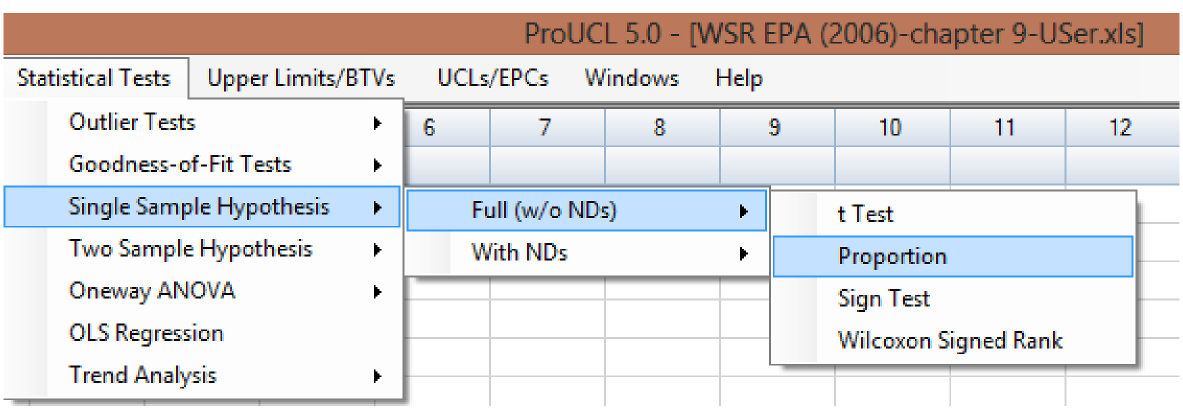
Output for Single-Sample t-Test (Full Data w/o NDs)

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)	

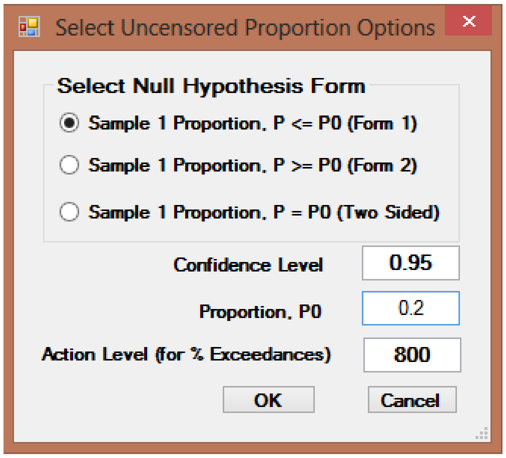
9.1.1.2 Single-Sample Proportion Test

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



2. 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.



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

Example 9-1b (continued). Consider the WSR data set described in EPA (2006a). One Sample proportion test results are summarized as follows.

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

User Selected Options	
Date/Time of Computation	3/17/2013 10:29:38 PM
From File	WSR EPA (2006)-chapter 9-USer.xls
Full Precision	OFF
Confidence Coefficient	95%
User Specified Proportion	0.200 (P0 of Exceedances of Action Level)
Action/compliance Limit	800.000
Select Null Hypothesis	Sample Proportion, P of Exceedances of Action Level >= User Specified Proportion (Form 2)
Alternative Hypothesis	Sample Proportion, P of Exceedances of Action Level < the User Specified Proportion

WSR1

One Sample Proportion 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
Number of Exceedances	8
Sample Proportion of Exceedances	0.8

H0: Sample Proportion >= 0.2 (Form 2)

P-Value Based Upon BD (Binomial Dist)	1

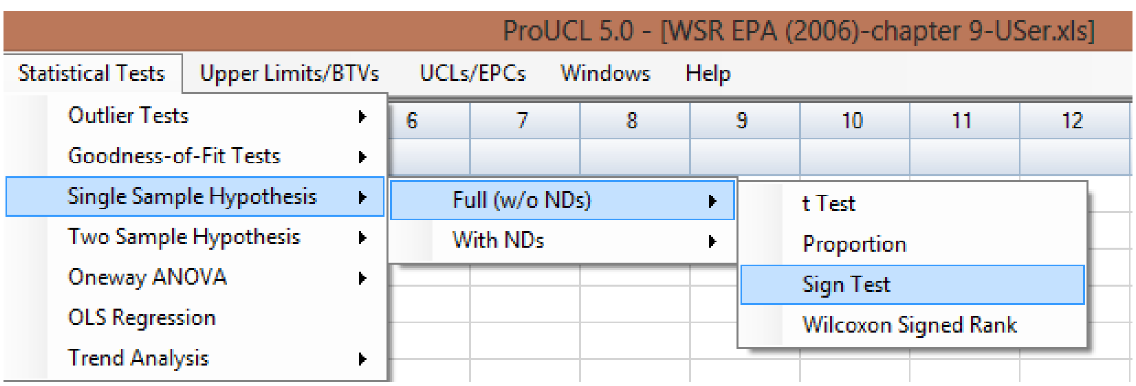
Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Sample Proportion >= 0.2

P-Value > Alpha (0.05)

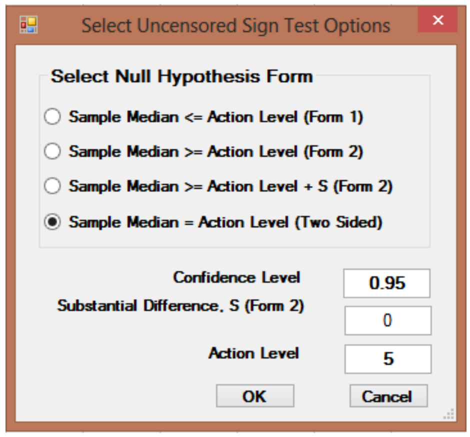
9.1.1.3 Single-Sample Sign Test

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



2. 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.



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

Example 9-1c (continued). Consider the WSR data set described in EPA (2006a). The Sign test results are summarized as follows.

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

From File	WSR EPA (2006)-chapter 9-USer.xls
Full Precision	OFF
Confidence Coefficient	95%
Substantial Difference	0.000
Action Level	5.000
Selected Null Hypothesis	Median = Action/compliance Limit (2 Sided Alternative)
Alternative Hypothesis	Median <> Action/compliance Limit

WSR2

One Sample Sign Test

Raw Statistics

Number of Valid Observations	49
Number of Distinct Observations	44
Minimum	1.09
Maximum	7.5
Mean	5.048
Median	5.55
SD	1.775
SE of Mean	0.254
Number Above Action Level	29
Number Equal Action Level	0
Number Below Action Level	20

H0: Sample Median = 5

Large Sample Z Test Statistic	1.286
P-Value	0.199

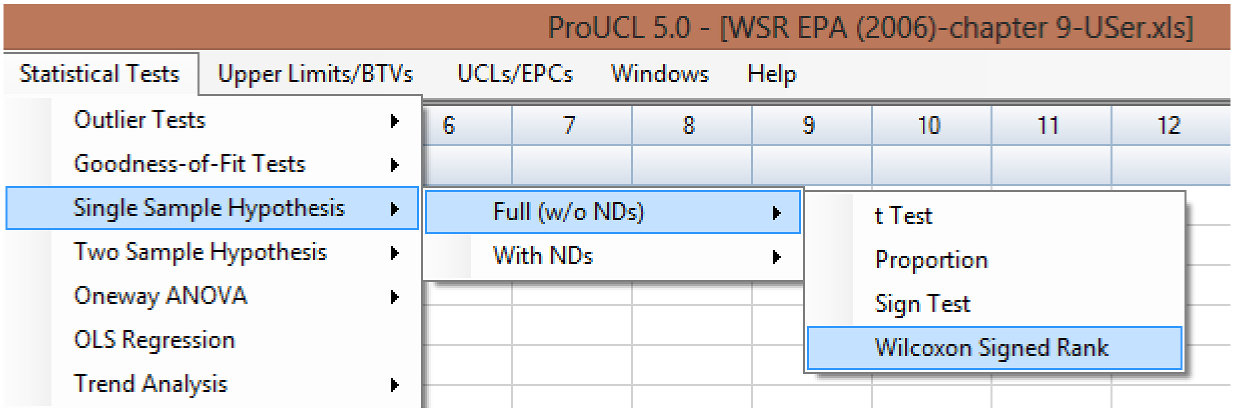
Conclusion with Alpha = 0.05

Do Not Reject H0 at the specified level of significance (0.05). Conclude Median = 5

P-Value > Alpha (0.05)

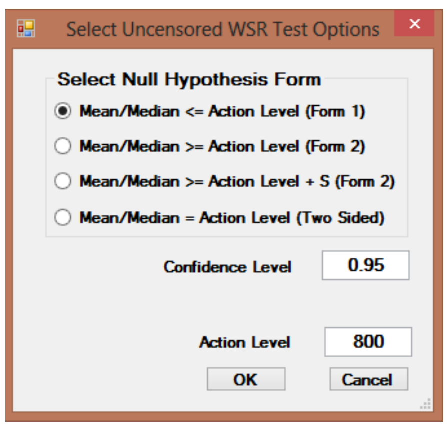
9.1.1.4 Single-Sample Wilcoxon Signed Rank (WSR) Test

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



2. 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.



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

Example 9-1d (continued). Consider the WSR data set described in EPA (2006a). One Sample WSR test results are summarized as follows.

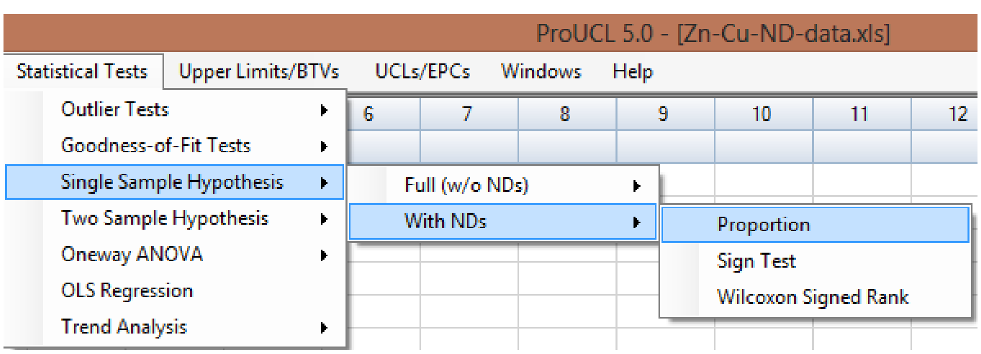
Output for Single-Sample Wilcoxon Signed Rank Test (Full Data without NDs)

Confidence Coefficient	95%
Substantial Difference	0.000
Action Level	800.000
Selected Null Hypothesis	Mean/Median <= Action Level (Form 1)
Alternative Hypothesis	Mean/Median > the Action Level
WSR1	
One Sample Wilcoxon Signed Rank 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
Number Above Action Level	8
Number Equal Action Level	0
Number Below Action Level	2
T-plus	50
T-minus	5
H0: Sample Mean/Median <= 800 (Form 1)	
Exact Test Statistic	50
Critical Value (0.05)	45
P-Value	0.0098
Conclusion with Alpha = 0.05	
Reject H0, Conclude Mean/Median > 800	
P-Value < Alpha (0.05)	

9.1.2 Single-Sample Hypothesis Testing for Data Sets with Nondetects

Most of the one-sample tests such as the Proportion test and the Sign test on data sets with ND values assume 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. Single-sample hypothesis tests for data sets with NDs are shown in the following ProUCL 5.0 screen shot.

1. Click on **Single Sample Hypothesis ► With NDs**

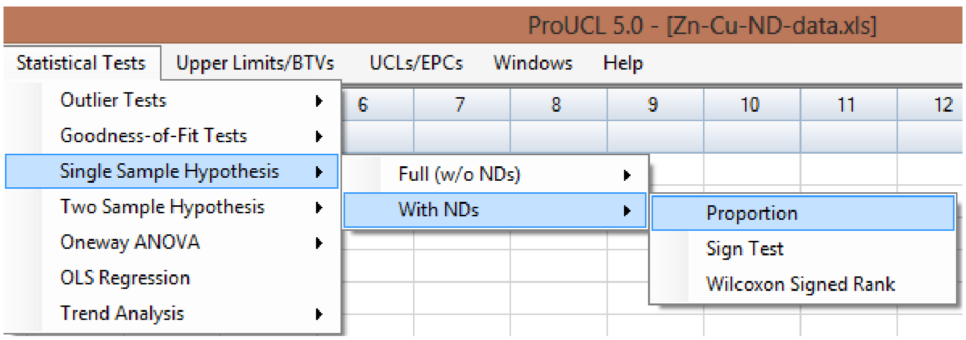


2. Select the With NDs option

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

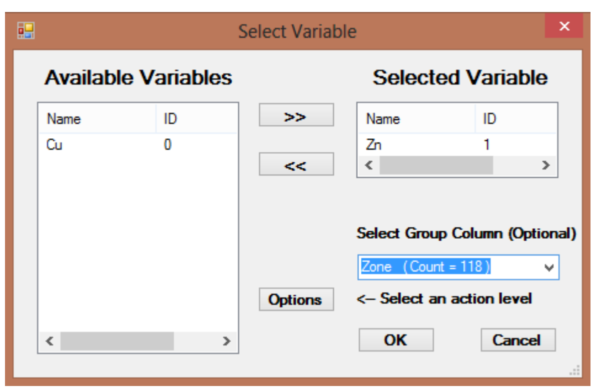
9.1.2.1 Single Proportion Test on Data Sets with NDs

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

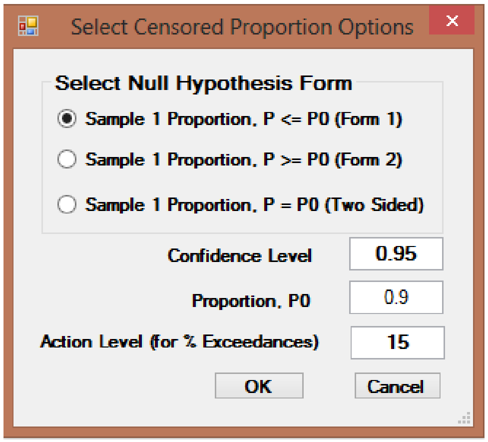


2. The **Select Variables** screen will appear.

- 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.



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



  - Specify the **Confidence Level**; default is **0.95**.
  - Specify meaningful values for **Proportion** and the **Action Level (=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 9-2a. Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed in the literature (Helsel, 2012, NADA in R [Helsel, 2013]). This data set is used here to illustrate the one sample proportion test on a data set with NDs. The output sheet generated by ProUCL 5.0 is described as follows.

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

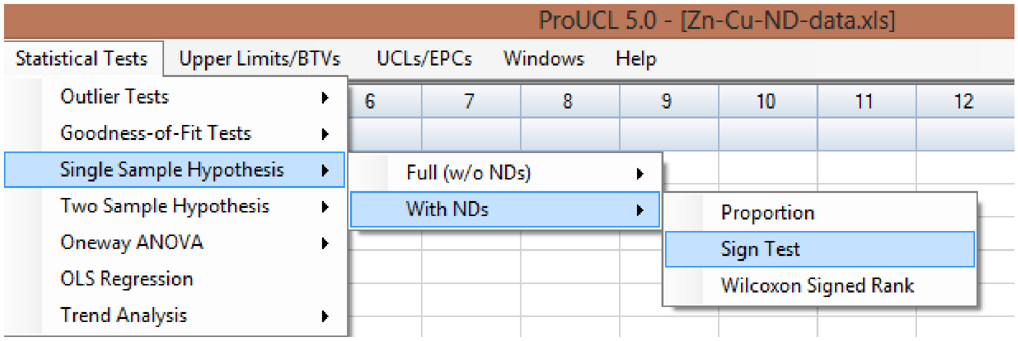
Conclusion with Alpha = 0.05

Do Not Reject H0, Conclude Sample Proportion <= 0.9

P-Value > Alpha (0.05)

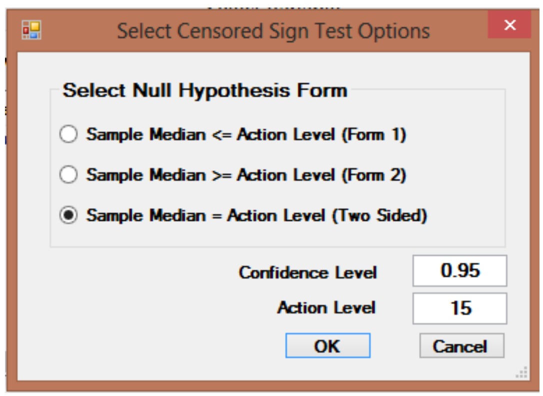
9.1.2.2 Single-Sample Sign Test with NDs

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



2. 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.



  - 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.

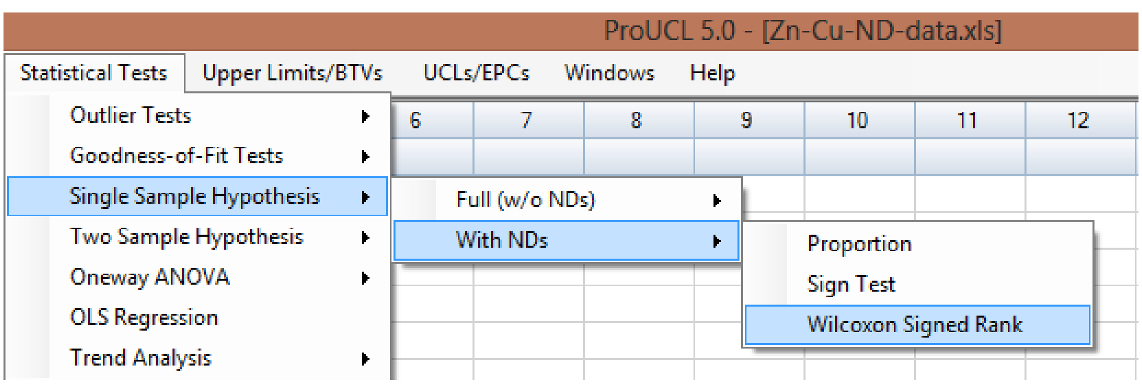
Example 9-2b (continued). Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed above. This data set is used here to illustrate the Single-Sample Sign test on a data set with NDs. The output sheet generated by ProUCL 5.0 is described as follows.

Output for Single-Sample Sign Test (Data with Nondetects)

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)	

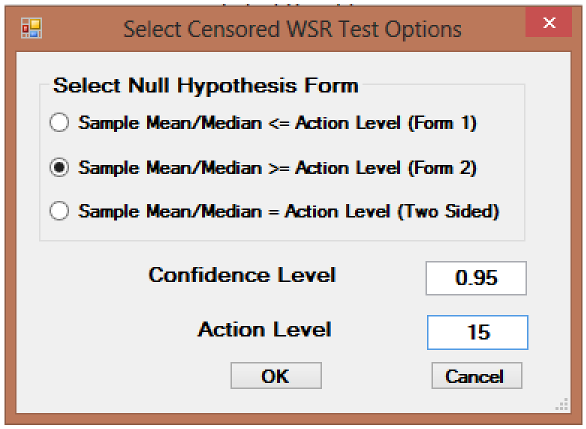
9.1.2.3 Single-Sample Wilcoxon Signed Rank Test with NDs

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



2. 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.



- 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 9-2c (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 5.0 is provided as follows.

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

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

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	
Minimum Detect	3	
Maximum Detect	90	
Mean of Detects	23.13	
Median of Detects	20	
SD of Detects	19.03	
Median of Processed Data used in WSR	18.5	
Number Above Action Level	27	
Number Equal Action Level	1	
Number Below Action Level	22	
T-plus	764	
T-minus	461	

H0: Sample Median >= 15 (Form 2)		
Large Sample z-Test Statistic	1.269	
Critical Value (0.05)	-1.645	
P-Value	0.898	
Conclusion with Alpha = 0.05		
Do Not Reject H0, Conclude Mean/Median >= 15		
P-Value > Alpha (0.05)		
Dataset contains multiple Non-Detect values!		
All NDs are replaced by their respective DL/2		

9.2 Two-Sample Hypotheses Testing Approaches

The two-sample hypotheses testing approaches available in ProUCL 5.0 are described in this section. Like Single-Sample Hypothesis, the Two-Sample Hypothesis options are available under the "Statistical Tests" module of ProUCL 5.0. These approaches are used to compare the parameters and distributions of the two populations (e.g., Background vs. AOC) based upon data sets collected from those populations. Both forms (Form 1 and Form 2, and Form 2 with Substantial Difference, S) of the two-sample hypothesis testing approaches are available in ProUCL 5.0. The methods are available for full-uncensored data sets as well as for data sets with ND observations with multiple detection limits.

- **Full (w/o NDs)** – performs parametric and nonparametric hypothesis tests on uncensored data sets consisting of all detected values. The following tests are available:
  - Student’s t and Satterthwaite tests to compare the means of two populations (e.g. Background versus AOC).
  - F-test to the check the equality of dispersions of two populations.
  - Two-sample nonparametric Wilcoxon-Mann-Whitney (WMW) test. This test is equivalent to Wilcoxon Rank Sum (WRS) test.
- **With NDs** – performs hypothesis tests on left-censored data sets consisting of detected and ND values. The following tests are available:
  - Wilcoxon-Mann-Whitney test. All observations (including detected values) below the highest detection limit are treated as ND (less than the highest DL) values.
  - Gehan’s test is useful when multiple detection limits may be present.

- Tarone-Ware test is useful when multiple detection limits may be present.

The details of these methods can be found in the ProUCL 5.0 Technical Guide 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.

Notes: As mentioned before, it is pointed out that, when one wants to use two-sample hypotheses tests on data sets with NDs, ProUCL 5.0 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 equal to '1'. This will allow the user to compare two groups (e.g., arsenic in background vs. site samples) with one of the groups having some NDs and the other group having all detected data.

9.2.1 Two-Sample Hypothesis Tests for Full Data

Full (w/o NDs): This option is used to analyze data sets consisting of all detected values. The following two-sample tests are available in ProUCL 5.0.

- Student's t and Satterthwaite tests to compare the means of two populations (e.g., Background versus AOC).
- F-test is also available to test the equality of dispersions of two populations.
- Two-sample nonparametric Wilcoxon-Mann-Whitney (WMW) test.

- **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 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 5.0 Technical Guide.

- **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 (see the ProUCL 5.0 Technical Guide for further details).

- **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.

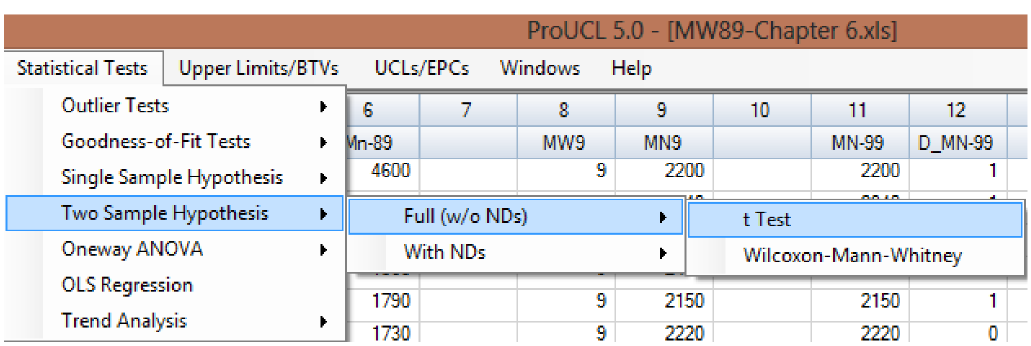
- **Two-Sample Nonparametric WMW Test**

    - 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 log-normally distributed. For large samples (e.g., ≥ 20), the distribution of the WMW test statistic can be approximated by a normal distribution.

Notes: The use of the tests listed above is not recommended on log-transformed data sets, especially when the parameters of interests are the population means. In practice, the cleanup and remediation decisions have to be made in 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.

1. Click on **Two Sample Hypothesis ► Full (w/o NDs)**



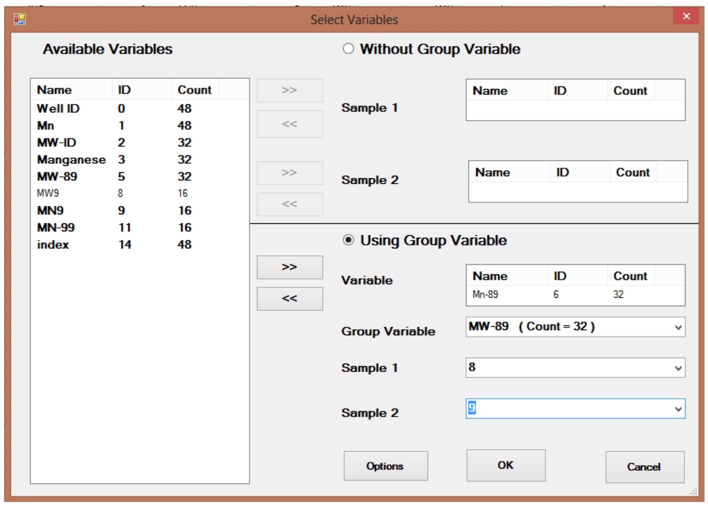
2. Select the **Full (w/o NDs)** option

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

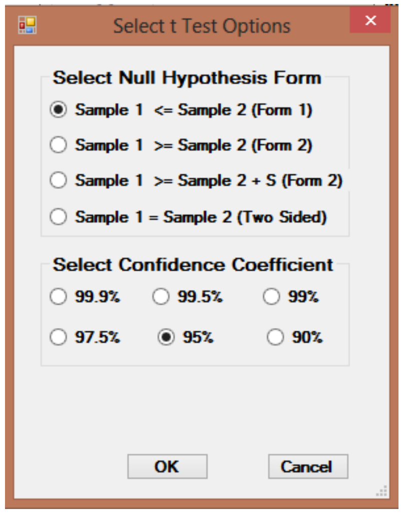
9.2.1.1 Two-Sample t-Test without NDs

1. Click on **Two Sample Hypothesis ► Full (w/o NDs) ► t Test**
2. The **Select Variables** screen will appear.

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



- **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) for the two populations (e.g., site vs. background) 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 group variable by using the **Group Variable**. The user should select an appropriate variable representing a group variable. The user can use letters, numbers, or alphanumeric labels for the group names.
  - o When the **Options** button is clicked, the following window will be shown.



- 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 9-3. Consider the manganese concentrations data set collected from three wells: MW1, an upgradient well, and MW8 and MW9 are two downgradient wells. The two-sample t-test results comparing Mn concentrations in MW8 vs. MW9 are described as follows.

