

## About ProUCL 5.1 Software

The development of ProUCL 5.1 software has been funded by USEPA to compute rigorous statistics to help the decision makers in making defensible decisions at polluted sites which are cost-effective and protective of human health and the environment. For data sets with and without nondetect (ND) observations, ProUCL computes statistics using parametric and nonparametric methods covering a wide-range of data skewness, data distributions, and sample sizes. ProUCL 5.1 software is compatible with Microsoft Office 8.0 and represents an upgrade of all previous versions of ProUCL. ProUCL 5.1 has been developed in Microsoft .NET Framework 4.0 using the C# programming language. To provide Excel-compatible Spreadsheet functionality, ProUCL uses FarPoint Spread 5 for .NET; and for graphics, ProUCL uses the development software package, ChartFx 7. Menu options in ProUCL 5.1 are similar to those of ProUCL 5.0. Most statistical tests: Outlier tests, Goodness-of-fit tests, Single and Two-Sample hypothesis tests, Oneway ANOVA, OLS regression, and Trend Analysis are available under the Statistical Tests module. Several enhancements have been made in UCLs/EPCs and Upper Limits/BTVs modules and a new statistic, an upper simultaneous limit (USL) has been incorporated in the Upper limits/BTVs module. For data sets with NDs, ProUCL 5.1 computes upper confidence limits (UCLs), upper prediction and tolerance limits (UPLs, UTLs), and USLs which adjust for data skewness. For data sets with NDs, ProUCL 5.1 also computes upper limits using KM estimates in gamma (normal, lognormal or nonparametric) UCL, UPL, and UTL equations provided the detected observations follow a gamma (normal, lognormal or nonparametric) distribution. ProUCL 5.1 outputs the confidence coefficient actually achieved by a nonparametric UTL for specified values of confidence coefficient, coverage probability, and sample size. The OLS and Trend Analysis modules handle missing events to compute trend test statistics and generate trend line graphs. All bugs and typographical errors found by the developers and users of ProUCL have been addressed in ProUCL 5.1.

The graphical methods in ProUCL include histograms, multiple quantile-quantile (Q-Q) plots, and side-by-side box plots. ProUCL has a couple of outlier tests: Dixon and Rosner tests which can be used on data sets with and without NDs. ProUCL has Goodness-of-Fit (GOF) tests for normal, lognormal, and gamma distributions which can be used for data sets with and without NDs. The GOF tests available in ProUCL include: Shapiro-Wilk and Lilliefors tests, Anderson-Darling and Kolmogorov-Smirnov tests. Single-sample hypothesis tests include: t-test, Sign, Wilcoxon Signed Rank, and proportion tests; and two-sample tests include: t-test, Wilcoxon Rank Sum (WRS) test, Gehan, and Tarone-Ware tests. ProUCL can compute parametric (normal, gamma, lognormal) and nonparametric (bootstrap, Chebyshev inequality, and Central Limit Theorem) decision statistics including UCLs of mean, upper percentiles, UPLs for future  $k$  ( $\geq 1$ ) observations, UTLs, and USLs. Oneway ANOVA module performs classical parametric and nonparametric Kruskal-Wallis tests. The Regression module performs the ordinary least squares (OLS) of regression on any bivariate data set including time-series data sets. The OLS module produces a regression table and generates a scatter plot displaying the OLS line with confidence and prediction intervals (bands) around the regression line. The Trend Analysis module performs Mann-Kendall and Theil-Sen tests on time-series data sets with missing values and generates trend graphs displaying parametric OLS regression and nonparametric Theil-Sen trend lines. The Time Series Plot option can be used to graphically and formally compare multiple time-series data sets. The Sample Sizes module is useful to develop data quality objectives (DQOs) based sampling designs and perform power evaluations needed to address statistical issues associated with the various site projects. ProUCL can process multiple variables by a group (optional) variable to perform the various tests (e.g., ANOVA, GOF, and trend tests) and compute decision statistics including UCLs, UPLs, UTLs, and USLs, a capability not available in several commercial software packages.

It is anticipated that the availability of the statistical methods in ProUCL covering a wide range of data sets of varying skewness and sample sizes will help the decision makers in making more informative and correct decisions at the various CERCLA and RCRA sites.

To facilitate the computation of UCL of mean based upon Incremental Sampling Methodology (ISM) samples (ITRC, 2012); the minimum sample size requirement has been lowered to 3, so that one can compute a UCL95 based upon actual ISM data sets of sizes  $