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

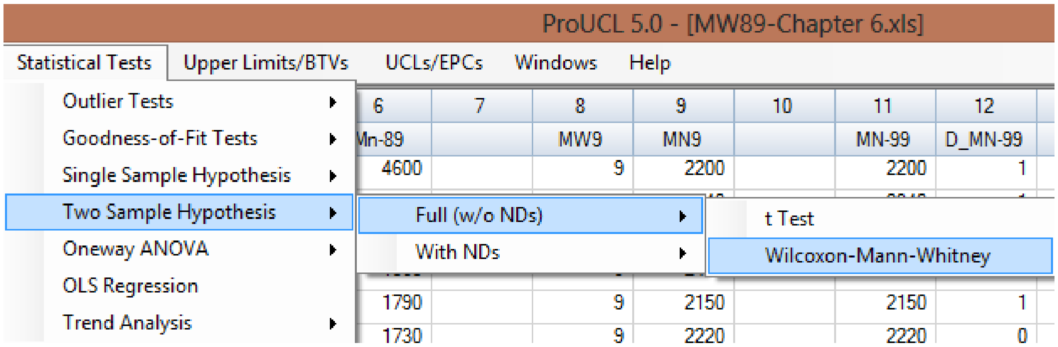
Numerator DF	Denominator DF	F-Test Value	P-Value
15	15	2.812	0.054

Conclusion with Alpha = 0.05

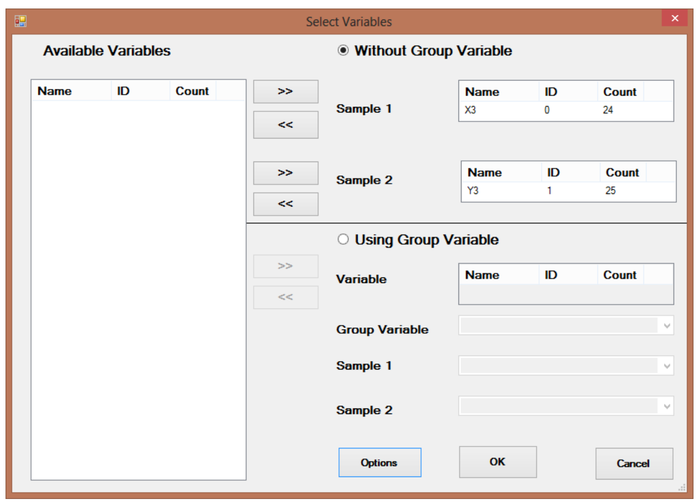
Two variances appear to be equal

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

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



2. The **Select Variables** screen will appear.

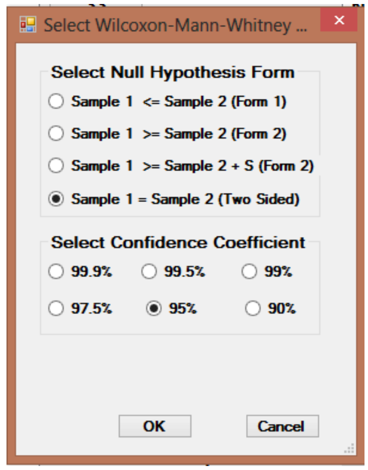


- Select variable(s) from the Select Variables screen.
- **Without Group Variable**: This option is used when the data values of the variable (arsenic) are given in separate columns.
- **With Group Variable**: This option is used when data of the variable (arsenic) are given in the same column. The values are separated into different samples (groups) by the values of an associated Group Variable.
- When the Group option is used, the user then selects a group variable by using the **Group Variable**. The user should select an appropriate variable representing a group variable. The user can use letters, numbers, or alphanumeric labels for the group names.

<u>Notes:</u> ProUCL 5.0 has been written using environmental terminology such as performing background versus site comparisons. However, all tests and procedures incorporated in ProUCL 5.0 can be used on data sets from any other application. For other applications such as comparing

a new treatment drug versus older treatment drug, the group variable may represent the two groups: Control Drug and New Drug.

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



  - Specify a **Substantial Difference, S** value. The default choice is **0**.
  - Choose the **Confidence 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 Sample 1 vs. Sample 2 comparison.

Example 9- 4. The two-sample Wilcoxon Mann Whitney (WMW) test results on a data set with ties are summarized as follows.

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

Substantial Difference	0.000
Selected Null Hypothesis	Sample 1 Mean/Median = Sample 2 Mean/Median (Two Sided Alternative)
Alternative Hypothesis	Sample 1 Mean/Median <> Sample 2 Mean/Median

Sample 1 Data: X3
Sample 2 Data: Y3

Raw Statistics

	Sample 1	Sample 2
Number of Valid Observations	24	25
Number of Distinct Observations	18	19
Minimum	5.687	1.85
Maximum	31.2	79.06
Mean	17.38	39.8
Median	17.56	44.63
SD	7.421	19.39
SE of Mean	1.515	3.878

Wilcoxon-Mann-Whitney (WMW) Test

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

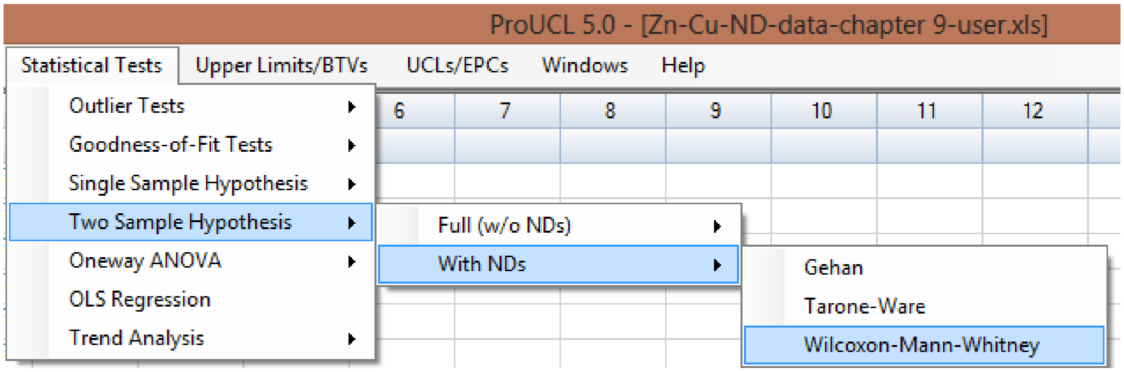
Sample 1 Rank Sum W-Stat	396
WMW U-Stat	96
Standardized WMW U-Stat	-4.083
Mean (U)	300
SD(U) - Adj ties	49.97
Lower Approximate U-Stat Critical Value (0.025)	-1.96
Upper Approximate U-Stat Critical Value (0.975)	1.96
P-Value (Adjusted for Ties)	4.4474E-5

Conclusion with Alpha = 0.05
Reject H0, Conclude Sample 1 <> Sample 2

P-Value < alpha (0.05)

9.2.2 Two-Sample Hypothesis Testing for Data Sets with Nondetects

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

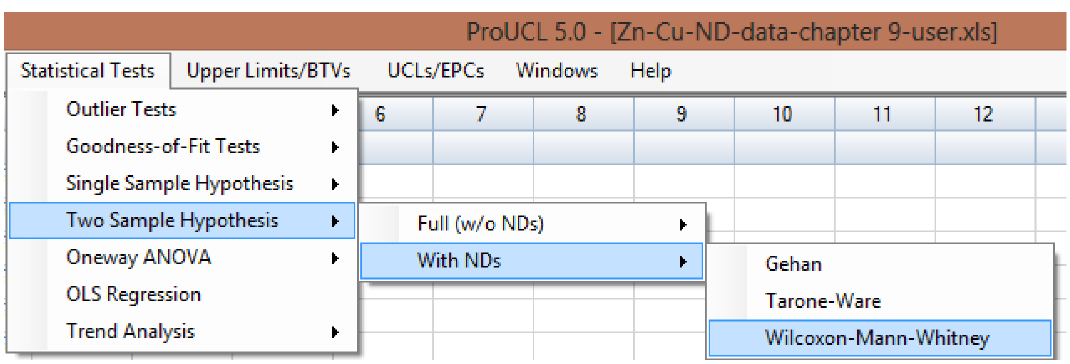


2. Select the **With NDs** option. A list of available tests will appear (shown above).

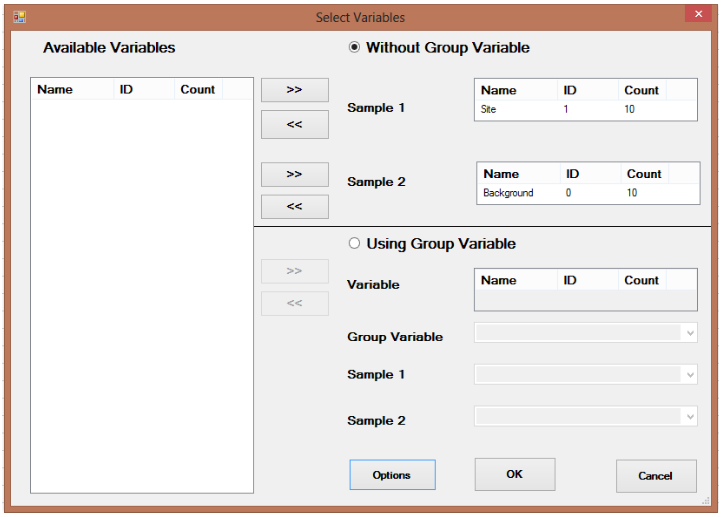
- To perform a Wilcoxon-Mann-Whitney test, click on **Wilcoxon-Mann-Whitney** from the drop-down menu list.
- To perform a Gehan test, click on **Gehan** from the drop-down menu.
- To perform a Tarone-Ware test, click on **Tarone-Ware** from the drop-down menu.

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

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



2. The **Select Variables** Screen shown below will appear.

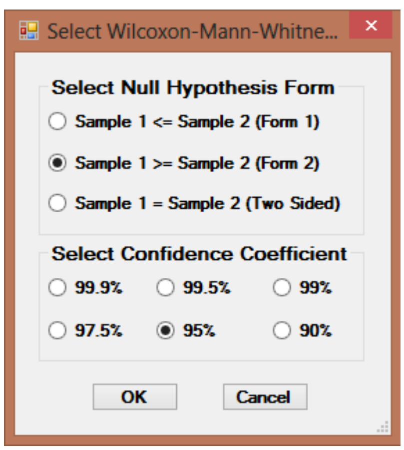


- Select variable(s) from the **Select Variables** screen.

- **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, ..., and so on.

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



  - 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 9-5. Consider a two sample data set with nondetects and multiple detection limits. Since the data sets have more than one detection limit, therefore it is not recommended to use the WMW test on this data set. However, sometimes, the users tend to use the WMW test on data sets with multiple detection limits. The WMW test results are summarized as follows:

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

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

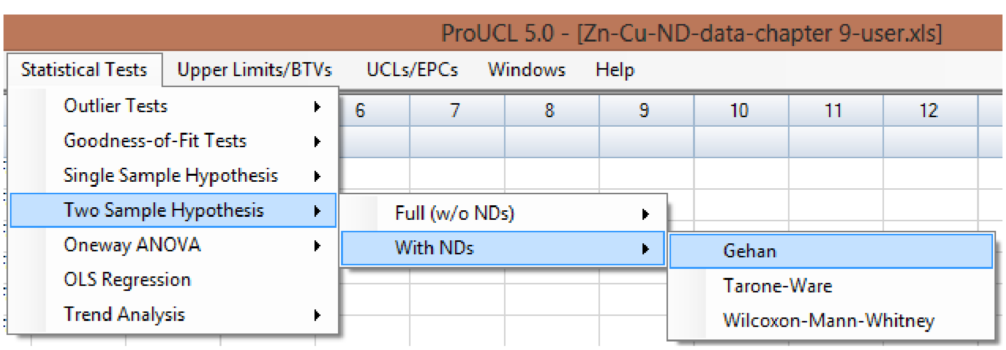
Sample 1 Rank Sum W-Stat	144.5
WMW U-Stat	78.5
Mean (U)	60.5
SD(U) - Adj ties	15.22
WMW U-Stat Critical Value (0.05)	35
Standardized WMW U-Stat	1.191
Approximate P-Value	0.883

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

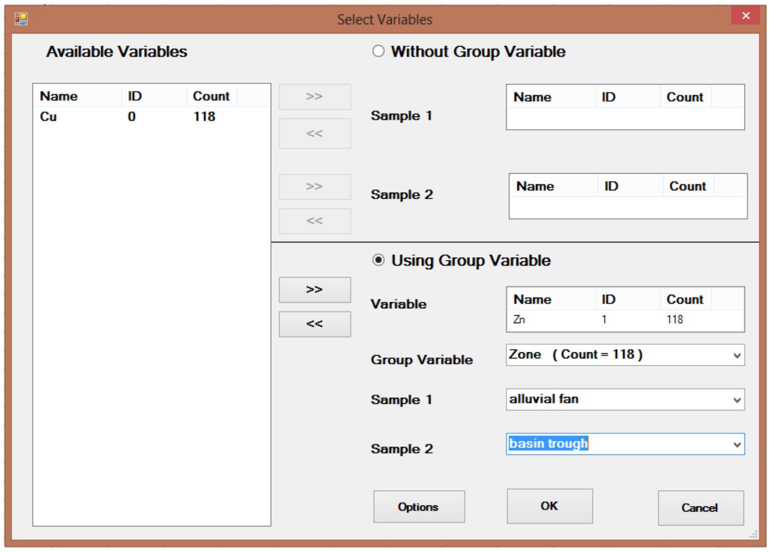
Notes**:** In the WMW test, all observations below the largest detection limit are considered as NDs (potentially including some detected values) and hence they all receive the same average rank. This action tends to reduce the associated power of the WMW test considerably. This in turn may lead to an incorrect conclusion.

9.2.2.2 Two-Sample Gehan Test for Data Sets with Nondetects

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



2. The **Select Variables** screen will appear.

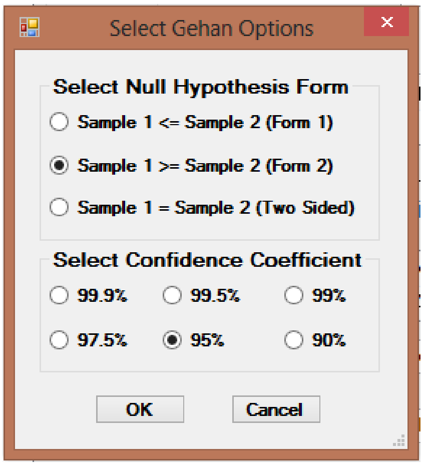


- Select variable(s) from the **Select Variables** screen.

- **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,....

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



- 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 9-6a. Consider the copper and zinc data set collected from two zones: Alluvial Fan and Basin Trough discussed in the literature (Helsel, 2012, NADA in R [2013]). This data set is used here to illustrate the Gehan two-sample test. The output sheet generated by ProUCL 5.0 is described as follows.

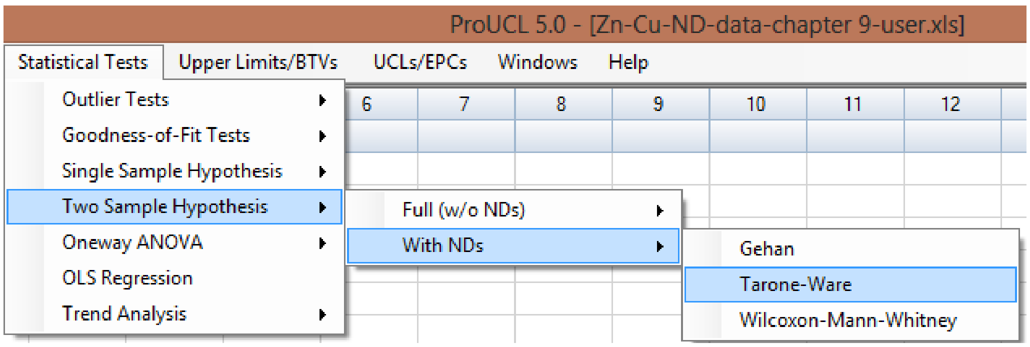
Output for Two-Sample Gehan Test (with Nondetects)

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
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		

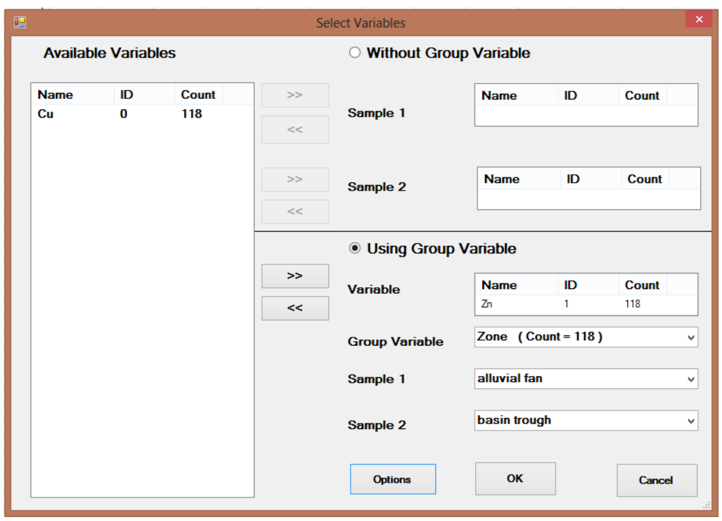
9.2.2.3 Two-Sample Tarone-Ware Test for Data Sets with Nondetects

The two-sample Tarone-Ware (T-W) test (1978) for data sets with NDs is new in ProUCL 5.0.

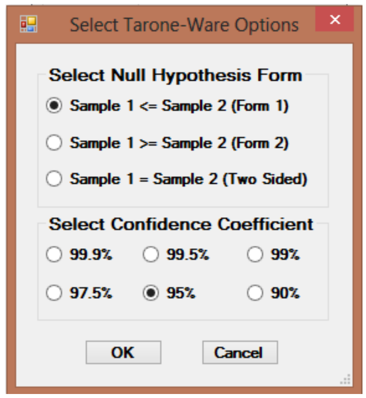
1. Click **Two Sample Hypothesis Testing ► Two Sample ► With NDs ► Tarone-Ware**



2. The **Select Variables** screen will appear.



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



- 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 9-6b (continued). Consider the copper and zinc data set used earlier. The data set is used here to illustrate the T-W two-sample test. The output sheet generated by ProUCL 5.0 is described as follows.

Output for Two-Sample Tarone-Ware Test (with Nondetects)

Confidence Coefficient	95%	
Selected Null Hypothesis	Sample 1 Mean/Median <= Sample 2 Mean/Median (Form 1)	
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
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	0.983	
Conclusion with Alpha = 0.05		
Do Not Reject H0, Conclude Sample 1 <= Sample 2		
P-Value >= alpha (0.05)		

Chapter 10

Computing Upper Limits to Estimate Background Threshold Values Based Upon Full Uncensored Data Sets and Left-Censored Data Sets with Nondetects

This chapter illustrates the computations of the various parametric and nonparametric statistics and upper limits that can be used as estimates of background threshold values (BTVs) and other not-to-exceed values. The BTV estimation methods are available for data sets with and without nondetect (ND) observations. Technical details about the computation of the various limits can be found in the associated ProUCL 5.0 Technical Guide. For each selected variable, this option computes various upper limits such as upper prediction limits (UPLs), upper tolerance limits (UTLs), upper simultaneous limits (USLs) and upper percentiles to estimate the BTVs that are used in site versus background evaluations.

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

- 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 nondetected observations with multiple detection limits.

The user specifies the confidence coefficient (probability) associated with each interval estimate. ProUCL accepts a confidence coefficient value in the interval (0.5, 1), 0.5 inclusive. The default choice is 0.95. For data sets with and without NDs, ProUCL 5.0 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.

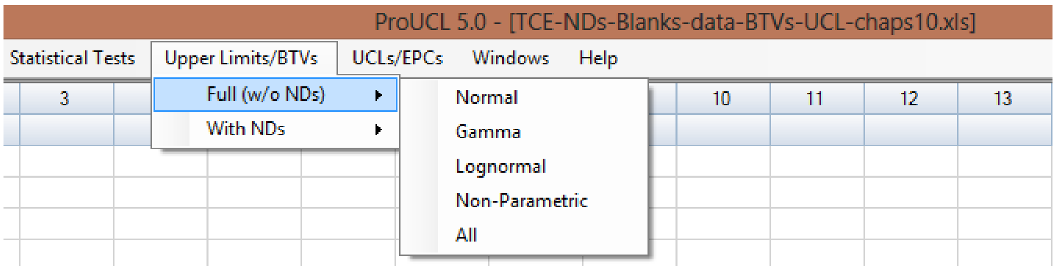
- Parametric and nonparametric USLs.

<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 nondetects, ProUCL outputs the 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 nondetects, 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 nondetect* 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.

10.1 Background Statistics for Full Data Sets without Nondetects

1. Click **Upper Limits/BTVs ► Full (w/o NDs)**

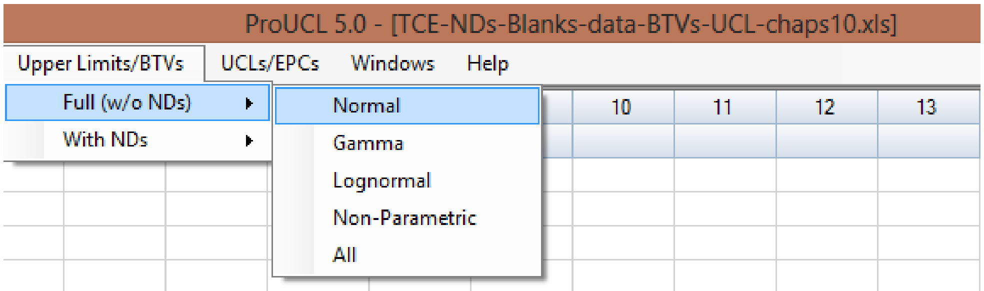


2. Select **Full (w/o NDs)**

- To compute the background statistics assuming the normal distribution, click on **Normal** from the drop-down menu list.
- To compute the background statistics assuming the gamma distribution, click on **Gamma** from the drop-down menu list.
- To compute the background statistics assuming the lognormal distribution, click on **Lognormal** from the drop-down menu list.
- To compute the background statistics using distribution-free nonparametric methods, click on **Non-Parametric** from the drop-down menu list.
- To compute and see all background statistics available in ProUCL 5.0, click on the **All** option from the drop-down menu list. ProUCL will display data distribution, all parametric and nonparametric background statistics in an Excel type spreadsheet. The user may use this output sheet to select the most appropriate statistic to estimate a BTV.

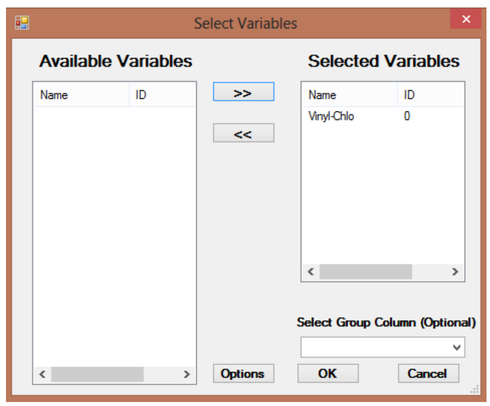
10.1.1 Normal or Lognormal Distribution

1. Click **Upper Limits/BTVs ► Full (w/o NDs) ► Normal or Lognormal**

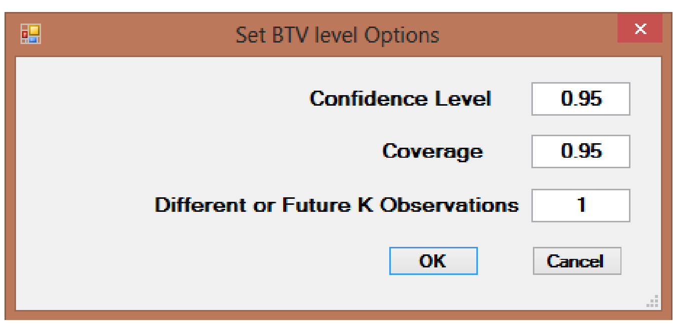


2. The **Select Variables** screen (Chapter 3) will appear.

- Select a variable(s) from the **Select Variables** screen.
- To compute BTV estimates by a group variable, 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 an appropriate group variable.



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



- Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.
- Specify the **Coverage** coefficient (for a percentile) needed to compute UTLs. **Coverage** represents a number in the interval (0.0, 1). The default choice is **0.95**. Remember, a UTL is an upper confidence limit (e.g., with confidence level = 0.95) for a 95% (e.g., with coverage = 0.95) percentile.

- Specify the **Different or Future K Observations**. The default choice is **1**. It is noted that when K = 1, the resulting interval will be a UPL for a single future observation. In the example shown above, a value of K = **1** has been used.

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

- Click on **OK** to continue or on **Cancel** button to cancel the Upper Limits/BTVs options.

Example 10-1a. Consider the real data set consisting of concentrations of several metals collected from a Superfund site. Aluminum concentrations follow a normal distribution and manganese concentrations follow a lognormal distribution. The normal and lognormal distribution based estimates of BTVs are summarized in the following two tables.

Aluminum - Output Screen for BTV Estimates Based upon a Normal Distribution (Full - Uncensored Data Set)

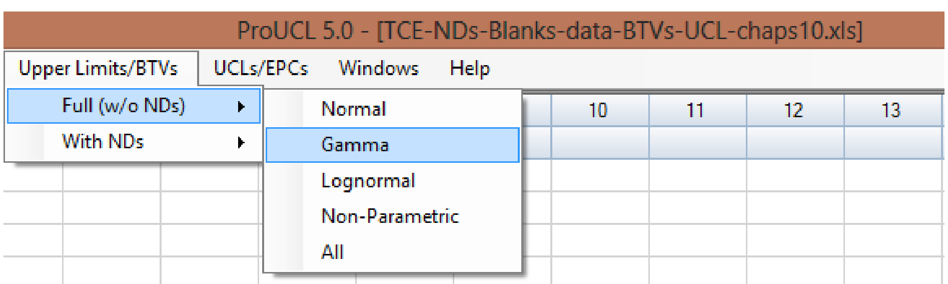
User Selected Options			
Date/Time of Computation	3/18/2013 9:26:21 PM		
From File	SuperFund.xls		
Full Precision	OFF		
Confidence Coefficient	95%		
Coverage	95%		
New or Future K Observations	1		
Aluminum			
General Statistics			
Total Number of Observations	24	Number of Distinct Observations	24
Minimum	1710	First Quartile	4058
Second Largest	15400	Median	7010
Maximum	16200	Third Quartile	10475
Mean	7789	SD	4264
Coefficient of Variation	0.547	Skewness	0.542
Mean of logged Data	8.798	SD of logged Data	0.61
Critical Values for Background Threshold Values (BTVs)			
Tolerance Factor K (For UTL)	2.309	d2max (for USL)	2.644
Normal GOF Test			
Shapiro Wilk Test Statistic	0.939	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.916	Data appear Normal at 5% Significance Level	
Lilliefors Test Statistic	0.109	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.181	Data appear Normal at 5% Significance Level	
Data appear Normal at 5% Significance Level			
Background Statistics Assuming Normal Distribution			
95% UTL with 95% Coverage	17635	90% Percentile (z)	13254
95% UPL (t)	15248	95% Percentile (z)	14803
95% USL	19063	99% Percentile (z)	17708

Manganese -Output Screen for BTV Estimates Based upon a Lognormal Distribution (Full-Uncensored Data Set)

	Lognormal Background Statistics for Uncensored Full Data Sets		
User Selected Options			
From File	SuperFund.xls		
Full Precision	OFF		
Confidence Coefficient	95%		
Coverage	95%		
New or Future K Observations	1		
Number of Bootstrap Operations	2000		
Manganese			
General Statistics			
Total Number of Observations	24	Number of Distinct Observations	23
Minimum	8.6	First Quartile	29.3
Second Largest	440	Median	71.25
Maximum	530	Third Quartile	122.5
Mean	113.8	SD	134.5
Coefficient of Variation	1.181	Skewness	2.17
Mean of logged Data	4.192	SD of logged Data	1.084
Critical Values for Background Threshold Values (BTVs)			
Tolerance Factor K (For UTL)	2.309	d2max (for USL)	2.644
Lognormal GOF Test			
Shapiro Wilk Test Statistic	0.972	**Shapiro Wilk Lognormal GOF Test**	
5% Shapiro Wilk Critical Value	0.916	Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.12	**Lilliefors Lognormal GOF Test**	
5% Lilliefors Critical Value	0.181	Data appear Lognormal at 5% Significance Level	
Data appear Lognormal at 5% Significance Level			
Background Statistics assuming Lognormal Distribution			
95% UTL with 95% Coverage	808.1	90% Percentile (z)	265.4
95% UPL (t)	440.6	95% Percentile (z)	393.5
95% USL	1162	99% Percentile (z)	823.5

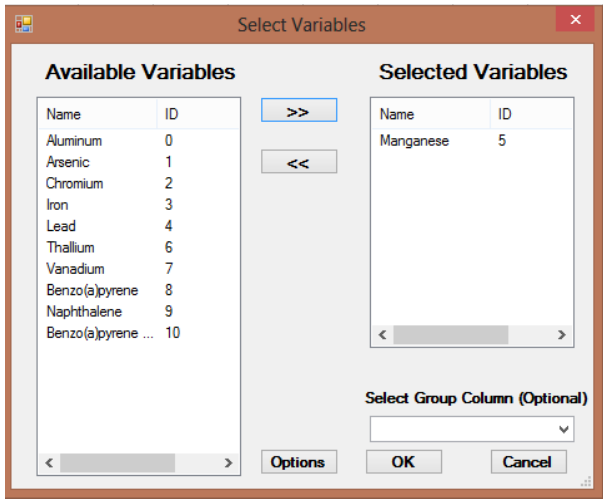
10.1.2 Gamma Distribution

1. Click **Upper Limits/BTVs ► Full (w/o NDs) ► Gamma**



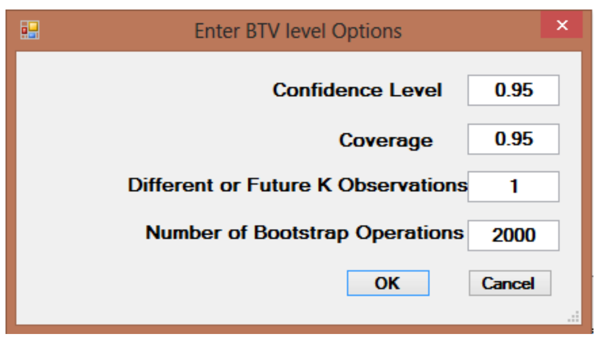
2. The **Select Variables** screen (Chapter 3) will appear.

- 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.



  - Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.

  - Specify the **Coverage** level; a number in interval (0.0, 1). Default choice is **0.95**.

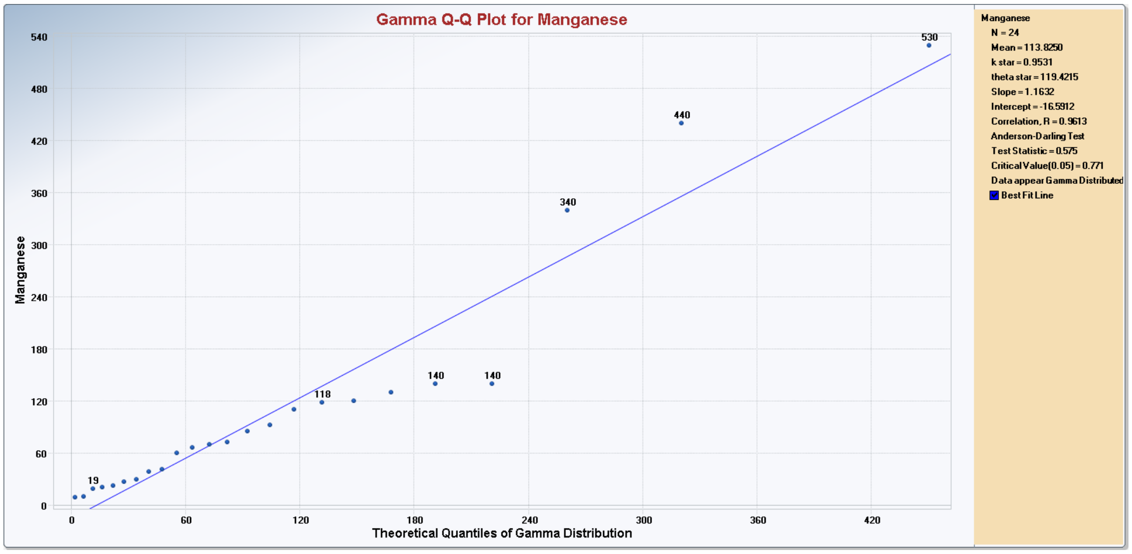
  - Specify the **Future K.** The default choice is **1.**

  - Specify the **Number of Bootstrap Operations**. The default choice is **2000**.

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

- Click on **OK** to continue or on **Cancel** button to cancel the Upper Limits/BTVs options.

Example 10-1b (continued). Manganese concentrations also follow a gamma distribution. The gamma distribution based BTV estimates are summarized in the following table generated by ProUCL 5.0. The Gamma GOF test is shown in the following figure.



Gamma GOF Test for Manganese Data Set

Manganese - Output Screen for BTV Estimates Based Upon a Gamma Distribution (Full-Uncensored Data Set)

Manganese			
General Statistics			
Total Number of Observations	24	Number of Distinct Observations	23
Minimum	8.6	First Quartile	29.3
Second Largest	440	Median	71.25
Maximum	530	Third Quartile	122.5
Mean	113.8	SD	134.5
Coefficient of Variation	1.181	Skewness	2.17
Mean of logged Data	4.192	SD of logged Data	1.084
Critical Values for Background Threshold Values (BTVs)			
Tolerance Factor K (For UTL)	2.309	d2max (for USL)	2.644
Gamma GOF Test			
A-D Test Statistic	0.575	**Anderson-Darling Gamma GOF Test**	
5% A-D Critical Value	0.771	Detected data appear Gamma Distributed at 5% Significance Level	
K-S Test Statistic	0.168	**Kolmogrov-Smirnoff Gamma GOF Test**	
5% K-S Critical Value	0.183	Detected data appear Gamma Distributed at 5% Significance Level	
Detected data appear Gamma Distributed at 5% Significance Level			
Gamma Statistics			
k hat (MLE)	1.058	k star (bias corrected MLE)	0.953
Theta hat (MLE)	107.6	Theta star (bias corrected MLE)	119.4
nu hat (MLE)	50.76	nu star (bias corrected)	45.75
MLE Mean (bias corrected)	113.8	MLE Sd (bias corrected)	116.6
Background Statistics Assuming Gamma Distribution			
95% Wilson Hilferty (WH) Approx. Gamma UPL	353.6	90% Percentile	265.2
95% Hawkins Wixley (HW) Approx. Gamma UPL	364.2	95% Percentile	346.8
95% WH Approx. Gamma UTL with 95% Coverage	503.8	99% Percentile	537
95% HW Approx. Gamma UTL with 95% Coverage	540.3		
95% WH USL	611.3	95% HW USL	672.3

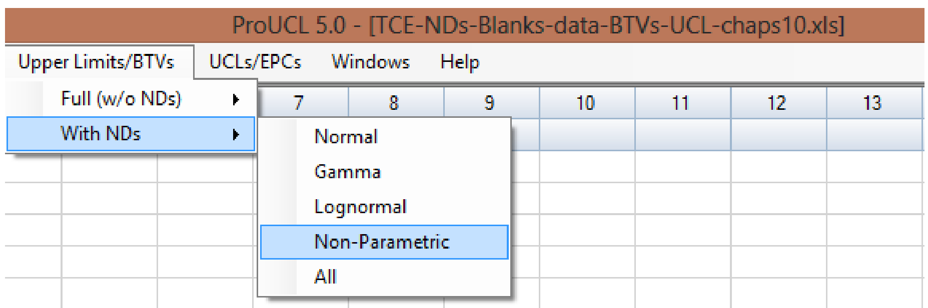
The mean manganese concentration is 113.8 with sd = 134.5, and the maximum value = 530. The UTL based upon a lognormal distribution is 808.1 which is significantly higher than the largest value of 530. It is noted that the *sd* of the log-transformed data is 1.084. By comparing BTV estimates computed using lognormal and gamma distributions, it is noted that the lognormal distribution based upper limits: UTL and UPL are significantly higher than those based upon a gamma distribution confirming the statements made earlier that the use of a lognormal distribution tends to yield inflated values of the upper limits used to estimate environmental parameters (e.g., BTVs, EPCs). These upper limits are summarized as follows.

	Lognormal	**Gamma (WH)**	**Gamma (HW)**
UTL95-95	808.1	504	540.3
UPL95	440.6	353.6	364.2

Mean = 113.8, Max value = 530.

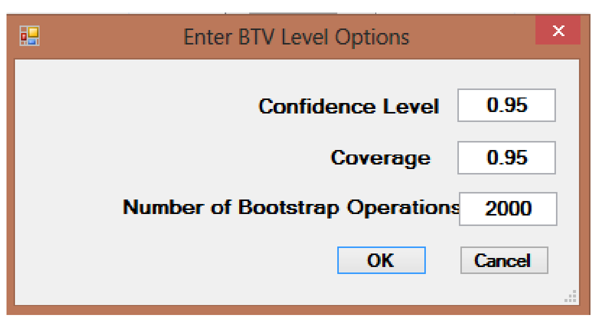
10.1.3 Nonparametric Methods

1. Click **Upper Limits/BTVs ► Full (w/o NDs) ► Non-Parametric**



2. The **Select Variables** screen (Chapter 3) 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.



- Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.

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

- Click **OK** button to continue or **Cancel** button to cancel the Upper Limits/BTVs options.

Example 10-2. Lead concentrations collected from the same Superfund site as used in Example 10-1 do not follow a discernible distribution. Nonparametric BTV estimates are summarized as follows.

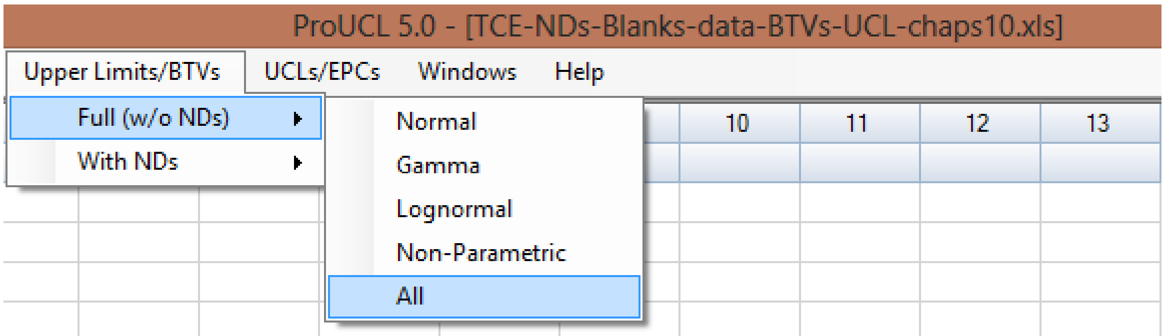
Lead - Output Screen for Nonparametric BTVs Estimates (Full-Uncensored Data Set)

	Nonparametric Background Statistics for Uncensored Full Data Sets		
User Selected Options			
From File	SuperFund.xls		
Full Precision	OFF		
Confidence Coefficient	95%		
Coverage	95%		
Number of Bootstrap Operations	2000		
Lead			
General Statistics			
Total Number of Observations	24	Number of Distinct Observations	18
Minimum	4.9	First Quartile	10.43
Second Largest	98.5	Median	14
Maximum	109	Third Quartile	19.25
Mean	22.49	SD	26.83
Coefficient of Variation	1.193	Skewness	2.665
Mean of logged Data	2.743	SD of logged Data	0.771
Critical Values for Background Threshold Values (BTVs)			
Tolerance Factor K (For UTL)	2.309	d2max (for USL)	2.644
Nonparametric Distribution Free Background Statistics			
Data do not follow a Discernible Distribution (0.05)			
Nonparametric Upper Limits for Background Threshold Values			
Order of Statistic, r	24	95% UTL with 95% Coverage	109
Approximate f	1.263	Confidence Coefficient (CC) achieved by UTL	0.708
95% Percentile Bootstrap UTL with 95% Coverage	109	95% BCA Bootstrap UTL with 95% Coverage	107.4
95% UPL	106.4	90% Percentile	44.81
90% Chebyshev UPL	104.6	95% Percentile	91.72
95% Chebyshev UPL	141.8	99% Percentile	106.6
95% USL	109		

To compute nonparametric upper limits providing the specified coverage (e.g., 0.95), sizes of the data sets should be fairly large (e.g., > 59). For details, consult the associated ProUCL 5.0 Technical Guide. In this example the sample size is only 24, and the confidence coefficient (CC) achieved by the nonparametric, UTL is only 0.71 which is significantly lower than the desired CC of 0.95.

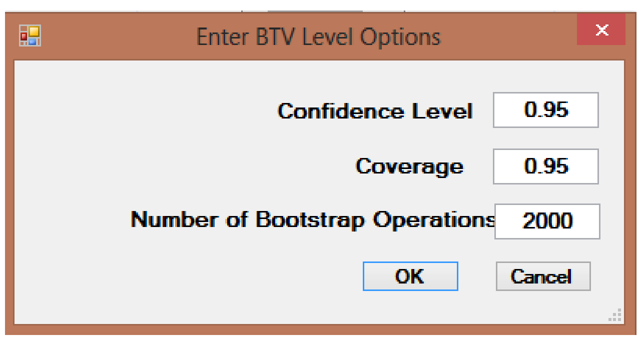
10.1.4 All Statistics Option

1. Click **Upper Limits/BTVs ► Full (w/o NDs) ► All**



2. The **Select Variables** screen (Chapter 3) 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.



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

- Click on **OK** to continue or on **Cancel** button to cancel the Upper Limits/BTVs options.

Example 10-1c (continued). The various BTV estimates based upon the manganese concentrations computed using the **All** option of ProUCL are summarized as follows. The **All** option computes and displays all available parametric and nonparametric BTV estimates. This option also informs the user about the distribution(s) of the data set. This option is specifically useful when one has to process many analytes (variables) without any knowledge about their probability distributions.

Manganese - Output Screen for All BTVs Estimates (Full-Uncensored Data Set)

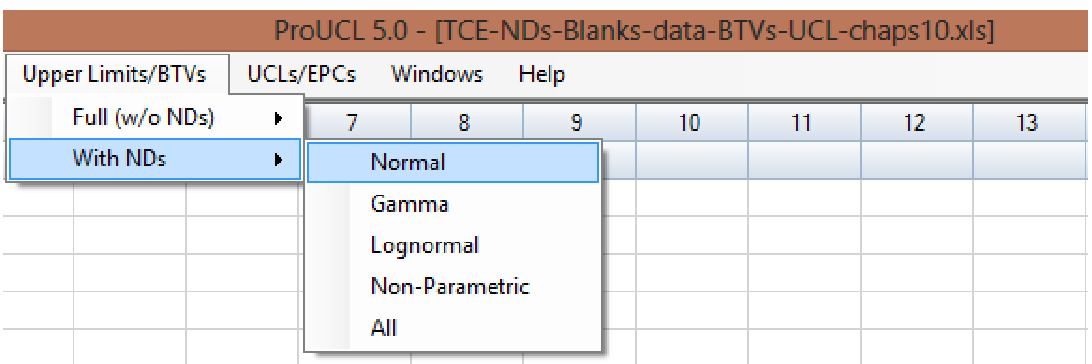
Manganese			
General Statistics			
Total Number of Observations	24	Number of Distinct Observations	23
Minimum	8.6	First Quartile	29.3
Second Largest	440	Median	71.25
Maximum	530	Third Quartile	122.5
Mean	113.8	SD	134.5
Coefficient of Variation	1.181	Skewness	2.17
Mean of logged Data	4.192	SD of logged Data	1.084
Critical Values for Background Threshold Values (BTVs)			
Tolerance Factor K (For UTL)	2.309	d2max (for USL)	2.644
Normal GOF Test			
Shapiro Wilk Test Statistic	0.697	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.916	Data Not Normal at 5% Significance Level	
Lilliefors Test Statistic	0.298	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.181	Data Not Normal at 5% Significance Level	
Data Not Normal at 5% Significance Level			
Background Statistics Assuming Normal Distribution			
95% UTL with 95% Coverage	424.3	90% Percentile (z)	286.1
95% UPL (t)	349	95% Percentile (z)	335
95% USL	469.3	99% Percentile (z)	426.6

Gamma GOF Test			
A-D Test Statistic	0.575	**Anderson-Darling Gamma GOF Test**	
5% A-D Critical Value	0.771	Detected data appear Gamma Distributed at 5% Significance Level	
K-S Test Statistic	0.168	**Kolmogrov-Smirnoff Gamma GOF Test**	
5% K-S Critical Value	0.183	Detected data appear Gamma Distributed at 5% Significance Level	
Detected data appear Gamma Distributed at 5% Significance Level			
Gamma Statistics			
k hat (MLE)	1.058	k star (bias corrected MLE)	0.953
Theta hat (MLE)	107.6	Theta star (bias corrected MLE)	119.4
nu hat (MLE)	50.76	nu star (bias corrected)	45.75
MLE Mean (bias corrected)	113.8	MLE Sd (bias corrected)	116.6
Background Statistics Assuming Gamma Distribution			
95% Wilson Hilferty (WH) Approx. Gamma UPL	353.6	90% Percentile	265.2
95% Hawkins Wixley (HW) Approx. Gamma UPL	364.2	95% Percentile	346.8
95% WH Approx. Gamma UTL with 95% Coverage	503.8	99% Percentile	537
95% HW Approx. Gamma UTL with 95% Coverage	540.3		
95% WH USL	611.3	95% HW USL	672.3

Lognormal GOF Test			
Shapiro Wilk Test Statistic	0.972	**Shapiro Wilk Lognormal GOF Test**	
5% Shapiro Wilk Critical Value	0.916	Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.12	**Lilliefors Lognormal GOF Test**	
5% Lilliefors Critical Value	0.181	Data appear Lognormal at 5% Significance Level	
Data appear Lognormal at 5% Significance Level			
Background Statistics assuming Lognormal Distribution			
95% UTL with 95% Coverage	808.1	90% Percentile (z)	265.4
95% UPL (t)	440.6	95% Percentile (z)	393.5
95% USL	1162	99% Percentile (z)	823.5
Nonparametric Distribution Free Background Statistics			
Data appear Gamma Distributed at 5% Significance Level			
Nonparametric Upper Limits for Background Threshold Values			
Order of Statistic, r	24	95% UTL with 95% Coverage	530
Approximate f	1.263	Confidence Coefficient (CC) achieved by UTL	0.708
95% Percentile Bootstrap UTL with 95% Coverage	530	95% BCA Bootstrap UTL with 95% Coverage	530
95% UPL	507.5	90% Percentile	280
90% Chebyshev UPL	525.5	95% Percentile	425
95% Chebyshev UPL	712	99% Percentile	509.3
95% USL	530		

10.2 Background Statistics with NDs

1. Click **Upper Limits/BTVs** ► **With NDs**

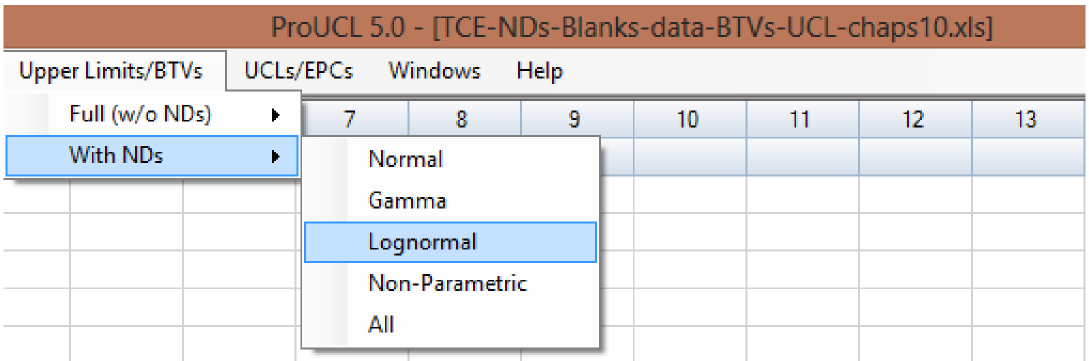


2. Select the **With NDs** option.

- To compute the background statistics assuming the normal distribution, click on **Normal** from the drop-down menu list.
- To compute the background statistics assuming the gamma distribution, click on **Gamma** from the drop-down menu list.
- To compute the background statistics assuming the lognormal distribution, click on **Lognormal** from the drop-down menu list.
- To compute the background statistics using distribution-free methods, click on **Non-Parametric** from the drop-down menu list.
- To compute all available background statistics in ProUCL 5.0, click on the **All** option from the drop-down menu list.

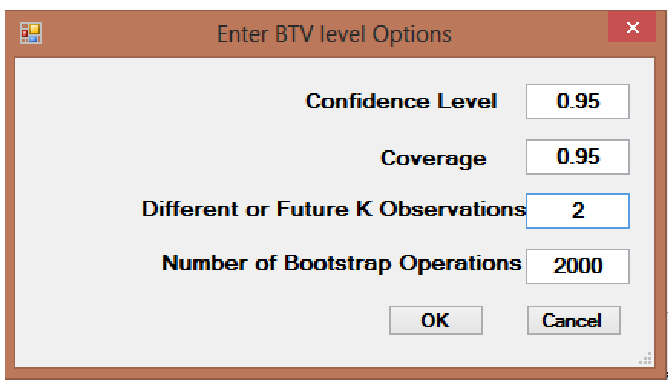
10.2.1 Normal or Lognormal Distribution

1. Click **Upper Limits/BTVs ► With NDs ► Normal or Lognormal**



2. The **Select Variables** screen (Chapter 3) 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.



  - 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**.
  - Specify the **Number of Bootstrap Operations.** The default choice is **2000.**
  - Click on the **OK** button to continue or on the **Cancel** button to cancel the option.

- Click on **OK** to continue or on **Cancel** button to cancel the Upper limits/BTVs options.

Example 10 -3a. Consider a small real TCE data set of size *n*=12 consisting of 4 ND observations. The detected data set of size 8 follows a normal as well as a lognormal distribution. The BTV estimates using the LROS method, normal and lognormal distribution on KM estimates, and nonparametric Chebyshev inequality and bootstrap methods on KM estimates are summarized in the following two tables. It is noted that upper limits including UTL95-95 and UPL95 based upon the robust LROS method yield much higher values than the other methods including KM estimates in normal and lognormal equations to compute the upper limits. It is noted that the detected data also follows a gamma distribution. The gamma distribution (of detected data) based BTV estimates are described in the next section.

TCE - Output Screen for BTV Estimates Computed Using Normal Distribution of Detected Data (Left-Censored Data Set with NDs)

User Selected Options	
From File	TCE-NDs-Blanks-data-BTVs-UCL-chaps.xls
Full Precision	OFF
Confidence Coefficient	95%
Coverage	95%
Different or Future K Observations	2

TCE

General Statistics			
Total Number of Observations	12	Number of Distinct Observations	9
Number of Missing Observations	2		
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**	
5% Shapiro Wilk Critical Value	0.818	Data Not Normal at 5% Significance Level	
Lilliefors Test Statistic	0.256	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.313	Detected Data appear Normal at 5% Significance Level	
Detected Data appear Approximate Normal at 5% Significance Level			
Kaplan Meier (KM) Background Statistics Assuming Normal Distribution			
Mean	2.188	SD	2.61
95% UTL95% Coverage	9.329	95% KM UPL (t)	7.067
95% KM UPL for Next 2 Observations	8.167	95% KM UPL for Mean of Next 2 Observations	5.768
95% KM Chebyshev UPL	14.03	90% KM Percentile (z)	5.533
95% KM Percentile (z)	6.481	99% KM Percentile (z)	8.26

Output Screen for BTV Estimates Computed Using a Lognormal Distribution of Detected Data (Left-Censored Data Set with NDs)

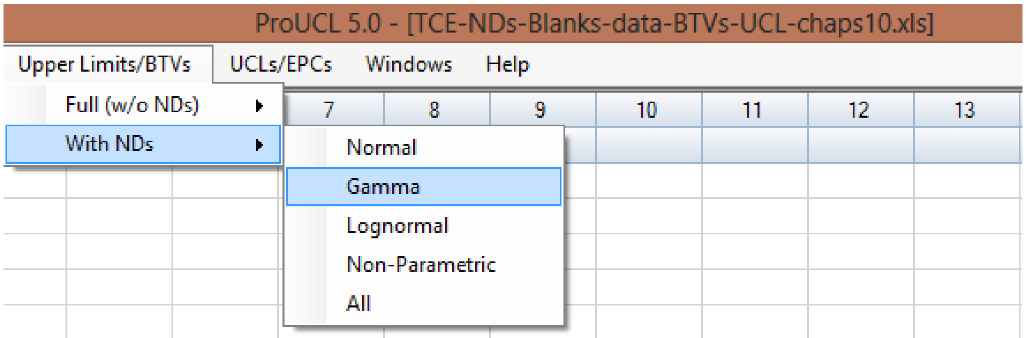
	Lognormal Background Statistics for Data Sets with Non-Detects
User Selected Options	
From File	TCE-NDs-Blanks-data-BTVs-UCL-chaps10.xls
Full Precision	OFF
Confidence Coefficient	95%
Coverage	95%
Different or Future K Observations	2
Number of Bootstrap Operations	2000

TCE

General Statistics			
Total Number of Observations	12	Number of Distinct Observations	9
Number of Missing Observations	2		
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
Lognormal GOF Test on Detected Observations Only			
Shapiro Wilk Test Statistic	0.865	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.818	Detected Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.258	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.313	Detected Data appear Lognormal at 5% Significance Level	
Detected Data appear Lognormal at 5% Significance Level			
Kaplan Meier (KM) Background Statistics Assuming Normal Distribution			
Mean	2.188	SD	2.61
95% UTL95% Coverage	9.329	95% KM UPL (t)	7.067
95% KM UPL for Next 2 Observations	8.167	95% KM UPL for Mean of Next 2 Observations	5.768
95% KM Chebyshev UPL	14.03	90% KM Percentile (z)	5.533
95% KM Percentile (z)	6.481	99% KM Percentile (z)	8.26
95% KM USL	8.152		
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
95% UPL for Next 2 Observations	25.78	95% UPL for Mean of 2 Observations	6.424
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

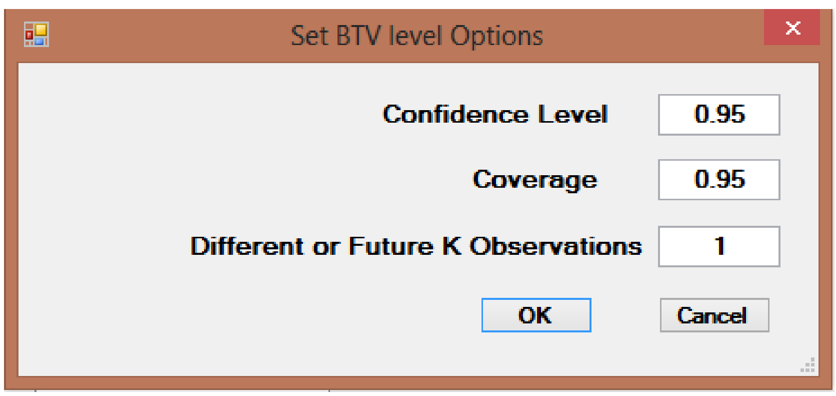
10.2.2 Gamma Distribution

1. Click **Upper Limits/BTVs ► With NDs ► Gamma**



2. The **Select Variables** screen (Chapter 3) 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.



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

- Click on **OK** to continue or on **Cancel** button to cancel the **Upper Limits/BTVs** options.

Example 10-3b (continued). It is noted that the detected TCE data considered in Example 10-3 also follows a gamma distribution. The gamma distribution based upper limits are summarized as follows.

TCE - Output Screen for BTV Estimates Computed Using Gamma Distribution of Detected Data (Left-Censored Data Set with NDs)

	Gamma Background Statistics for Data Sets with Non-Detects
User Selected Options	
From File	TCE-NDs-Blanks-data-BTVs-UCL-chaps10.xls
Full Precision	OFF
Confidence Coefficient	95%
Coverage	95%
TCE	

	General Statistics		
Total Number of Observations	12	Number of Distinct Observations	9
Number of Missing Observations	2		
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
	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	**Kolmogrov-Smirnoff 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 (2k)	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 detected data is small such as < 0.1		
	For such situations, GROS method tends to yield inflated values of UCLs and BTVs		
	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 (2k)	2.956	90% Percentile	5.709
95% Percentile	8.668	99% Percentile	16.26

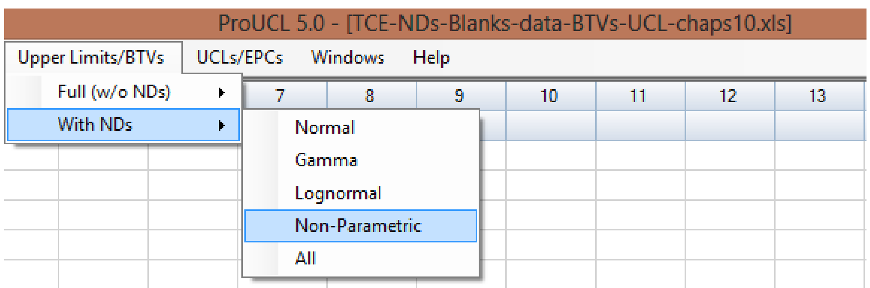
		The following statistics are computed using Gamma ROS Statistics on Imputed Data			
		Upper Limits using Wilson Hilferty (WH) and Hawkins Wixley (HW) Methods			
	WH	HW		WH	HW
95% Approx. Gamma UTL with 95% Coverage	19.62	27.19	95% Approx. Gamma UPL	9.793	11.66
95% Gamma USL	13.95	17.89	95% UPL for Next 2 Observations	14.01	18
		The following statistics are computed using gamma distribution and KM estimates			
		Upper Limits using Wilson Hilferty (WH) and Hawkins Wixley (HW) Methods			
	k hat (KM)	0.702		nu hat (KM)	16.86
	WH	HW		WH	HW
95% Approx. Gamma UTL with 95% Coverage	11.34	11.95	95% Approx. Gamma UPL	6.88	6.896
95% Gamma USL	8.836	9.063			

Kaplan Meier (KM) Background Statistics Assuming Normal Distribution			
Mean	2.188	SD	2.61
95% UTL95% Coverage	9.329	95% KM UPL (t)	7.067
95% KM UPL for Next 2 Observations	8.167	95% KM UPL for Mean of Next 2 Observations	5.768
95% KM Chebyshev UPL	14.03	90% KM Percentile (z)	5.533
95% KM Percentile (z)	6.481	99% KM Percentile (z)	8.26
95% KM USL	8.152		

The detected data set does not follow a normal distribution based upon the S-W test, but follows a normal distribution based upon the Lilliefors test. Since the detected data set is of small size (=8), the normal GOF conclusion is suspect. The detected data follow a gamma distribution. There 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 to compute the various 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.

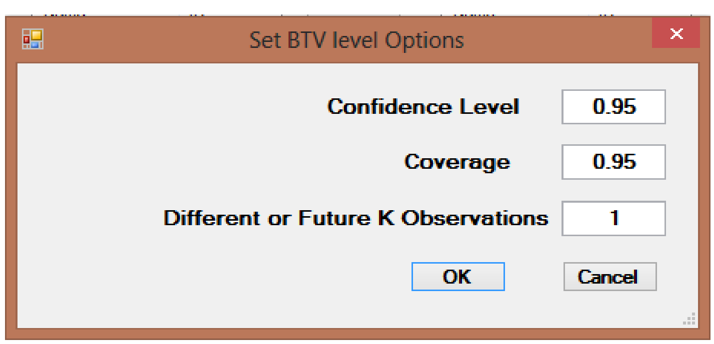
10.2.3 Nonparametric Methods (with NDs)

1. Click **Upper Limits/BTVs► With NDs ► Non-Parametric**



2. The **Select Variables** screen (Chapter 3) 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.



- Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.

- Specify the **Coverage** level; a number in interval (0.0, 1). Default choice is **0.95**.

- 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 Limit/BTVs option.

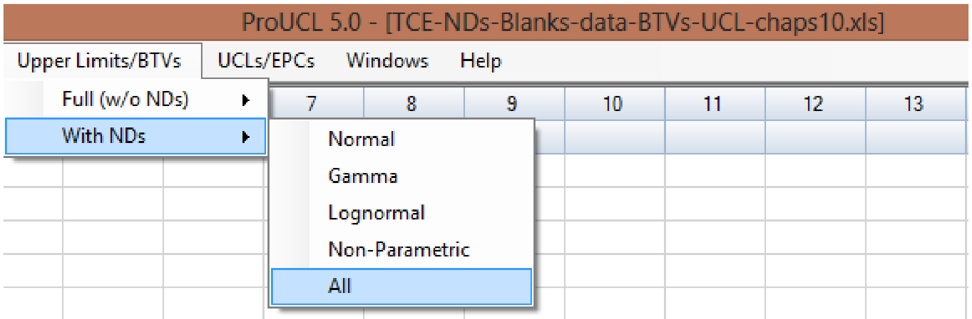
Example 10-3c (continued). The nonparametric upper limits based the TCE data considered in Example 10-3 are summarized in the following table.

TCE - Output Screen for Nonparametric BTV Estimates (Left-Censored Data Set with NDs)

TCE			
General Statistics			
Total Number of Observations	12	Number of Distinct Observations	9
Number of Missing Observations	2		
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
Nonparametric Distribution Free Background Statistics			
Data appear to follow a Discernible Distribution at 5% Significance Level			
Kaplan Meier (KM) Background Statistics Assuming Normal Distribution			
Mean	2.188	SD	2.61
95% UTL95% Coverage	9.329	95% KM UPL (t)	7.067
95% KM Chebyshev UPL	14.03	90% KM Percentile (z)	5.533
95% KM Percentile (z)	6.481	99% KM Percentile (z)	8.26
95% KM USL	8.152		
Nonparametric Uppper Limits for BTVs(no distinction made between detects and nondetects)			
Order of Statistic, r	12	95% UTL with95% Coverage	9.29
Approximate f	0.632	Confidence Coefficient (CC) achieved by UTL	0.46
95% UPL	9.29	95% USL	9.29

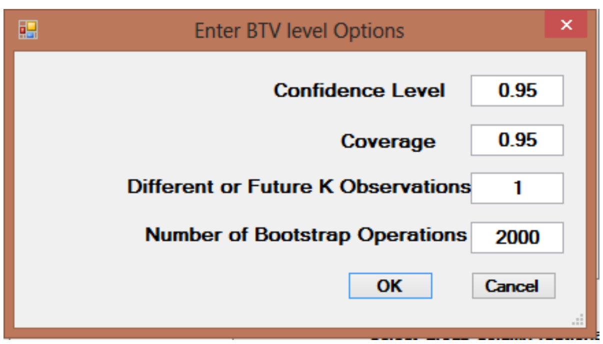
10.2.4 All Statistics Option

1. Click **Upper Limits/BTVs ► With NDs ► All**



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

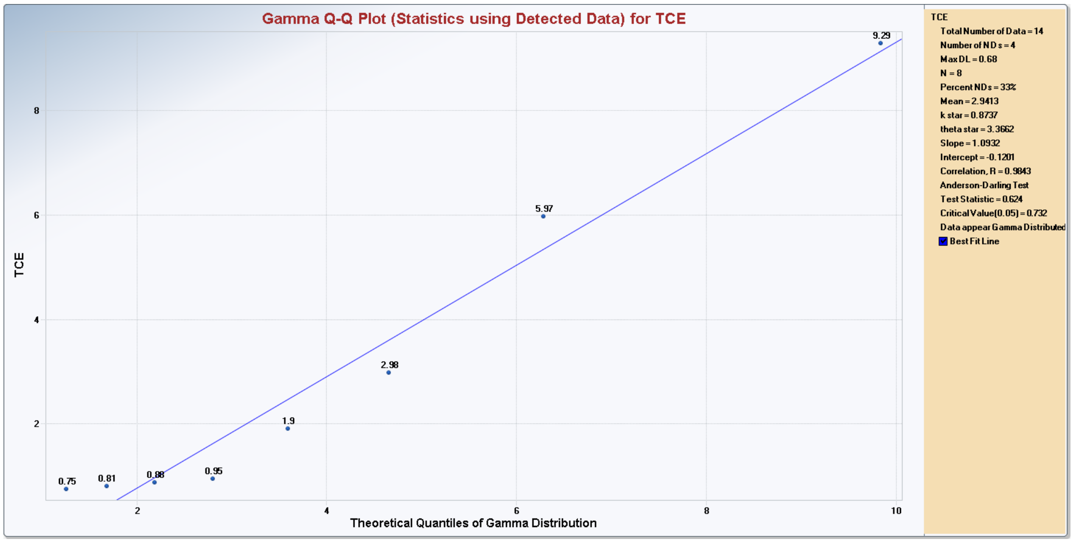
- Select a variable(s) from the **Select Variables** screen.
- If needed, select a group variable by clicking the arrow below the **Select Group Column (Optional)** to obtain a drop-down list of variables, and select a proper group variable.
- When the **Option** button is clicked, the following window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



  - 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.

Example 10-3d (continued). BTV estimates using the **All** option for the TCE data 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 high-lighted in the output table provided after the gamma GOF Q-Q plot.



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

TCE			
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**	
5% Shapiro Wilk Critical Value	0.818	Data Not Normal at 5% Significance Level	
Lilliefors Test Statistic	0.256	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.313	Detected Data appear Normal at 5% Significance Level	
Detected Data appear Approximate Normal at 5% Significance Level			
Kaplan Meier (KM) Background Statistics Assuming Normal Distribution			
Mean	2.188	SD	2.61
95% UTL95% Coverage	9.329	95% KM UPL (t)	7.067
95% KM Chebyshev UPL	14.03	90% KM Percentile (z)	5.533
95% KM Percentile (z)	6.481	99% KM Percentile (z)	8.26
95% KM USL	8.152		

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	Kolmogrov-Smirnoff 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 (2k)	5.492

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			
The following statistics are computed using gamma distribution and KM estimates					
Upper Limits using Wilson Hilferty (WH) and Hawkins Wixley (HW) Methods					
k hat (KM)	0.702		nu hat (KM)	16.86	
	WH	HW		WH	HW
95% Approx. Gamma UTL with 95% Coverage	11.34	11.95	95% Approx. Gamma UPL	6.88	6.896
95% Gamma USL	8.836	9.063			

Lognormal GOF Test on Detected Observations Only			
Shapiro Wilk Test Statistic	0.865	Shapiro Wilk GOF Test	
5% Shapiro Wilk Critical Value	0.818	Detected Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.258	Lilliefors GOF Test	
5% Lilliefors Critical Value	0.313	Detected Data appear Lognormal at 5% Significance Level	
Detected Data appear Lognormal at 5% 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

Nonparametric Uppper Limits for BTVs(no distinction made between detects and nondetects)			
Order of Statistic, r	12	95% UTL with95% Coverage	9.29
Approximate f	0.632	Confidence Coefficient (CC) achieved by UTL	0.46
95% UPL	9.29	95% USL	9.29

Note: Even though the data set failed the Shapiro-Wilk test of normality, based upon Lilliefors test it was concluded that the data set follows a normal distribution. Therefore instead of saying that the data set does not follow a normal distribution, ProUCL outputs that the data set follows an approximate normal distribution. In practice the two tests can lead to different conclusions, especially when the data set is of small size. In such instances, it is suggested that the user supplements test results with graphical displays to derive the final conclusion.

As noted, detected data follow a gamma as well as a lognormal distribution. The various upper limits using Gamma ROS and Lognormal ROS methods and Gamma and Lognormal distribution on KM estimates are summarized as follows.

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 (2012), EPA (2009)-LROS
UTL95-95 (ROS)	19.62	WH- ProUCL(ROS)	50.54	Helsel (2012), EPA (2009)-LROS
UPL95 (KM)	6.88	WH - ProUCL (KM-Gamma)	7.06	KM-Lognormal EPA (2009)
UTL95-95 (KM)	11.34	WH - ProUCL (KM-Gamma)	15.25	KM- Lognormal EPA(2009)

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.

Chapter 11

Computing Upper Confidence Limits (UCLs) of Mean Based Upon Full-Uncensored Data Sets and Left-Censored Data Sets with Nondetects

Several parametric and nonparametric UCL methods for full-uncensored and left-censored data sets consisting of ND observations with multiple detection limits (DLs) are available in ProUCL 5.0. 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 5.0 Technical Guide, Singh, Singh, and Engelhardt, 1997; Singh, Singh, and Iaci (2002); and Singh, Maichle, and Lee (USEPA, 2006).

In ProUCL 5.0, two choices are available to compute UCL statistics:

- <u>Full (w/o NDs):</u> Computes UCLs for full-uncensored data sets without any nondetects.

- <u>With NDs:</u> 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 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.

    - The user is responsible for selecting an appropriate choice for the data distribution: normal, gamma, lognormal, or nonparametric. It is desirable that user determines data distribution using the Goodness-of-Fit test option prior to using the UCL option. The UCL output sheet also informs the user if data are normal, gamma, lognormal, or a non-discernible distribution. Program computes statistics depending on the user selection.

    - For data sets, which are not normal, one may try the gamma UCL next. The program will offer you advice if you chose the wrong UCL option.

    - For data sets, which are neither normal nor gamma, one may try the lognormal UCL. The program will offer you advice if you chose the wrong UCL option.

- Data sets that are not normal, gamma, or lognormal are classified as distribution-free nonparametric data sets. The user may use nonparametric UCL option for such data sets. The program will offer you advice if you chose the wrong UCL option.

- The program also provides the **All** option. By selecting this option, 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.

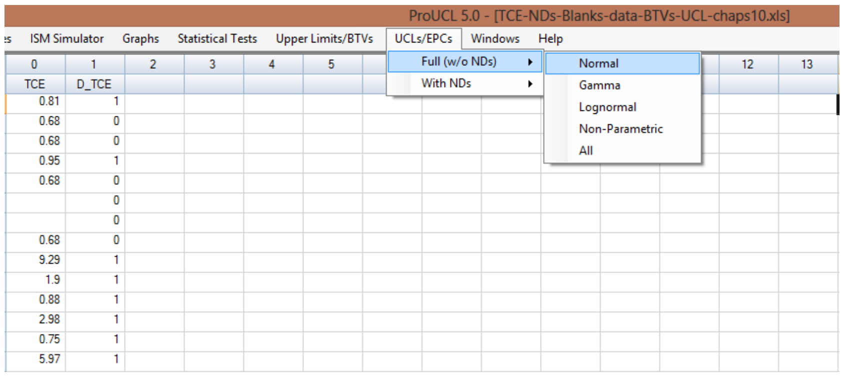
<u>Notes:</u> Like all other methods, it is recommended that the user identify a few low probability (coming from extreme tails) outlying observations that may be present in the data set. Outliers distort statistics of interest including summary statistics, data distributions, test statistics, UCLs and BTVs. Decisions based upon distorted statistics may be misleading and incorrect. The objective is to compute decision statistics based upon the majority of the data set representing the main dominant population. The project team should decide about the disposition (to include or not to include) of outliers before computing estimates the EPC terms and BTVs. To determine the influence of outliers on UCLs and background statistics, the project team may want to compute statistics twice: once using the data set with outliers, and once using the data set without outliers.

<u>Note on Computing Lower Confidence Limits (LCLs) of Mean:</u> 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 nondetects, 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 nondetect* 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.

11.1 UCLs for Full (w/o NDs) Data Sets

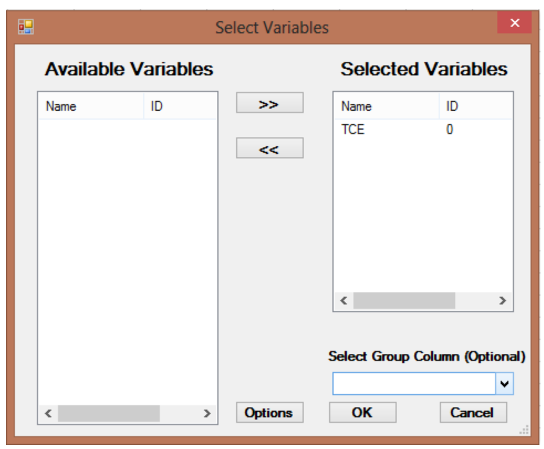
11.1.1 Normal Distribution (Full Data Sets without NDs)

1. Click **UCLs/EPCs ► Full (w/o NDs) ► Normal**

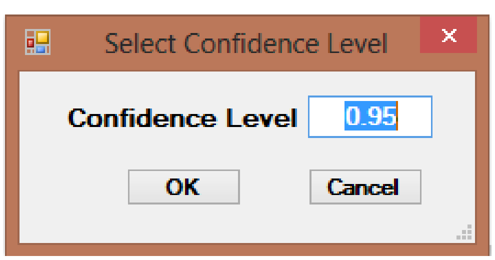


2. The **Select Variables** screen (Chapter 3) will appear.

- 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 to select a group variable.



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



- o Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.

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

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

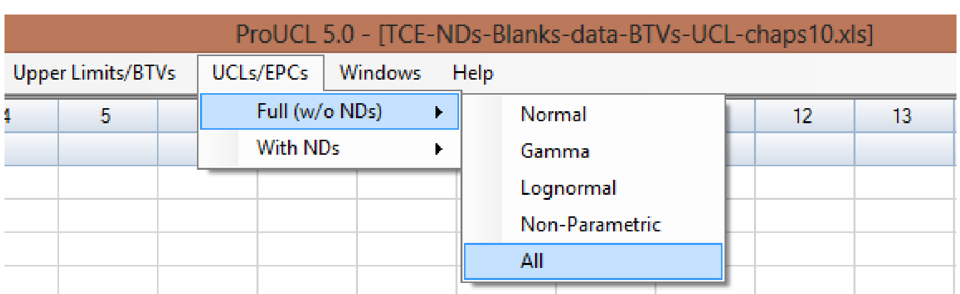
Example 11-1. Consider the real data set consisting of concentrations of several metals collected from a Superfund site; vanadium concentrations follow a normal distribution. The normal distribution based 95% UCLs of mean are summarized in the following table.

Vanadium - Output Screen for Normal Distribution (Full Data w/o NDs)

Normal UCL Statistics for Uncensored Full Data Sets			
User Selected Options			
Date/Time of Computation	3/25/2013 3:53:05 PM		
From File	SuperFund.xls		
Full Precision	OFF		
Confidence Coefficient	95%		
Vanadium			
General Statistics			
Total Number of Observations	20	Number of Distinct Observations	17
		Number of Missing Observations	0
Minimum	7.2	Mean	17.34
Maximum	32	Median	16.5
SD	8.075	SD of logged Data	0.492
Coefficient of Variation	0.466	Skewness	0.429
Normal GOF Test			
Shapiro Wilk Test Statistic	0.925	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.905	Data appear Normal at 5% Significance Level	
Lilliefors Test Statistic	0.146	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.198	Data appear Normal at 5% Significance Level	
Data appear Normal at 5% Significance Level			
Assuming Normal Distribution			
95% Normal UCL		**95% UCLs (Adjusted for Skewness)**	
95% Student's-t UCL	20.46	95% Adjusted-CLT UCL (Chen-1995)	20.49
		95% Modified-t UCL (Johnson-1978)	20.49
Suggested UCL to Use			
95% Student's-t UCL	20.46		

11.1.2 Gamma, Lognormal, Nonparametric, All Statistics Option (Full Data without NDs)

1. Click **UCLs/EPCs ► Full (w/o NDs) ► Gamma, Lognormal, Non-Parametric, or All**

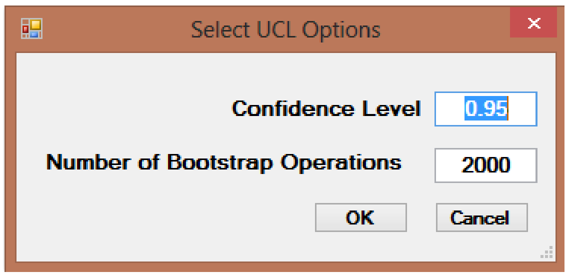


2. The **Select Variables** screen (Chapter 3) will appear.

- 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.

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



  - Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive.

  - Specify the **Number of Bootstrap Operations (runs)**. Default choice is **2000**.

  - Click on **OK** button to continue or on **Cance**l button to cancel the UCLs option.

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

Example 11-2: This skewed data set of size *n*=25 with mean=44.09 was used in Chapter 2 of the Technical Guide. The data follows a lognormal and a gamma distribution. The data are: 0.3489, 0.8526, 2.5445, 2.5602, 3.3706, 4.8911, 5.0930, 5.6408, 7.0407, 14.1715, 15.2608, 17.6214, 18.7690, 23.6804, 25.0461, 31.7720, 60.7066, 67.0926, 72.6243, 78.8357, 80.0867, 113.0230, 117.0360, 164.3302, and 169.8303. UCLs based upon **Gamma**, **Lognormal**, **Non-parametric**, and **All** options are summarized in the following tables.

Output Screen for Gamma Distribution Based UCLs (Full (w/o NDs))

X			
	General Statistics		
Total Number of Observations	25	Number of Distinct Observations	25
		Number of Missing Observations	0
Minimum	0.349	Mean	44.09
Maximum	169.8	Median	18.77
SD	51.34	SD of logged Data	1.68
Coefficient of Variation	1.164	Skewness	1.294
	Gamma GOF Test		
A-D Test Statistic	0.374	Anderson-Darling Gamma GOF Test	
5% A-D Critical Value	0.794	Data appear Gamma Distributed at 5% Significance Level	
K-S Test Statistic	0.113	Kolmogrov-Smirnoff Gamma GOF Test	
5% K-S Critical Value	0.183	Data appear Gamma Distributed at 5% Significance Level	
Data appear Gamma Distributed at 5% Significance Level			
	Gamma Statistics		
k hat (MLE)	0.643	k star (bias corrected MLE)	0.592
Theta hat (MLE)	68.58	Theta star (bias corrected MLE)	74.42
nu hat (MLE)	32.15	nu star (bias corrected)	29.62
MLE Mean (bias corrected)	44.09	MLE Sd (bias corrected)	57.28
		Approximate Chi Square Value (0.05)	18.2
Adjusted Level of Significance	0.0395	Adjusted Chi Square Value	17.59
	Assuming Gamma Distribution		
95% Approximate Gamma UCL (use when n>=50)	71.77	95% Adjusted Gamma UCL (use when n<50)	74.27
	Suggested UCL to Use		
95% Adjusted Gamma UCL	74.27		

Output Screen for Lognormal Distribution Based UCLs (Full (w/o NDs))

X			
General Statistics			
Total Number of Observations	25	Number of Distinct Observations	25
		Number of Missing Observations	0
Minimum	0.349	Mean	44.09
Maximum	169.8	Median	18.77
SD	51.34	Std. Error of Mean	10.27
Coefficient of Variation	1.164	Skewness	1.294
Lognormal GOF Test			
Shapiro Wilk Test Statistic	0.948	**Shapiro Wilk Lognormal GOF Test**	
5% Shapiro Wilk Critical Value	0.918	Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.135	**Lilliefors Lognormal GOF Test**	
5% Lilliefors Critical Value	0.177	Data appear Lognormal at 5% Significance Level	
Data appear Lognormal at 5% Significance Level			
Lognormal Statistics			
Minimum of Logged Data	-1.053	Mean of logged Data	2.835
Maximum of Logged Data	5.135	SD of logged Data	1.68
Assuming Lognormal Distribution			
95% H-UCL	229.2	90% Chebyshev (MVUE) UCL	140.6
95% Chebyshev (MVUE) UCL	176.3	97.5% Chebyshev (MVUE) UCL	225.8
99% Chebyshev (MVUE) UCL	323		
Suggested UCL to Use			
Data appear Gamma, May want to try Gamma_Distribution			

Output Screen for Nonparametric UCLs (Full (w/o NDs))

Nonparametric Distribution Free UCLs			
95% CLT UCL	60.98	95% Jackknife UCL	61.66
95% Standard Bootstrap UCL	60.44	95% Bootstrap-t UCL	65
95% Hall's Bootstrap UCL	62.14	95% Percentile Bootstrap UCL	61.42
95% BCA Bootstrap UCL	63.63		
90% Chebyshev(Mean, Sd) UCL	74.89	95% Chebyshev(Mean, Sd) UCL	88.85
97.5% Chebyshev(Mean, Sd) UCL	108.2	99% Chebyshev(Mean, Sd) UCL	146.3
Suggested UCL to Use			
Data appear Gamma, May want to try Gamma_Distribution			

Output Screen for All Statistics Option (Full [w/o NDs])

X			
General Statistics			
Total Number of Observations	25	Number of Distinct Observations	25
		Number of Missing Observations	0
Minimum	0.349	Mean	44.09
Maximum	169.8	Median	18.77
SD	51.34	Std. Error of Mean	10.27
Coefficient of Variation	1.164	Skewness	1.294
Normal GOF Test			
Shapiro Wilk Test Statistic	0.799	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.918	Data Not Normal at 5% Significance Level	
Lilliefors Test Statistic	0.245	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.177	Data Not Normal at 5% Significance Level	
Data Not Normal at 5% Significance Level			
Assuming Normal Distribution			
95% Normal UCL		**95% UCLs (Adjusted for Skewness)**	
95% Student's-t UCL	61.66	95% Adjusted-CLT UCL (Chen-1995)	63.82
		95% Modified-t UCL (Johnson-1978)	62.1
Gamma GOF Test			
A-D Test Statistic	0.374	**Anderson-Darling Gamma GOF Test**	
5% A-D Critical Value	0.794	Detected data appear Gamma Distributed at 5% Significance Level	
K-S Test Statistic	0.113	**Kolmogrov-Smirnoff Gamma GOF Test**	
5% K-S Critical Value	0.183	Detected data appear Gamma Distributed at 5% Significance Level	
Detected data appear Gamma Distributed at 5% Significance Level			

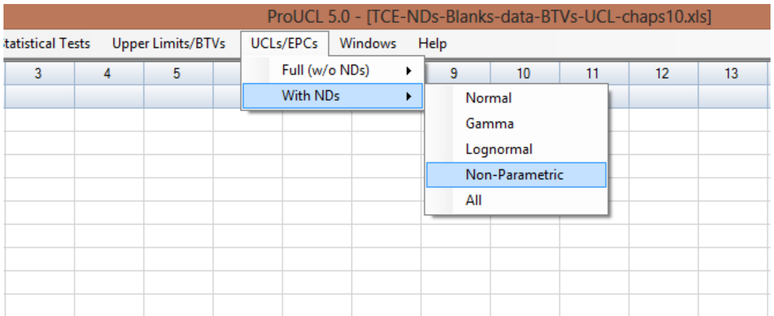
Gamma Statistics			
k hat (MLE)	0.643	k star (bias corrected MLE)	0.592
Theta hat (MLE)	68.58	Theta star (bias corrected MLE)	74.42
nu hat (MLE)	32.15	nu star (bias corrected)	29.62
MLE Mean (bias corrected)	44.09	MLE Sd (bias corrected)	57.28
		Approximate Chi Square Value (0.05)	18.2
Adjusted Level of Significance	0.0395	Adjusted Chi Square Value	17.59
Assuming Gamma Distribution			
95% Approximate Gamma UCL (use when n>=50)	71.77	95% Adjusted Gamma UCL (use when n<50)	74.27
Lognormal GOF Test			
Shapiro Wilk Test Statistic	0.948	**Shapiro Wilk Lognormal GOF Test**	
5% Shapiro Wilk Critical Value	0.918	Data appear Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.135	**Lilliefors Lognormal GOF Test**	
5% Lilliefors Critical Value	0.177	Data appear Lognormal at 5% Significance Level	
Data appear Lognormal at 5% Significance Level			
Lognormal Statistics			
Minimum of Logged Data	-1.053	Mean of logged Data	2.835
Maximum of Logged Data	5.135	SD of logged Data	1.68
Assuming Lognormal Distribution			
95% H-UCL	229.2	90% Chebyshev (MVUE) UCL	140.6
95% Chebyshev (MVUE) UCL	176.3	97.5% Chebyshev (MVUE) UCL	225.8
99% Chebyshev (MVUE) UCL	323		
Nonparametric Distribution Free UCL Statistics			
Data appear to follow a Discernible Distribution at 5% Significance Level			
Nonparametric Distribution Free UCLs			
95% CLT UCL	60.98	95% Jackknife UCL	61.66
95% Standard Bootstrap UCL	60.45	95% Bootstrap-t UCL	65.83
95% Hall's Bootstrap UCL	63.51	95% Percentile Bootstrap UCL	61.84
95% BCA Bootstrap UCL	64.96		
90% Chebyshev(Mean, Sd) UCL	74.89	95% Chebyshev(Mean, Sd) UCL	88.85
97.5% Chebyshev(Mean, Sd) UCL	108.2	99% Chebyshev(Mean, Sd) UCL	146.3
Suggested UCL to Use			
95% Adjusted Gamma UCL	74.27		

Notes: Once again, the statistics summarized above demonstrate the merits of using the gamma distribution based UCL of mean to estimate EPC terms. The use of a lognormal distribution tends to yield unrealistic UCLs of no practical merit (e.g., Lognormal UCL = 229.2 and the maximum = 169.8 in the above example). The results summarized in the above tables suggest that the use of a gamma distribution (when a data set follows a gamma distribution) cannot be dismissed just because it is easier (Helsel and Gilroy, 2012) to use a lognormal distribution to model skewed data sets.

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.

11.2 UCL for Left-Censored Data Sets with NDs

1. Click **UCLs/EPCs ► With NDs**



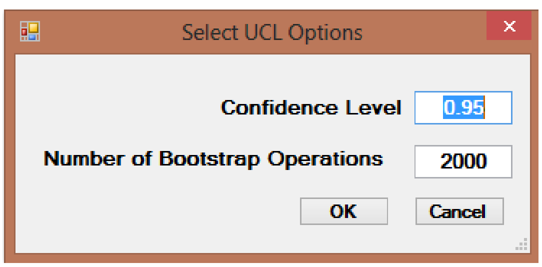
2. Choose the **Normal, Gamma, Lognormal, Non-Parametric,** or **All** option.

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

   - Select a variable(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.



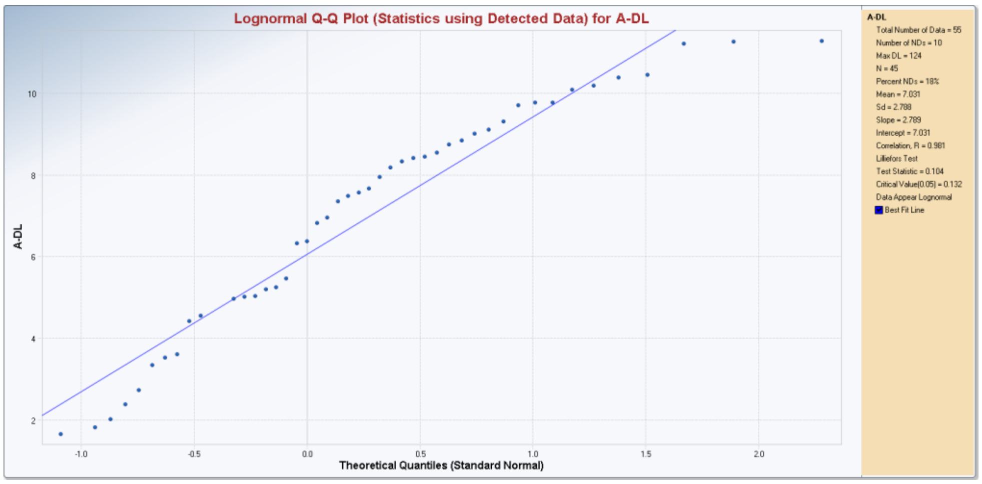
     - Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.

     - 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 11-3. 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. 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 high-lighted in the output screen provided after the lognormal GOF Q-Q plot.



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**	
5% Shapiro Wilk Critical Value	0.945	Detected Data Not Normal at 5% Significance Level	
Lilliefors Test Statistic	0.298	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.132	Detected Data Not Normal at 5% Significance Level	
Detected Data Not Normal at 5% Significance Level			
Kaplan-Meier (KM) Statistics using Normal Critical Values and other Nonparametric UCLs			
Mean	8638	Standard Error of Mean	2488
SD	18246	95% KM (BCA) UCL	13562
95% KM (t) UCL	12802	95% KM (Percentile Bootstrap) UCL	13040
95% KM (z) UCL	12731	95% KM Bootstrap t UCL	15221
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	**Kolmogrov-Smirnoff 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
MLE Mean (bias corrected)	10556	MLE Sd (bias corrected)	19216
Gamma Kaplan-Meier (KM) Statistics			
k hat (KM)	0.224	nu hat (KM)	24.66
Approximate Chi Square Value (24.66, α)	14.35	Adjusted Chi Square Value (24.66, β)	14.14
95% Gamma Approximate KM-UCL (use when n>=50)	14844	95% Gamma Adjusted KM-UCL (use when n<50)	15066
GROS Statistics using imputed NDs			
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
MLE Mean (bias corrected)	8637	MLE Sd (bias corrected)	20215
		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 (use when n>=50)	15896	95% Gamma Adjusted UCL (use when n<50)	16167
Lognormal GOF Test on Detected Observations Only			
Shapiro Wilk Test Statistic	0.939	**Shapiro Wilk GOF Test**	
5% Shapiro Wilk Critical Value	0.945	Detected Data Not Lognormal at 5% Significance Level	
Lilliefors Test Statistic	0.104	**Lilliefors GOF Test**	
5% Lilliefors Critical Value	0.132	Detected Data appear Lognormal at 5% Significance Level	
Detected Data appear Approximate Lognormal at 5% 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	13090
95% BCA Bootstrap UCL	14069	95% Bootstrap t UCL	15524
95% H-UCL (Log ROS)	1855231		
UCLs using Lognormal Distribution and KM Estimates when Detected data are Lognormally Distributed			
KM Mean (logged)	6.03	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		
Suggested UCL to Use			
95% KM (Chebyshev) UCL	19483	95% GROS Approximate Gamma UCL	15896
95% Approximate Gamma KM-UCL	14844		

Detected data follow a gamma as well as a lognormal distribution. 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.

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 (2009)

- The results summarized in the above table re-iterate 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.

Chapter 12

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) depend upon certain decision parameters including the confidence coefficient, *(1-α)* and the specified error margin (difference), Δ from the unknown true population mean, μ. Similarly, for hypotheses testing approaches, sample size determination formulae depend upon pre-specified values of the decision parameters selected while 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 not significant from both human health and cost-effectiveness point of view.

Both parametric (assuming normality) and nonparametric (distribution free) sample size determination formulae as described in guidance documents (e.g., MARSSIM 2000; EPA [2002c, 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]).

<u>New in ProUCL 5.0:</u> The Sample size module in ProUCL 5.0 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.

<u>Planning stage before collecting data:</u> 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

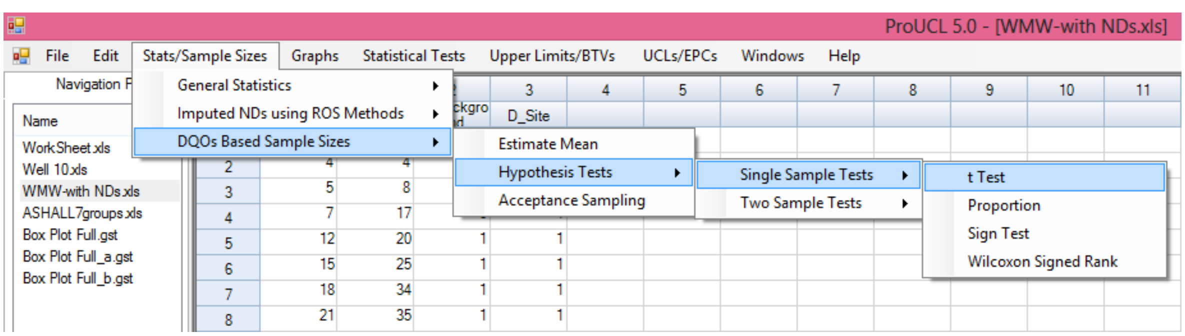
to collect the number of samples thus obtained. The detailed discussion of the sample size determination approaches during the planning stage can be found in EPA [2006a] and MARSSIM [2000].

Power assessment stage after performing a statistical method: Often, in practice, environmental samples/data sets are collected without taking the DQOs process into consideration. 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 can assess the power associated with the test in retrospect. That is for specified DQOs and decision errors (Type I error and power of the test [=1-Type II error]) and using the sample standard deviation computed based upon the already collected data, the minimum sample size needed to perform the test for specified values of the decision parameters is computed.

- 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.

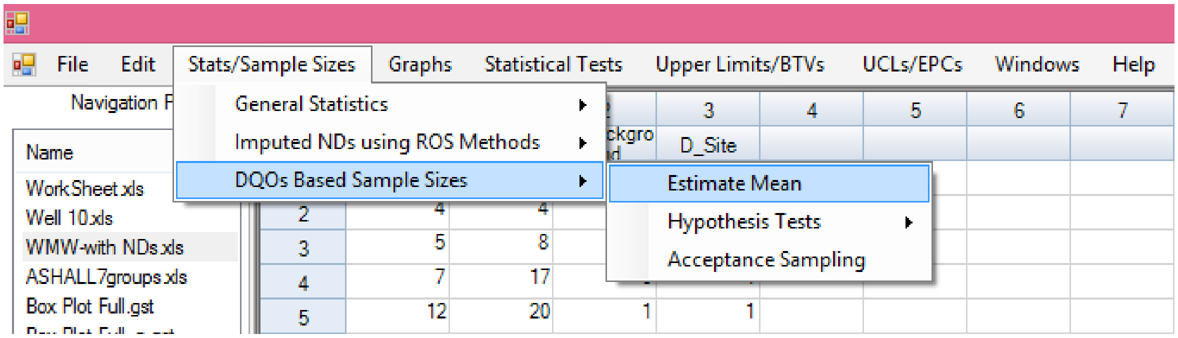
- It should be pointed out that there could be differences in the sample sizes computed in two different stages due to the differences in the values of the estimated variability. Specifically, the preliminary estimate of the variance computed using information from similar sites could be significantly 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.

Sample size determination methods in ProUCL can be used for both stages. The only difference will be in the input value of the standard deviation/variance. It is user's responsibility to input a correct value for the standard deviation during the two stages.

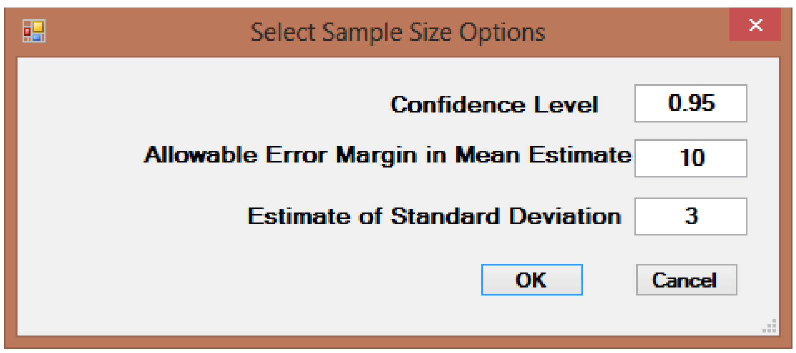


12.1 Estimation of Mean

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

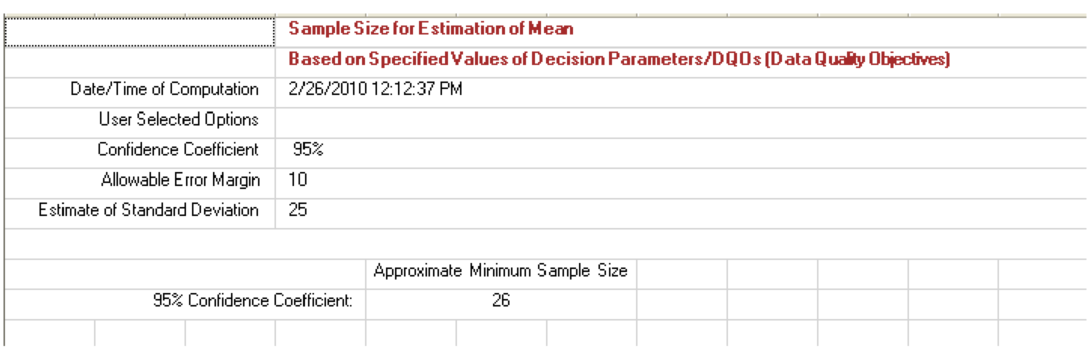


2. The following options window is shown.



- Specify the **Confidence Coefficient**. Default is **0.95**.
- Specify the **Estimate of standard deviation**. Default is **3.**
- Specify the **Allowable Error Margin in Mean Estimate**. Default is **10**.
- Click on **OK** button to continue or on **Cancel** button to cancel the options.

Output Screen for Sample sizes for Estimation of Mean (CC = 95%, *sd* = 25, Error Margin = 10)

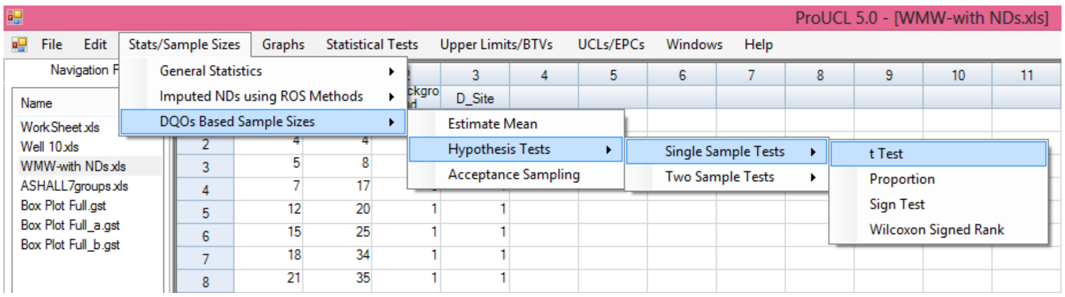


	Sample Size for Estimation of Mean	
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)	
Date/Time of Computation	2/26/2010 12:12:37 PM	
User Selected Options		
Confidence Coefficient	95%	
Allowable Error Margin	10	
Estimate of Standard Deviation	25	
		Approximate Minimum Sample Size
95% Confidence Coefficient:		26

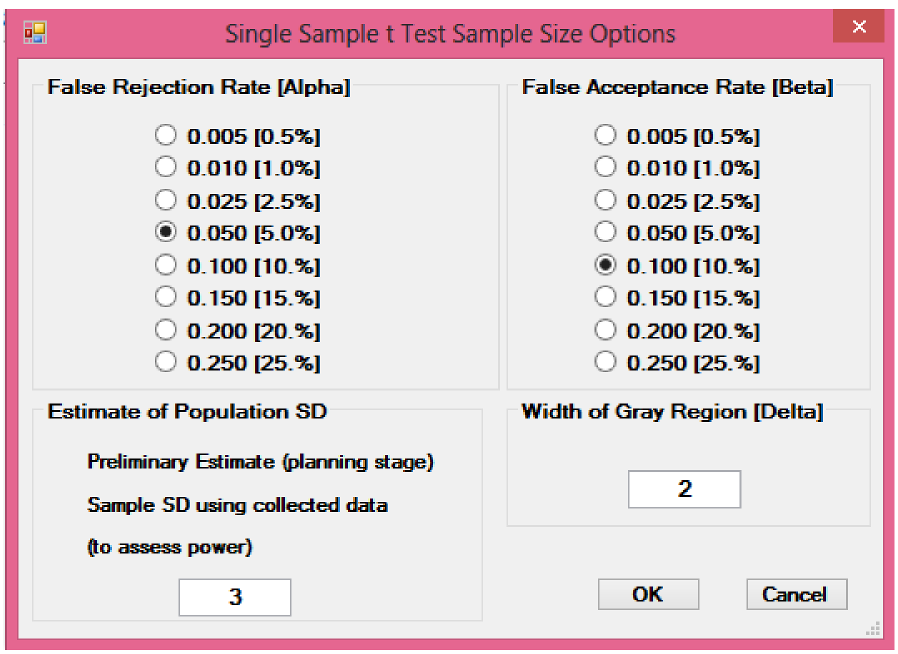
12.2 Sample Sizes for Single-Sample Hypothesis Tests

12.2.1 Sample Size for Single-Sample t-Test

1. Click **DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► t Test**



- The following options window is shown.



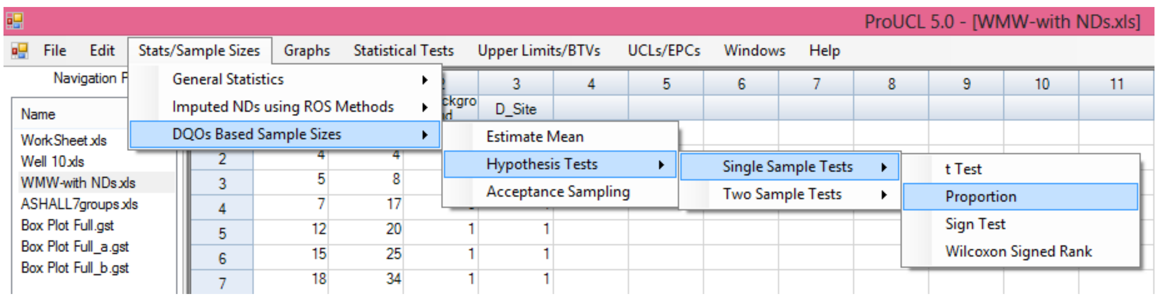
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation**. Default is **3.**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on **OK** button to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

Output Screen for Sample Sizes for Single-Sample t-Test (α = 0.05, β = 0.2, *sd* = 10.41, Δ = 10) Example from EPA 2006a (page 49)

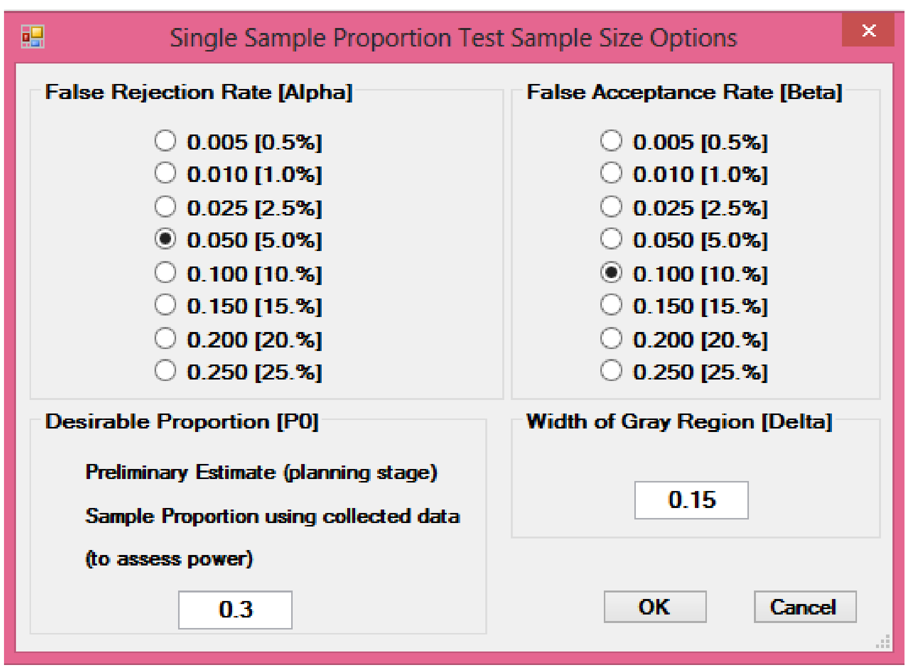
	Sample Sizes for Single Sample t Test						
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)						
Date/Time of Computation	2/26/2010 12:41:58 PM						
User Selected Options							
False Rejection Rate [Alpha]	0.05						
False Acceptance Rate [Beta]	0.2						
Width of Gray Region [Delta]	10						
Estimate of Standard Deviation	10.41						
	Approximate Minimum Sample Size						
Single Sided Alternative Hypothesis:	9						
Two Sided Alternative Hypothesis:	11						

12.2.2 Sample Size for Single-Sample Proportion Test

1. Click **DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► Proportion**



2. The following options window is shown.



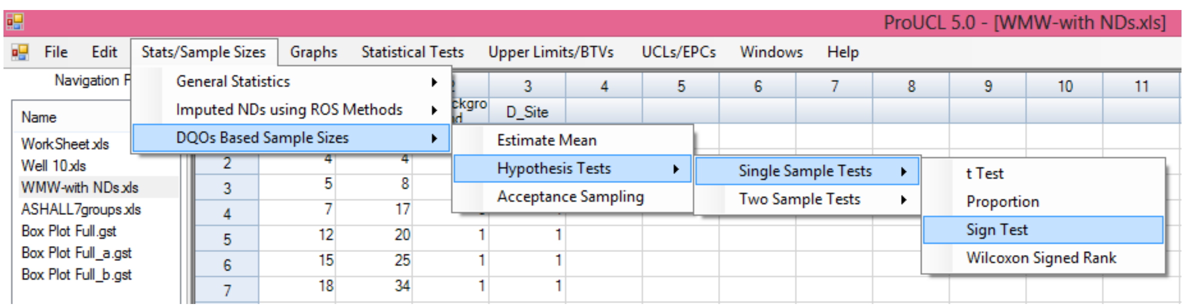
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Desirable Proportion (P0)**. Default is **0.3.**
- Specify the **Width of the Gray Region (Delta)**. Default is **0.15**.
- Click on **OK** button to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

Output Screen for Sample 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)

12.2.3 Sample Size for Single-Sample Sign Test

1. Click **DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► Sign Test**



2. The following options window is shown.

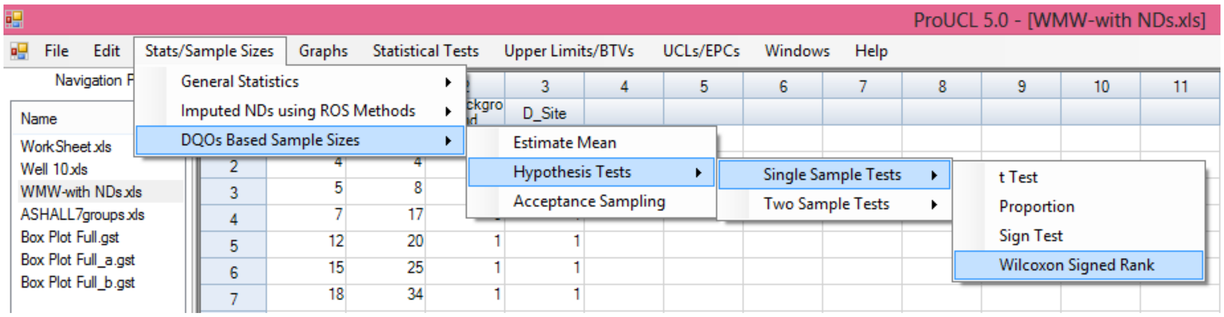
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on **OK** button to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

Output Screen for Sample Sizes for Single-Sample Sign Test (Default Options)

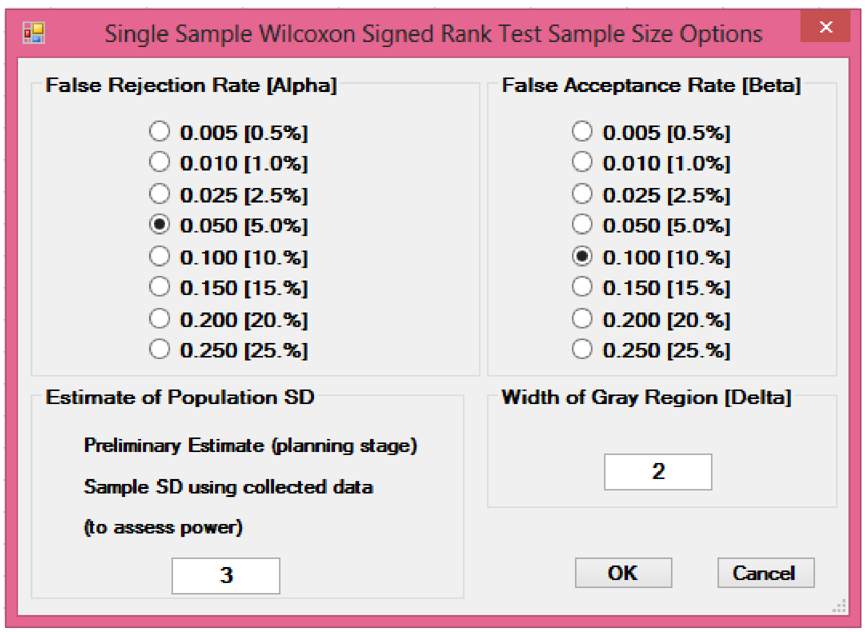
	Sample Sizes for Single Sample Sign Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:15:27 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.1
Width of Gray Region [Delta]	2
Estimate of Standard Deviation	3
	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	35
Two Sided Alternative Hypothesis:	43

12.2.4 Sample Size for Single-Sample Wilcoxon Signed Rank Test

1. Click **DQOs Based Sample Sizes ► Hypothesis Tests► Single Sample Tests► Wilcoxon Signed Rank**



2. The following options window is shown.



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

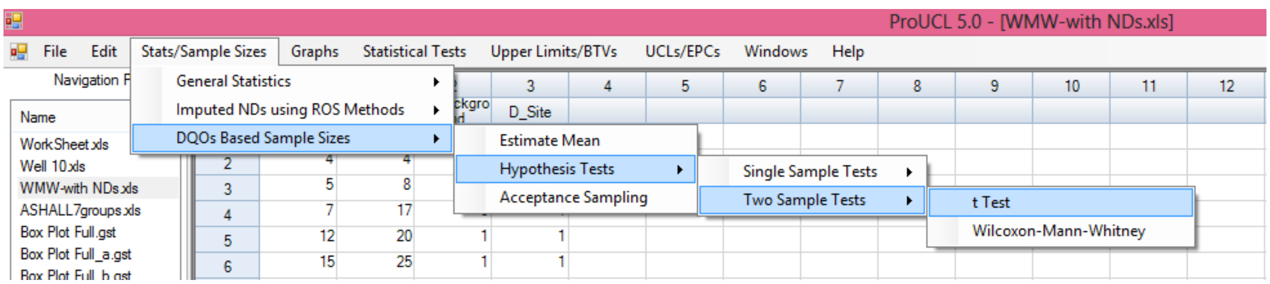
Output Screen for Sample Sizes for Single-Sample WSR Test (α = 0.1, β = 0.2, *sd* = 130, Δ = 100) Example from EPA 2006a (page 65)

	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					

12.3 Sample Sizes for Two-Sample Hypothesis Tests

12.3.1 Sample Size for Two-Sample t-Test

1. Click **DQOs Based Sample Sizes ► Hypothesis Tests► Two Sample Tests► t Test**



2. The following options window is shown.

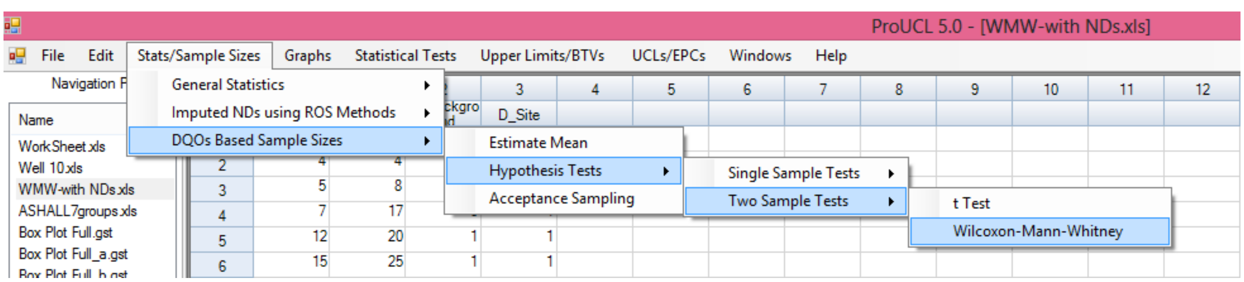
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click on **OK** button to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

Output Screen for Sample Sizes for Two-Sample t-Test (α = 0.05, β = 0.2, s_p = 1.467, Δ = 2.5) Example from EPA 2006a (page 68)

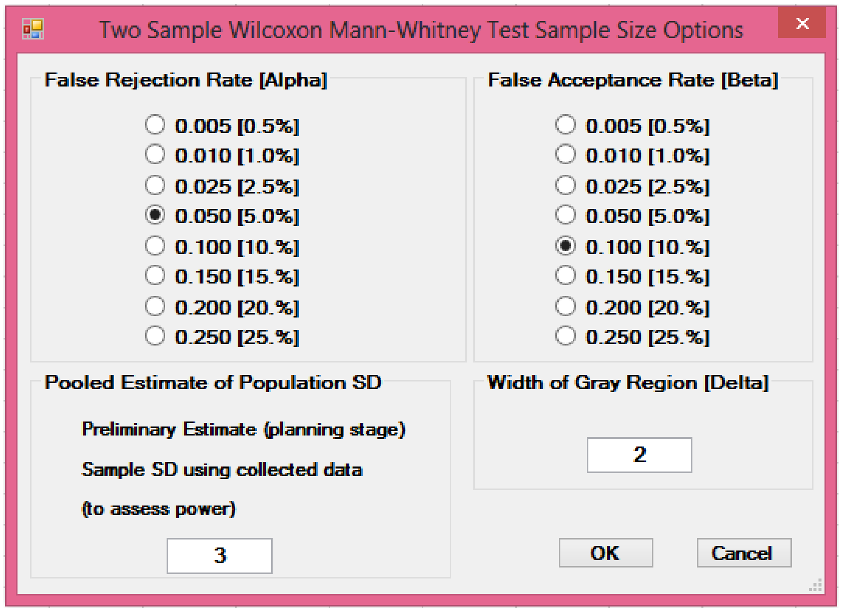
	Sample Sizes for Two Sample t Test					
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)					
Date/Time of Computation	2/26/2010 1:17:57 PM					
User Selected Options						
False Rejection Rate [Alpha]	0.05					
False Acceptance Rate [Beta]	0.2					
Width of Gray Region [Delta]	2.5					
Estimate of Pooled SD	1.467					
	Approximate Minimum Sample Size					
Single Sided Alternative Hypothesis:	5					
Two Sided Alternative Hypothesis:	7					

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

1. Click **DQOs Based Sample Sizes ► Hypothesis Tests► Two Sample Tests► Wilcoxon-Mann-Whitney**



2. The following options window is shown.



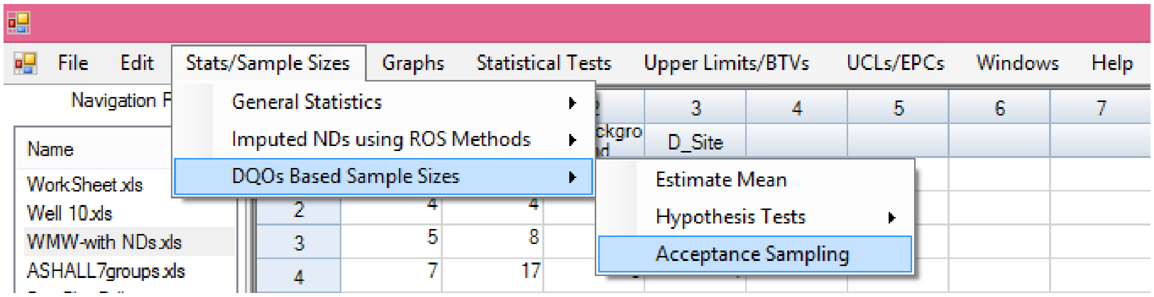
- Specify the **False Rejection Rate (Alpha)**. Default is **0.05**.
- Specify the **False Acceptance Rate (Beta)**. Default is **0.1**.
- Specify the **Estimate of standard deviation of WMW Test Statistic**. Default is **3**
- Specify the **Width of the Gray Region (Delta)**. Default is **2**.
- Click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

Output Screen for Sample Sizes for Single-Sample WMW Test (Default Options)

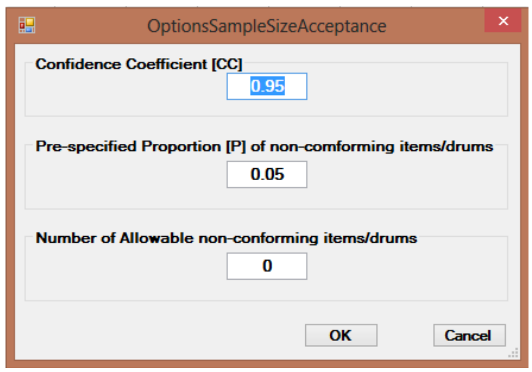
	Sample Sizes for Two Sample Wilcoxon-Mann-Whitney Test
	Based on Specified Values of Decision Parameters/DQOs (Data Quality Objectives)
Date/Time of Computation	2/26/2010 12:18:47 PM
User Selected Options	
False Rejection Rate [Alpha]	0.05
False Acceptance Rate [Beta]	0.1
Width of Gray Region [Delta]	2
Estimate of Standard Deviation	3
	Approximate Minimum Sample Size
Single Sided Alternative Hypothesis:	46
Two Sided Alternative Hypothesis:	56

12.4 Sample Sizes for Acceptance Sampling

1. Click **DQOs Based Sample Sizes ► Acceptance Sampling**



2. The following options window is shown.



- 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.

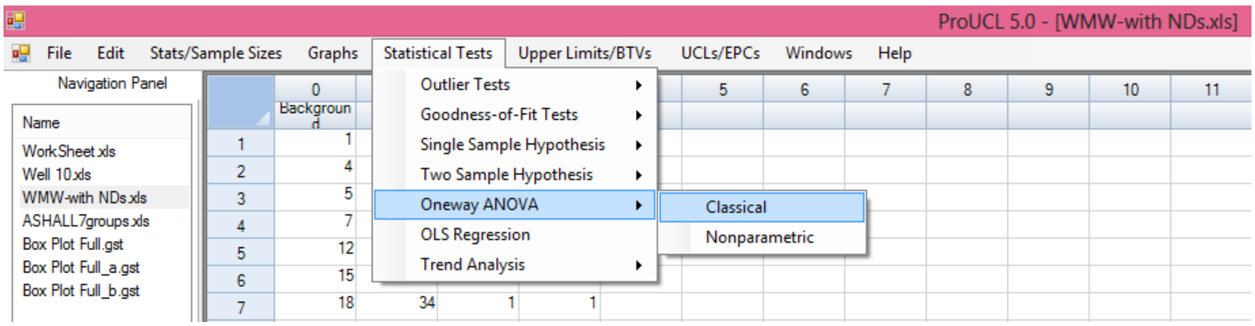
Output Screen for Sample Sizes for Acceptance Sampling (Default Options)

	Acceptance Sampling for Pre-specified Proportion of Non-conforming Items
	Based on Specified Values of Decision Parameters/DQOs
Date/Time of Computation	2/26/2010 12:20:34 PM
User Selected Options	
Confidence Coefficient	0.95
Pre-specified proportion of non-conforming items in the lot	0.05
Number of allowable non-conforming items in the lot	0
	Approximate Minimum Sample Size
Exact Binomial/Beta Distribution	59
Approximate Chisquare Distribution (Tukey-Scheffe)	59

Chapter 13

Analysis of Variance

Oneway 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. Oneway ANOVA is often used to perform inter-well comparisons in groundwater monitoring projects. Classical Oneway ANOVA is a generalization of the two-sample t-test (Hogg and Craig, 1995); and nonparametric ANOVA, Kruskal-Wallis test (Hollander and Wolfe, 1999) is a generalization of the two- sample Wilcoxon Mann Whitney test. Theoretical details of Oneway ANOVA are given in the ProUCL Technical Guide. Oneway ANOVA is available under the Statistical Tests module of ProUCL 5.0. It is advised to use these tests on raw data in the original scale without transforming the data (e.g., using a log-transformation).



13.1 Classical Oneway ANOVA

1. Click **Oneway ANOVA ► Classical**

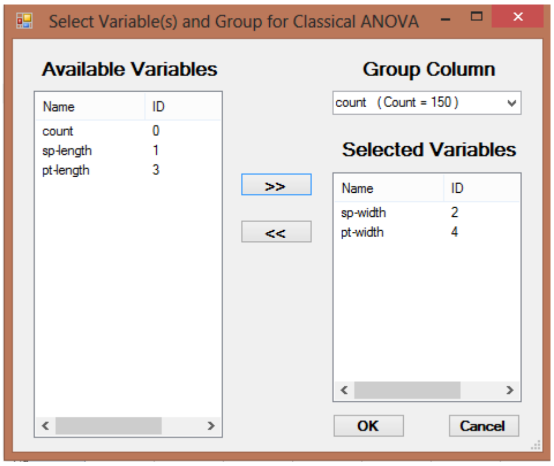
The data file used should follow the format as shown below; the data file should consist of a group variable defining the various groups (stacked data) to be evaluated using the **Oneway ANOVA** module. The **Oneway ANOVA** module can process multiple variables simultaneously.

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

2. 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 13-1a. 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. Oneway ANOVA results with conclusions for the variable sepal-width (sp-width) are shown as follows:



Output for a Classical Oneway 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			

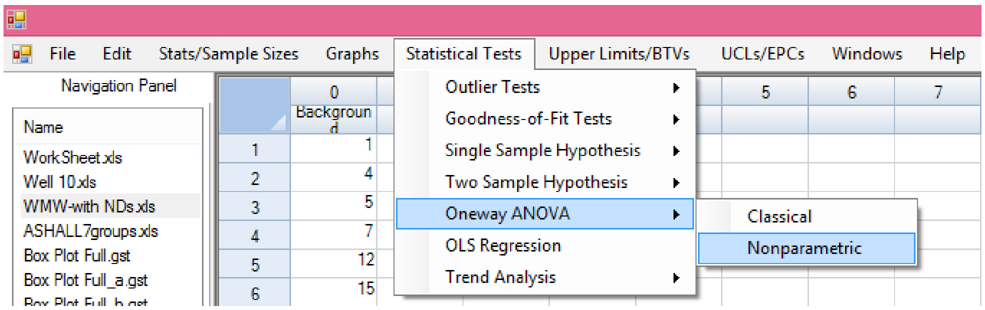
Pooled Standard Deviation	0.34
R-Sq	0.401

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.

13.2 Nonparametric ANOVA

Nonparametric Oneway 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.

1. Click Oneway **ANOVA ► Nonparametric**



Like classical Oneway 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 **Oneway ANOVA** module.

2. 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 13-1b (continued). Nonparametric Oneway ANOVA results with conclusion for sepal-length (sp-length) are shown as follows.

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.

Chapter 14

Ordinary Least Squares of Regression and Trend Analysis

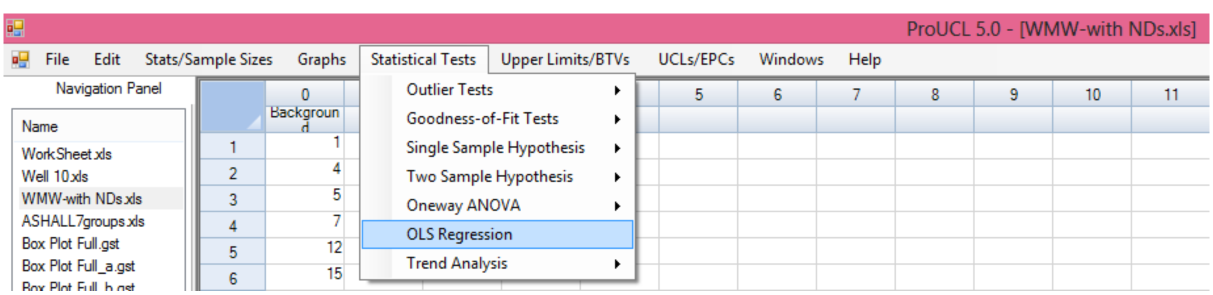
The OLS of regression and trend tests are often used to determine trends potentially present in constituent concentrations at polluted sites, especially in GW monitoring applications. The OLS regression and two nonparametric trend tests: Mann-Kendall test and Theil-Sen test are available under the Statistical Tests module of ProUCL 5.0. 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 5.0.

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,...) 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 to use the Theil-Sen test 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 5.0 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.

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).

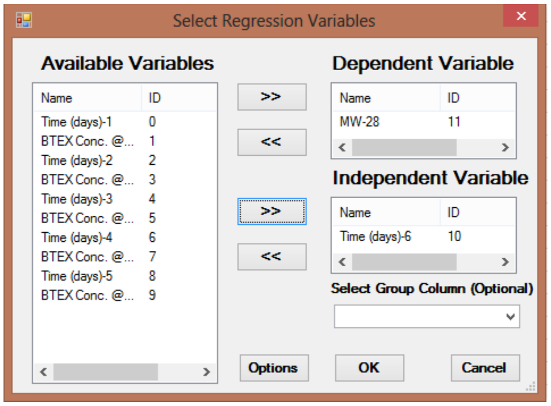
14.1 Simple Linear Regression

1. Click **Statistical Tests► OLS Regression.**



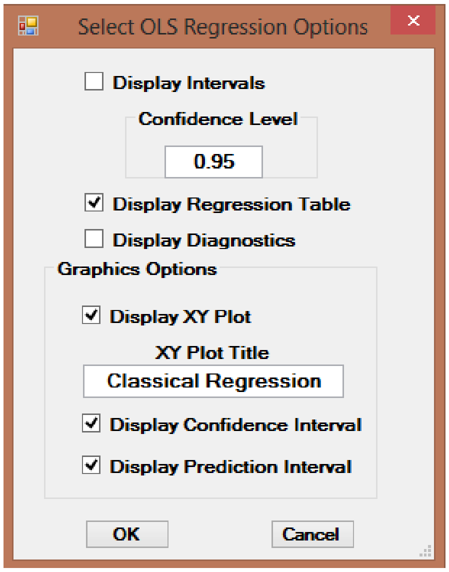
2. The **Select Regression Variables** screen will appear.

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



- 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.



  - 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.

- o Select **Confidence Interval** and **Prediction Interval** to display the confidence and the prediction bands around the regression line.

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

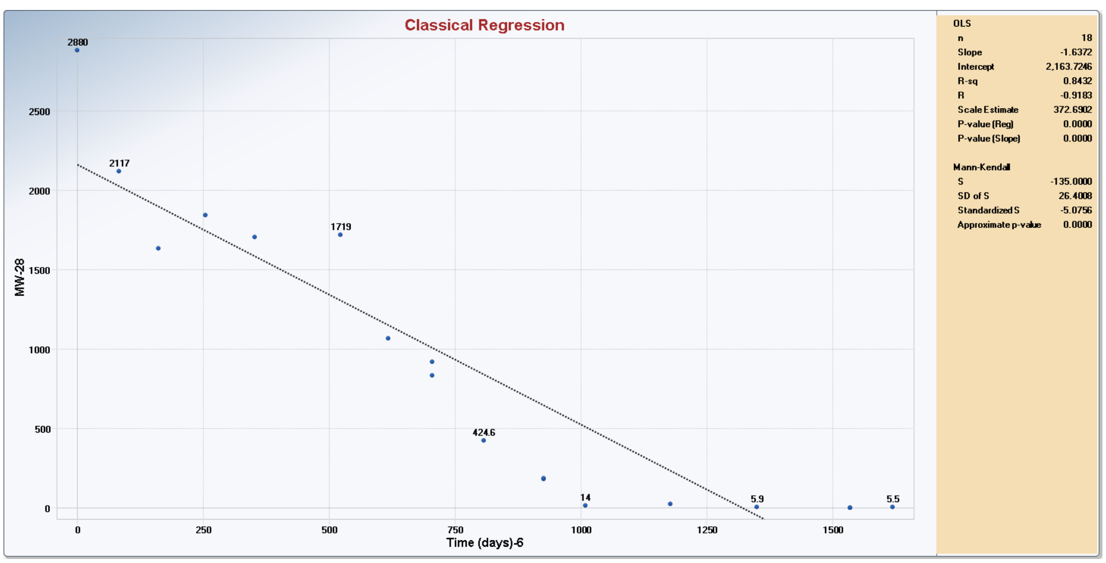
- Click **OK** to continue or **Cancel** to cancel the OLS Regression.

  - o 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.

  - o The above options will also generate an Excel-Type output sheet. A partial output sheet is shown below following the OLS Regression Graph.

Example 14-1a. Consider analyte concentrations, X collected from a groundwater (GW) monitoring well, MW-28 over a certain period of time. 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.

OLS Regression Graph without Regression and Prediction Intervals



OLS Regression Graph with Regression and Prediction Intervals

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
Time (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

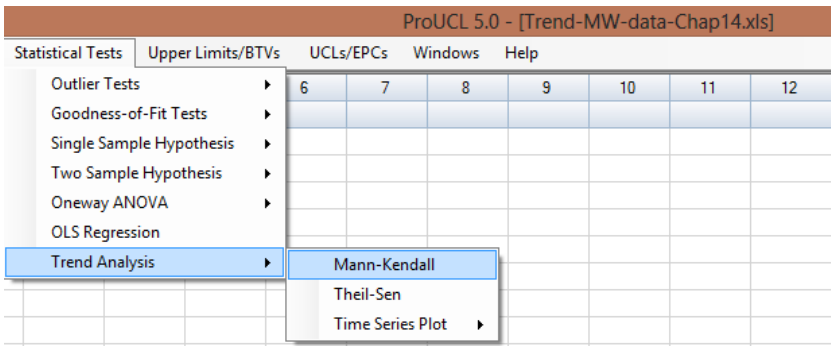
Regression Table

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

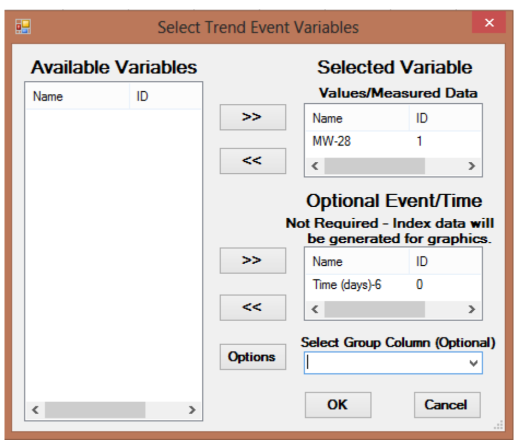
Verifying Normality of Residuals: As shown in the above partial output, ProUCL displays residuals including standardized residuals on the OLS output sheet. Those residuals can be imported (copying and pasting) in an excel file to assess the normality of those OLS residuals. The parametric trend evaluations 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 to determine potential trends present in a time series data set.

14.2 Mann-Kendall Test

1. Click **Statistical Tests ►Trend Analysis ► Mann-Kendall.**



2. The **Select Trend Event Variables** screen will appear.

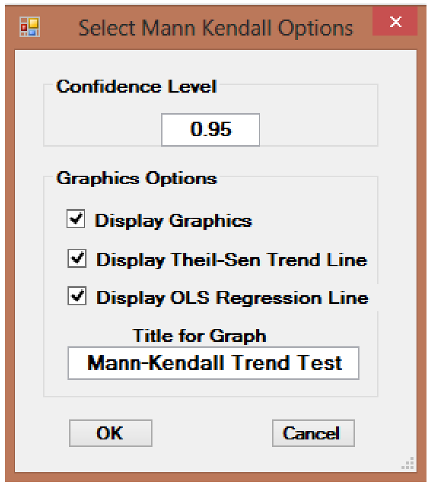


- Select the **Event/Time** variable. This variable is optional to perform the Mann-Kendall (M-K) Test; however, for graphical display it is suggested to provide a valid Event/Time variable (numerical values only). If the user wants to generate a graphical display without providing an Event/Time variable, ProUCL generates an index variable to represent sampling events.

- 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.



  - Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.

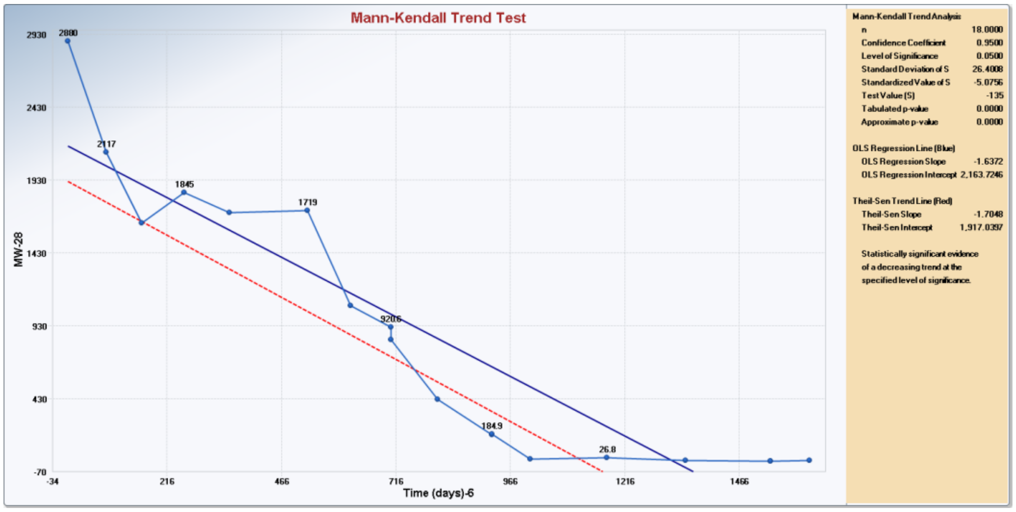
  - Select the trend lines to be displayed: **OLS Regression Line** and/or **Theil-Sen Trend Line**. If only **Display Graphics** is chosen, a time series plot will be generated.

  - 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.

14- 1b (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.

Mann Kendall Test Trend Graph Displaying all Selected Options



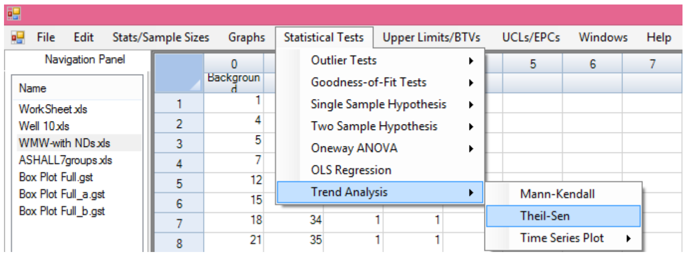
Mann-Kendall Trend Test Output Sheet

	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.	

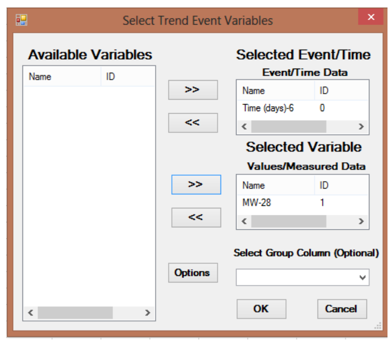
14.3 Theil – Sen Test

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.

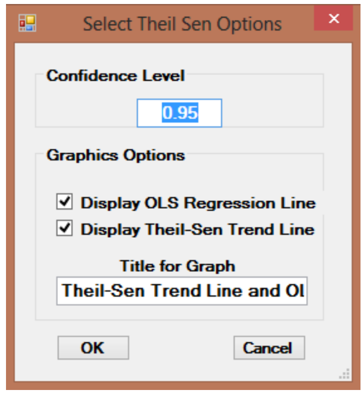
1. Click **Statistical Tests ►Trend Analysis ► Theil-Sen.**



2. The **Select Variables** screen will appear.



- 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.

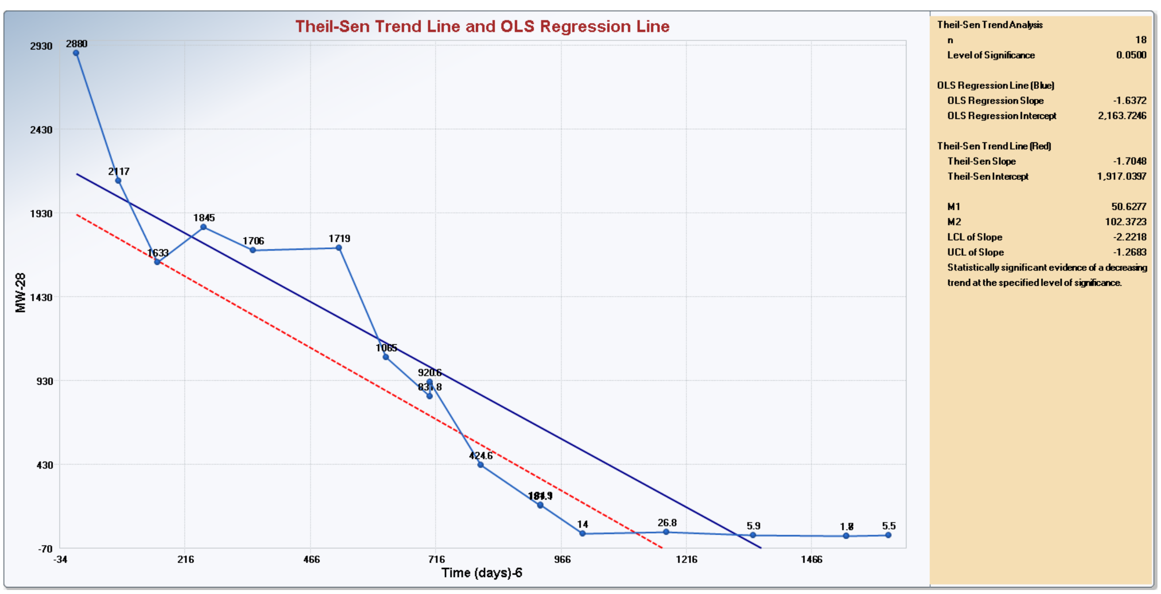


- o Specify the **Confidence Level**; a number in the interval (0.5, 1), 0.5 inclusive. The default choice is **0.95**.
- o Select the trend lines to be displayed: **OLS Regression Line** and/or **Theil-Sen Trend Line.**
- o Click on **OK** button to continue or on **Cancel** button to cancel the option.

- Click **OK** to continue or **Cancel** to cancel the Theil-Sen Test.

14- 1c (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 Trend Graph displaying all Selected Options



Theil-Sen Trend Test Output Sheet

Date/Time of Computation	3/27/2013 2:19:55 PM		
From File	Trend-MW-data-Chap14.xls	**Approximate inference for Theil-Sen Trend Test**	
Full Precision	OFF	Mann-Kendall Statistic (S)	-137
Confidence Coefficient	0.95	Standard Deviation of S	26.4
Level of Significance	0.05	Standardized Value of S	-5.151
		Approximate p-value	1.2930E-7
MW-28		Number of Slopes	153
		Theil-Sen Slope	-1.705
General Statistics		Theil-Sen Intercept	1917
Number of Events	18	M2'	98.21
Number Values Reported (n)	18	One-sided 95% upper limit of Slope	-1.365
Minimum	1.7	95% LCL of Slope (0.025)	-2.222
Maximum	2880	95% UCL of Slope (0.975)	-1.268
Mean	864.6		
Geometric Mean	174.8	Statistically significant evidence of a decreasing	
Median	628.2	trend at the specified level of significance.	
Standard Deviation	913.1		

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.

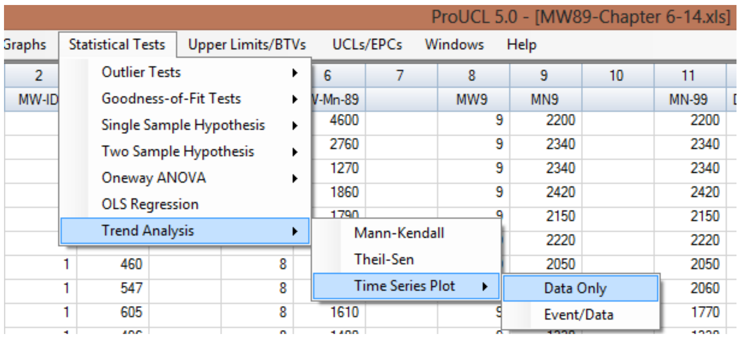
Averaging of Multiple Measurements at Sampling Events: In practice, when multiple observations are collected/reported at one or more sampling events (times), one or more pairwise slopes may become infinite resulting in a failure to compute the Theil-Sen test statistic. In such cases, the user may want 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 5.0 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.

14.4 Time Series Plots

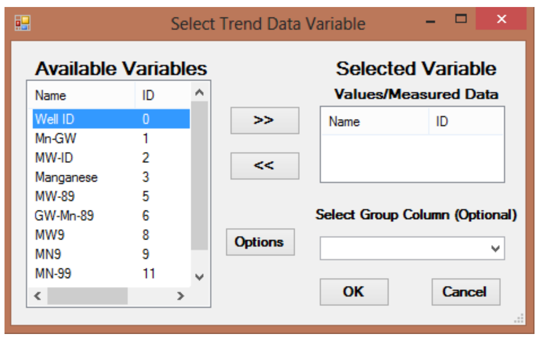
This option of the Trend Analysis module can be used to determine and compare trends in multiple groups over the same period of time.

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 be 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 year, ...). 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.

1. Click **Statistical Tests ►Trend Analysis ► Time Series Plots**

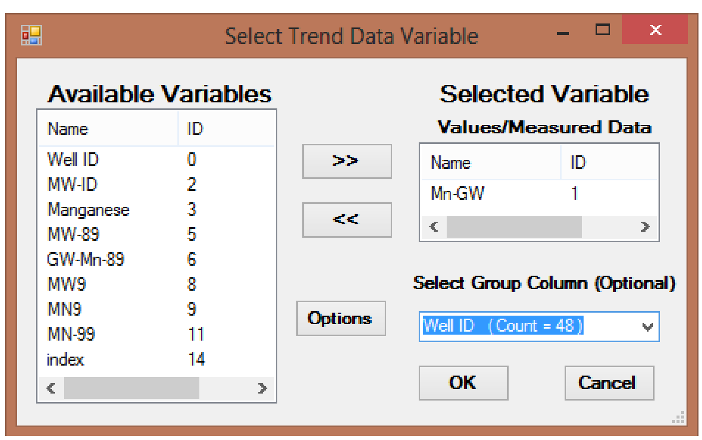


2. When the **Data Only** option is clicked, the following window is shown:

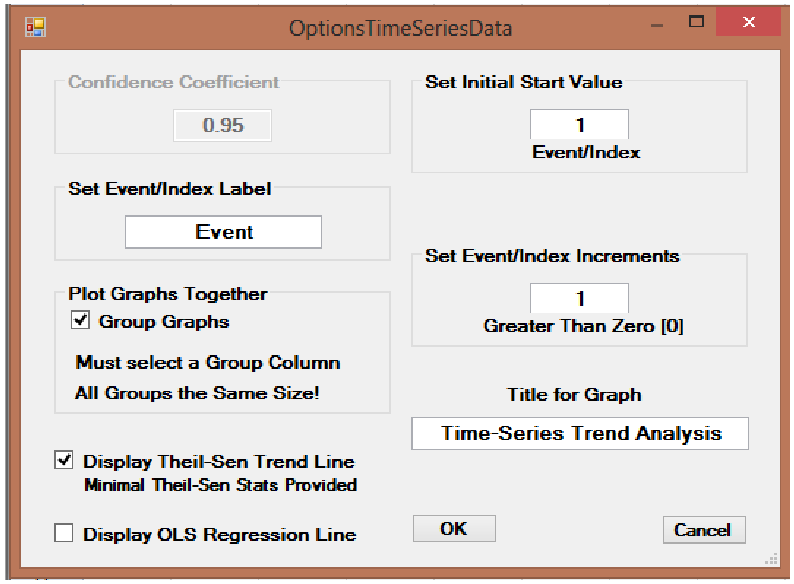


This option is used on the measured data only. The user selects a variable with measured values which are used in generating a time series plot. The time series plot option is specifically useful when data come from multiple groups (monitoring wells during the same period of time).

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



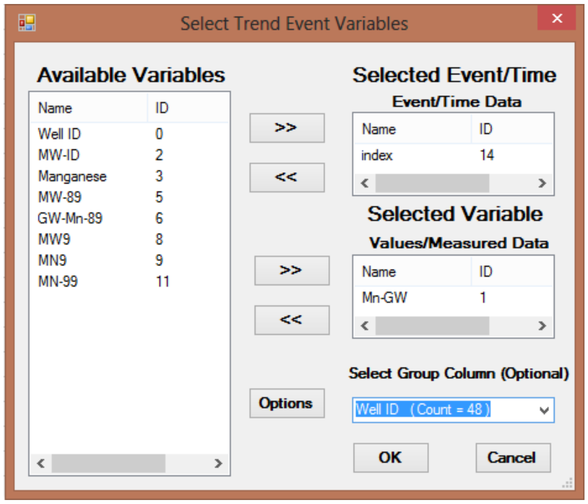
- When the **Options** button is clicked, the following window will be shown.



The user can select to display graphs individually or together for all groups 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.

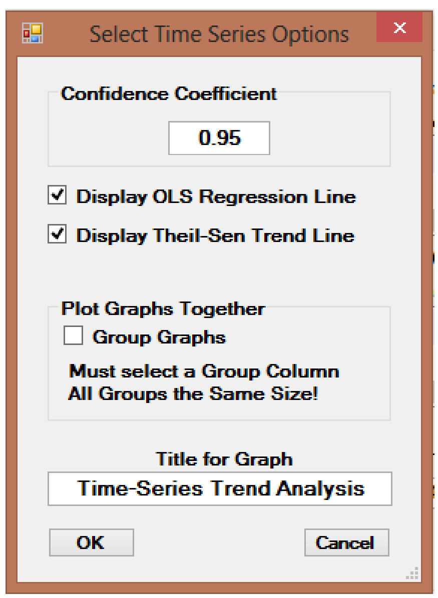
3. When the **Event/Data** option is clicked, the following window is shown:



- Select a group variable (is 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.

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



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.

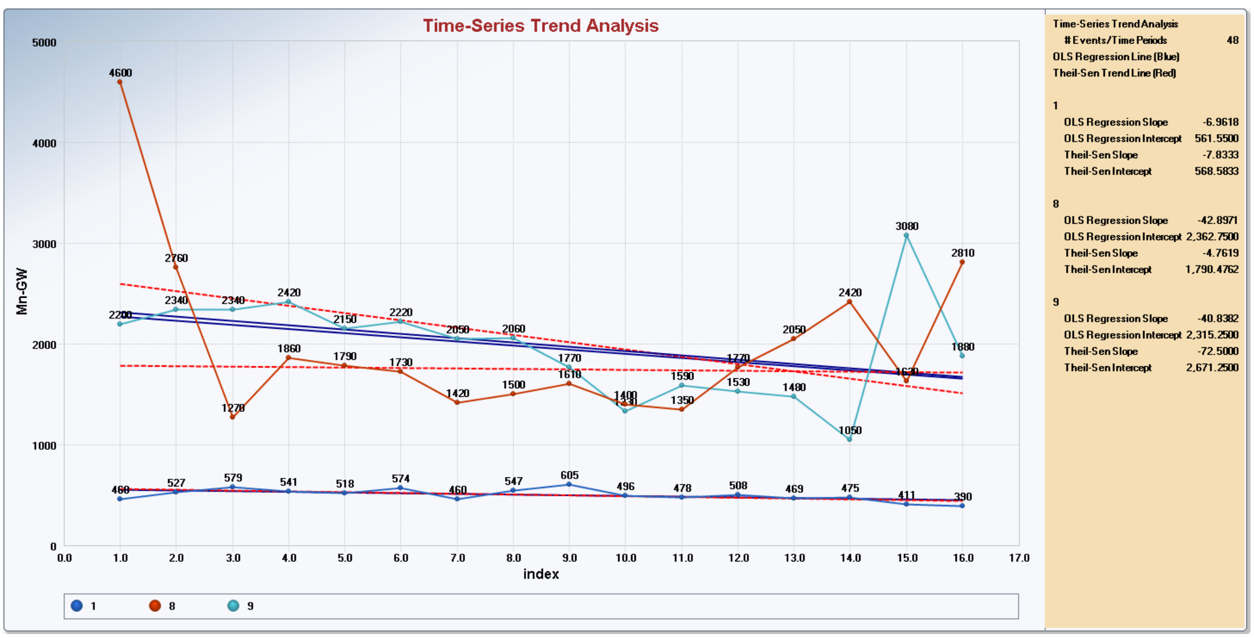
- Specify the lines (**Regression** and/or **Theil-Sen**) to be displayed on the time series plot.
- Select **Plot Graphs Together** option for comparing time series trends for more than one group on the same graph.
- If this option is not selected but a **Group Variable** is selected, different graphs will be plotted for each group.
- Click on **OK** button to continue or on **Cance**l button to cancel the options.

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

Notes: 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 14-2. 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). Some trend statistics are displayed in the side panel.

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



Chapter 15

Background Incremental Sample Simulator (BISS) Simulating BISS Data from a Large Discrete Background Data

The Background Incremental Sample Simulator (BISS) module has been incorporated in ProUCL5.0 at the request of the Office of Superfund Remediation and Technology Innovation (OSRTI). However, this module is currently under further investigation and research, and therefore it is not available for general public use. This module may be released in a future version of the ProUCL software, along with strict conditions and guidance for how it is applied. The main text for this chapter is not included in this document for release to general public. Only a brief placeholder write-up is provided here.

The following scenario describes the Site or project conditions under which the BISS module could be useful: Suppose there is a long history of soil sample collection at a Site. In addition to having a large amount of Site data, a robust background data set (at least 30 samples from verified background locations) has also been collected. Comparison of background data to on-Site data has been, and will continue to be, an important part of this project's decision-making strategy. All historical data is from discrete samples, including the background data. There is now a desire to switch to incremental sampling for the Site. However, guidance for incremental sampling makes it clear that it is inappropriate to compare discrete sample results to incremental sample results. That includes comparing a Site's incremental results directly to discrete background results.

One option is to recollect all background data in the form of incremental samples from background decision units (DUs) that are designed to match Site DUs in geology, area, depth, target soil particle size, number of increments, increment sample support, etc. If project decision-making uses a background threshold value (BTV) strategy to compare Site DU results one at a time against background, then an appropriate number (the default is no less than 10) of background DU incremental samples would need to be collected to determine the BTV for the population of background DUs. However, if the existing discrete background data show background concentrations to be low (in comparison to Site concentrations) and fairly consistent (relative standard deviation, RSD <1), there is a second option described as follows.

When a robust discrete background data set that meets the above conditions already exists, the following is an alternative to automatically recollecting ALL background data as incremental samples.

Step 1. Identify 3 background DUs and collect at least 1 incremental sample from each for a minimum of 3 background incremental samples.

Step 2. Enter the discrete background data set ($n \geq 30$) and the ≥ 3 background incremental samples into the BISS module (the BISS module will not run unless both data sets are entered).

- The BISS module will generate a specified (default is 7) simulated incremental samples from the discrete data set.

- The module will then run a t-test to compare the simulated background incremental data set (e.g., with $n = 7$) to the actual background incremental data set ($n \geq 3$).

- If the t-test finds no difference between the 2 data sets, the BISS module will combine the 2 data sets and determine the statistical distribution, mean, standard deviation, potential UCLs and potential BTVs for the combined data set. Only this information will be supplied to the general user. The individual values of the simulated incremental samples will not be provided.

- If the t-test finds a difference between the actual and simulated data sets, the BISS module will not combine the data sets nor provide a BTV.

- In both cases, the BISS module will report summary statistics for the actual and simulated data sets.

Step 3. If the BISS module reported out statistical analyses from the combined data set, select the BTV to use with Site DU incremental sample results. Document the procedure used to generate the BTV in project reports. If the BISS module reported that the simulated and actual data sets were different, the historical discrete data set cannot be used to simulate incremental results. Additional background DU incremental samples will need to be collected to obtain a background DU incremental data set with the number of results appropriate for the intended use of the background data set.

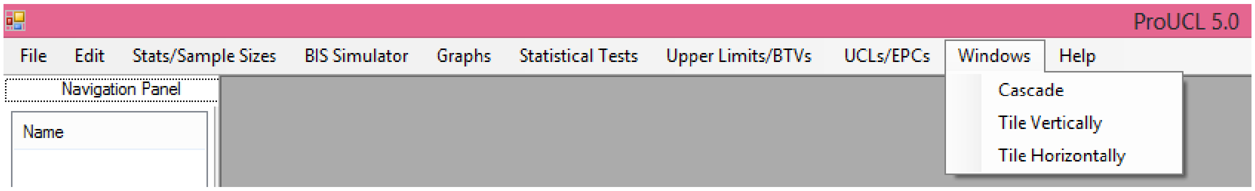
The objective of the BISS module is to take advantage of the information provided by the existing background discrete samples. The availability of a large discrete data set collected from the background areas with geological formations and conditions comparable to the Site DU(s) of interest is a requirement for successful application of this module. There are fundamental differences between incremental and discrete samples. For example, the sample supports of discrete and incremental samples are very different. Sample support has a profound effect on sample results so samples with different sample supports should not be compared directly, or compared with great caution.

Since incremental sampling is a relatively new approach, the performance of the BISS module requires further investigation. If you would like to try this strategy for your project, or if you have questions, contact Deana Crumbling, crumbling.deana@epa.gov.

Chapter 16

Windows

The Windows Menu performs typical Windows program options.



Click on the **Window** menu to reveal the drop-down options shown above.

The following Window drop-down menu options are available:

- Cascade option: arranges windows in a cascade format.**{**tc "7. Window " \l 2**}** This is similar to a typical Windows program option.

- Tile option: resizes each window vertically or horizontally and then displays all open windows. This is similar to a typical Windows program option.

- The drop-down options list also includes a list of all open windows with a check mark in front of the active window. Click on any of the windows listed to make that window active. This is especially useful if you have many windows (e.g., >40) open; the navigation panel only holds the first 40 windows.

Chapter 17

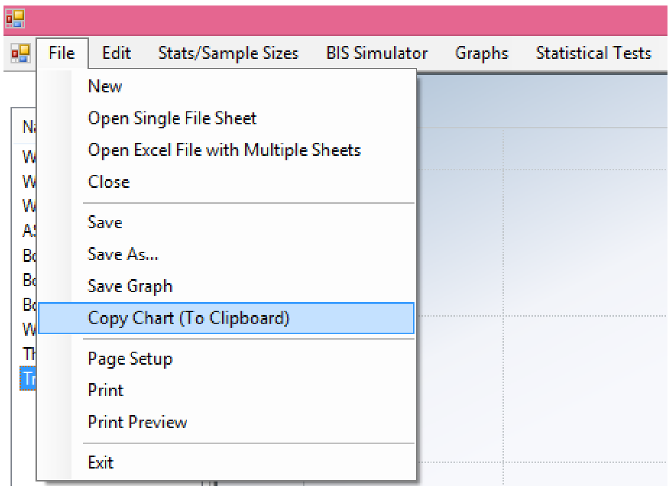
Handling the Output Screens and Graphs

17.1 Copying and Saving Graphs

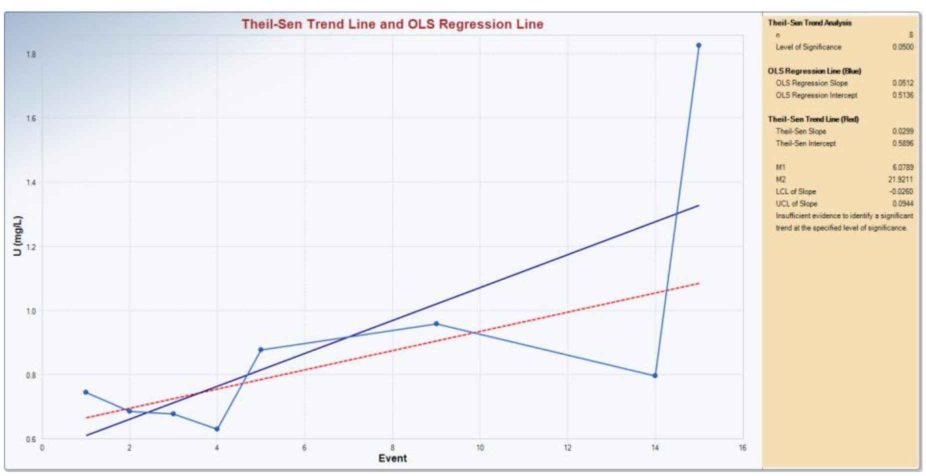
Graphs can be copied into Word, Excel, or PowerPoint files in two ways.

1. Click the **Copy Chart (To Clipboard)** shown below; a graph must be present to be copied to the clipboard.

File ► Copy Chart (To Clipboard)

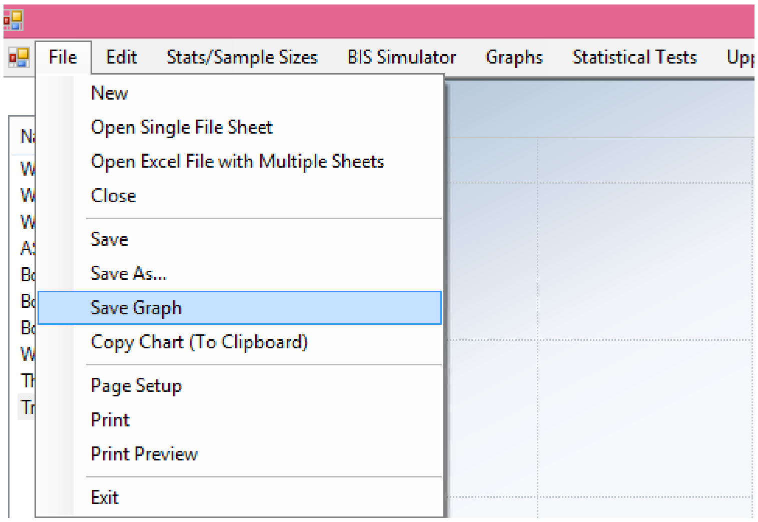


Once the user has clicked **Copy Chart (To Clipboard)**, the graph is ready to be imported (pasted) into most Microsoft office applications (e.g., Word, Excel, and PowerPoint) by clicking the **Edit ► Paste** option in those Microsoft applications as shown below.



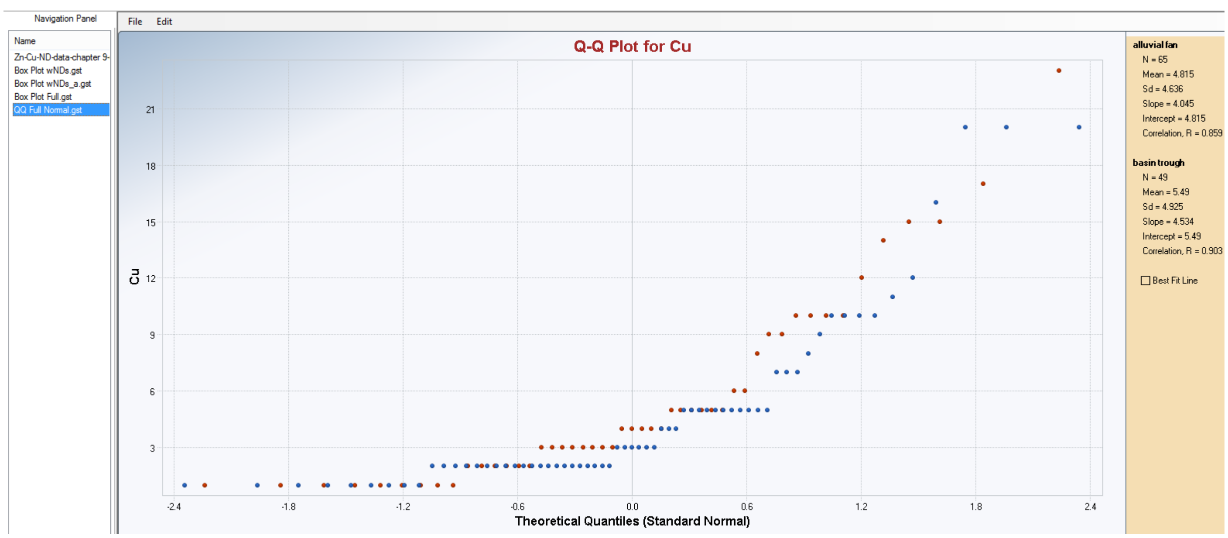
2. Graphs can be saved using the **Save Graph** Option in the **Navigation Panel** as a **Bitmap file** with .bmp extension. The user can import the saved bitmap file into a desired document such as a word document or a PowerPoint presentation by using the **Copy** and **Paste** options available in the selected Microsoft application.

File ► Save Graph

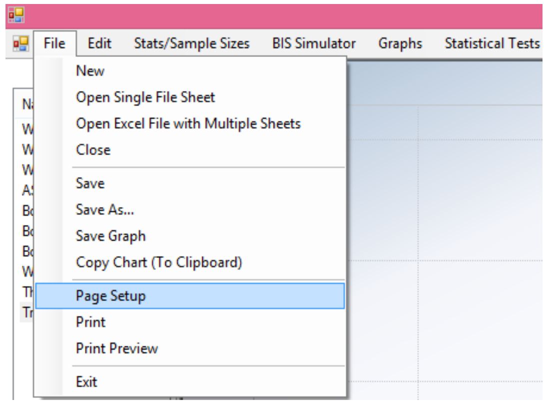


17.2 Printing Graphs

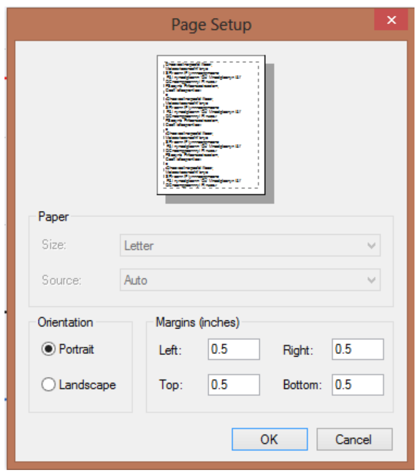
1. Click the graph you want to print in the **Navigation Panel**.



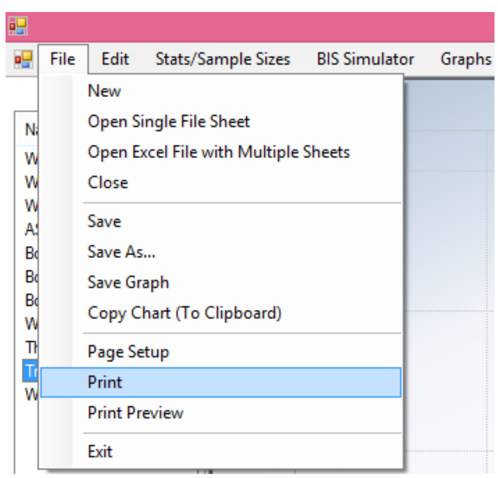
2. Click **File ► Page Setup**.



3. Check the button next to **Portrait** or **Landscape** (shown below), and click **OK**. In some cases, with larger headings and captions, it may be desirable to use the **Landscape** printing option.

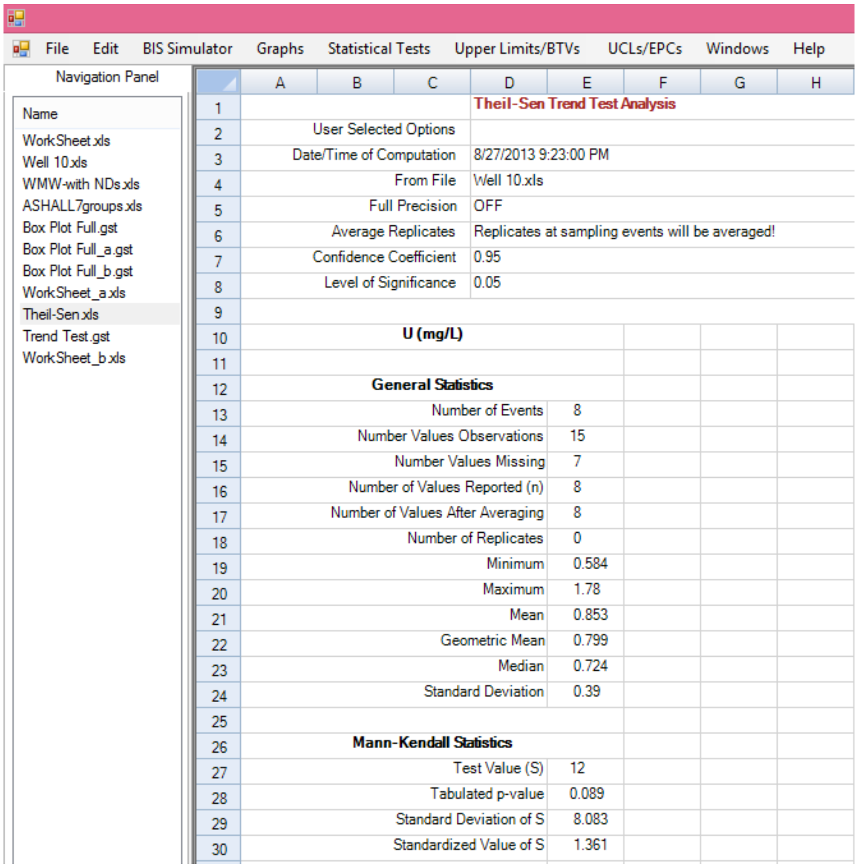


4. Click **File ► Print** to print the graph, and **File ► Print Preview** to preview (optional) the graph before printing.

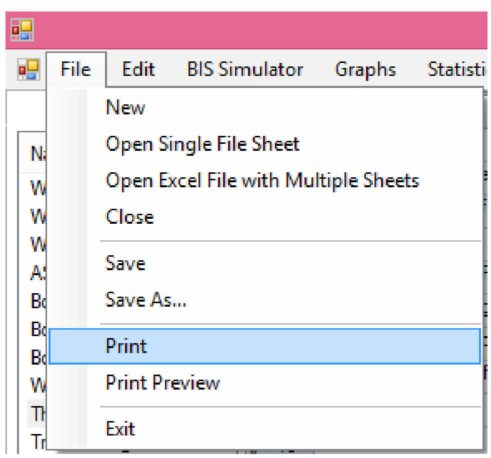


17.3 Printing Non-graphical Outputs

1. Click/Highlight the output you want to save or print in the **Navigation Panel**.



2. Click **File ► Print** or **File ► Print Preview** if you wish to see the preview before printing.

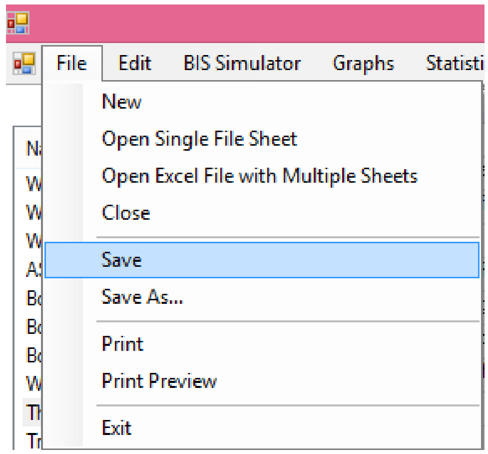


17.4 Saving Output Screens as Excel Files

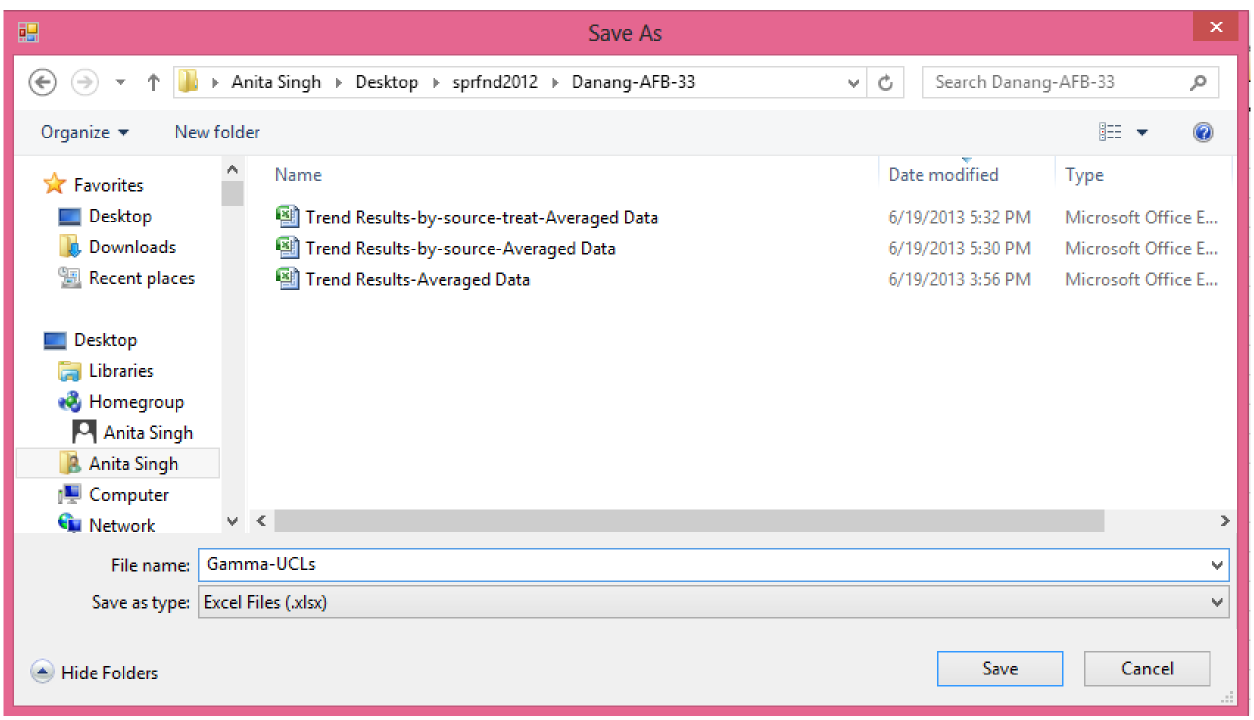
ProUCL 5.0 saves output files and data files as Excel files with .xls or .xlsx extensions.

1. Click on the output you want to save in the **Navigation Panel List**.

2. Click **File ► Save** or **File► Save As**



3. Enter the desired file name you want to use, and click **Save**, and save the file in the desired folder using your browser as shown below.



Chapter 18

Summary and Recommendations to Compute a 95% UCL for Full Uncensored and Left-Censored Data Sets with NDs

This chapter briefly summarizes recommendations and the process to compute upper confidence limits of the population mean based upon data sets with and without ND observations. The recommendations are made based upon the simulation studies summarized in Singh, Singh, and Engelhardt (1997, 1999); Singh, Singh, and Iaci (2002); Singh and Singh (2003); and Singh, Maichle, and Lee (2006). Some details can be found in Chapters 2 and 4 of the associated ProUCL 5.0 Technical Guide. Depending upon the data size, data distribution (e.g., normal, gamma, lognormal, and nonparametric), and data skewness, ProUCL suggests using one or more 95% UCL to estimate the population mean. If needed, the user may want to consult a statistician for additional insight.

18.1 Computing UCL95s of the Mean Based Upon Uncensored Full Data Sets

- Formal GOF tests and GOF Q-Q plots are used first to determine the data distribution so that appropriate parametric or nonparametric UCL95s can be computed.

- For a normally or approximately normally distributed data set, the user is advised to use Student's t-distribution-based UCL of the mean. Student's t UCL or modified-t-statistic based UCL can be used to estimate the EPC term when the data set is symmetric (e.g., skewness $= \left|\hat{k}_3\right|$ is smaller than 0.2-0.3) or mildly skewed; that is, when σ or $\hat{\sigma}$ is less than 0.5.

- For gamma or approximately gamma distributed data sets, the user is advised to: 1) use the approximate gamma UCL when k>1 and n≥50; 2) use the adjusted gamma UCL when k>1 and n<50; 3) use the bootstrap-t method or Hall's bootstrap method when khat ≤ 1 and the sample size, n < 15-20; and 4) use the adjusted gamma UCL (if available) for khat ≤ 1 and sample size, n ≥ 15. If the adjusted gamma UCL is not available, then use the approximate gamma UCL as an estimate of the EPC term. When bootstrap-t method or Hall's bootstrap method yields an erratic inflated UCL (e.g., when outliers are present) result, the UCL may be computed using the adjusted gamma UCL (if available) or the approximate gamma UCL.

- For lognormally distributed data sets, ProUCL recommends a UCL computation method based upon the sample size, n, and standard deviation of the log-transformed data, $\hat{\sigma}$. These suggestions are summarized in Table 2-10 of the ProUCL 5.0 Technical Guide.

- For nonparametric data sets, which are not normally, lognormally, or gamma distributed, a nonparametric UCL is used to estimate the EPC term. Methods used to estimate EPC terms based upon nonparametric data sets are summarized in Table 2-11 of the ProUCL 5.0 Technical Guide. For example for mildly skewed nonparametric data sets of smaller sizes (e.g., <30), one may use a modified-t UCL or BCA bootstrap UCL; and for larger samples one may use a CLT-UCL, adjusted-CLT UCL, or a BCA bootstrap UCL. These nonparametric UCLs computation methods do not provide desired coverage to the mean for moderately skewed to highly skewed data sets.

- For moderately skewed to highly skewed nonparametric data sets, the use of a Chebyshev (Mean, *Sd*) UCL is suggested. It is noted that for extremely skewed data sets (e.g., with $\hat{\sigma}$ exceeding 3.0), even a Chebyshev inequality-based 99% UCL of the mean fails to provide the desired coverage (e.g., 0.95) of the population mean.

- For highly skewed data sets with $\hat{\sigma}$ exceeding 3.0, 3.5, it is suggested to pre-process the data. It is very likely that the data consist of outliers and/or come from multiple populations. The population partitioning methods may be used to identify mixture populations present in the data set. For defensible conclusions, the decision statistics such as EPC terms may be computed separately for each of the identified sub-population present in the mixture data set.

18.2 Computing UCLs Based Upon Left-Censored Data Sets with Nondetects

The parametric maximum likelihood estimation (MLE) methods assume that there is only one detection limit; therefore parametric MLE methods (e.g., Cohen's MLE method) are not available in ProUCL 5.0. Since it is not easy to verify (perform goodness-of-fit) the distribution of a left-censored data set consisting of detects and NDs with multiple detection limits, some poor performing estimation methods including the parametric MLE methods and the winsorization method are not retained in ProUCL 5.0. In ProUCL 5.0, emphasis is given on the use of nonparametric UCL computation methods and hybrid parametric methods based upon KM estimates which account for data skewness in the computation of UCL95. It is recommended to avoid the use of transformations (to achieve symmetry) while computing the upper limits based upon left-censored data sets. It is not easy to correctly interpret the statistics computed in the transformed scale. Moreover, the results and statistics computed in the original scale do not suffer from transformation bias. Like full uncensored data sets, when the standard deviation of the log-transformed data becomes >1.0, avoid the use of a lognormal model even when the data appear to be lognormally distributed. Its use often results in unrealistic statistics of no practical merit (Singh, Singh, and Engelhard 1997; Singh, Singh, and Iaci, 2002). It is also recommended to identify potential outliers representing observations coming from population(s) different from the main dominant population and investigate them separately. Decisions about the disposition of outliers should be made by all interested members of the project team.

- It is recommended to avoid the use of the DL/2 (t) UCL method, as the DL/2 UCL does not provide the desired coverage (for any distribution and sample size) for the population mean, even for censoring levels as low as 10%, 15%. This is contrary to the conjecture and assertion (e.g., EPA 2006a) made that the DL/2 method can be used for lower (e.g., ≤ 20%) censoring levels. The coverage provided by the DL/2 (t) method deteriorates fast as the censoring intensity increases. The DL/2 (t) method is not recommended by the authors or developers of this text and ProUCL 5.0.

- The use of the KM estimation method is a preferred method as it can handle multiple detection limits. Therefore, the use of KM estimates is suggested to compute the decision statistics based upon methods which adjust for data skewness. Depending upon the data set size, distribution of the detected data, and data skewness, the various nonparametric and hybrid KM UCL95 methods including KM (BCA), bootstrap-t KM UCL, Chebyshev KM UCL, Gamma-KM UCL based upon the KM estimates provide good coverages for the population mean. All of these methods are available in ProUCL 5.0.

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.

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 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.

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.

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).

p^{th} Percentile or p^{th} 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 then 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: This is similar to a 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: A sample here 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 characteristic under study (e.g., lead concentrations). It can also be measured in terms of the standard deviation of log-transformed data. The greater is 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 merits, 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 is 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.

Uncensored Data Set: A data set without any censored observations is called an uncensored data set.

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 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

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 M. Engelhardt. 1991. *Statistical Analysis of Reliability and Life Testing Models*, Theory and Methods. 2^{nd} 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 (Navy). 1998. Procedural Guidance for Statistically Analyzing Environmental Background Data, Naval Facilities Engineering Command.

Department of Navy. 1999. *Handbook for Statistical Analysis of Environmental Background Data*. Prepared by SWDIV and EFA West of Naval Facilities Engineering Command. July 1999.

Department of Navy. 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.

Department of Navy (Navy). 2003. Guidance for Environmental Background Analysis, Naval Facilities Engineering Command, Volume II: Sediment, UG-2054-ENV.

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

Draper NR & 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.

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.

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

Gillian, Mimmack, Manas, G. and Meyer, D. 2001. *Introductory Statistics for Business.* Pearsons South Africa, p. 21.

Gilliespie, B.W., Chen, Q., *et al*. 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. 2012. *Statistics for Censored Environmental Data Using Minitab and R.* Second Edition. John Wiley and Sons, NY.

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

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

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

Hinton, S.W. 1993. *Δ 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.

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.

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

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 Ingram Olkin, Harold Hotelling, et alia. Stanford University Press. pp. 278–292.

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.*

Natrella, M.G. 1963. *Experimental Statistics*. National Bureau of Standards, Hand Book No. 91, U.S. Government Printing Office, Washington, DC.

Kunter, M. J., C. J. Nachtsheim, J. Neter, and Li W. 2004. *Applied Linear Statistical Methods.* Fifth Edition. McGraw-Hill/Irwin.

Palachek, A.D., D.R. Weier, T.R. Gatliffe, D.M. Splett, and D.K. Sullivan. 1993. Statistical Methodology for Determining Contaminants of Concern by Comparison of Background and Site Data with Applications to Operable Unit 2, SA-93-010, Internal Report, Statistical Applications, EG&G Rocky Flats Inc., Rocky Flats Plant, Golden, CO.

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. 1982. *An extension of Shapiro and Wilk's W test for normality to large samples*. Applied Statistics, 31, 115–124.

Royston, P. 1982a. *Algorithm AS 181: The W test for Normality*. Applied Statistics, 31, 176–180

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.

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. 1978. *Kolmogorov-Smirnov Test Statistic for the Gamma Distribution with Unknown Parameters*, Dissertation, Department of Statistics, Temple University, Philadelphia, PA.

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, 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.

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. 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., Singh, A.K., and G. Flatman 1994. *Estimation of Background Levels of Contaminants*. Math Geology, Vol. 26, No, 3, 361-388.

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.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, Anita, 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. and Nocerino, J.M. 2002. *Robust Estimation of the Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations,* Vol. 60, pp 69-86.

Singh, A., Singh, A.K., and Iaci, R.J. 2002. *Estimation of the Exposure Point Concentration Term Using a Gamma Distribution*, EPA/600/R-02/084, October 2002.

Singh, A. and Singh, A.K. 2003. *Estimation of the Exposure Point Concentration Term (95% UCL) Using Bias-Corrected Accelerated (BCA) Bootstrap Method and Several other methods for Normal, Lognormal, and Gamma Distributions*. Draft EPA Internal Report.

Singh, A. 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 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

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.

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.

U.S. Environmental Protection Agency (EPA). 1991. *A Guide: Methods for Evaluating the Attainment of Cleanup Standards for Soils and Solid Media*. Publication EPA/540/R95/128.

U.S. Environmental Protection Agency (EPA). 1992a. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. Publication EPA 9285.7-081, May 1992.

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.

U.S. Environmental Protection Agency (EPA). 1994. *Statistical Methods for Evaluating the Attainment of Cleanup Standards*, EPA 230-R-94-004, Washington, DC.

U.S. Environmental Protection Agency (EPA). 1996. *A Guide: Soil Screening Guidance: Technical Background Document*. Second Edition, Publication 9355.4-04FS.

U.S. Environmental Protection Agency (EPA). 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.

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.

U.S. Environmental Protection Agency (EPA). 2002c. *RCRA Waste Sampling, Draft Technical Guidance – Planning, Implementation and Assessment*. EPA 530-D-02-002, 2002.

U.S. Environmental Protection Agency (EPA). 2004. *ProUCL Version 3.1, Statistical Software*. National Exposure Research Lab, EPA, Las Vegas Nevada, October 2004. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

U.S. Environmental Protection Agency (EPA). 2006a, *Guidance on Systematic Planning Using the Data Quality Objective Process*, EPA QA/G-4, EPA/240/B-06/001. Office of Environmental Information, Washington, DC. Download from: http://www.epa.gov/quality/qs-docs/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: http://www.epa.gov/quality/qs-docs/g9s-final.pdf

U.S. Environmental Protection Agency (EPA). 2007. *ProUCL Version 4.0 Technical Guide*. EPA 600-R-07-041, 2007.

U.S. Environmental Protection Agency (EPA). 2009. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance*. EPA 530-R-09-007, 2009.

U.S. Environmental Protection Agency (EPA). 2009a. *ProUCL Version 4.00.04 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. Down load from: http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

U.S. Environmental Protection Agency (EPA). 2009b. *ProUCL4.00.04 Facts Sheet*. *Statistical Software for Environmental Applications for Data Sets with and without nondetect observations.* National Exposure Research Lab, EPA, Las Vegas, Nevada, 2009. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

U.S. Environmental Protection Agency (EPA). 2009c. *Scout 2008 – A Robust Statistical Package*, Office of Research and Development, February 2009. http://www.epa.gov/esd/databases/scout/abstract.htm#Scout2008v101 http://www.epa.gov/nerlesd1/databases/datahome.htm

U.S. Environmental Protection Agency (EPA). 2010. *Scout 2008 User Guide (Draft)* EPA/600/R-08/038, Office of Research and Development, April 2010. http://www.epa.gov/nerlesd1/databases/datahome.htm

U.S. Environmental Protection Agency (EPA). 2010a. *A Quick Guide to the Procedures in Scout (Draft)*, Office of Research and Development, April 2010. http://www.epa.gov/nerlesd1/databases/datahome.htm http://www.epa.gov/esd/databases/scout/abstract.htm

U.S. Environmental Protection Agency (EPA). 2010b. *ProUCL Version 4.00.05 User Guide (Draft)*. EPA/600/R-07/041, May 2010. http://www.epa.gov/osp/hstl/tsc/software.htm

U.S. Environmental Protection Agency (EPA). 2010c. ProUCL Version 4.00.05 Technical Guide (Draft). EPA/600/R-07/041, May, 2010. http://www.epa.gov/osp/hstl/tsc/software.htm

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. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

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. http://www.epa.gov/osp/hstl/tsc/softwaredocs.htm

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.

United States
Environmental Protection
Agency

Office of Research
and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

EPA/600/R-07/038
Februaryl 2009
www.epa.gov

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE ☐ ; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35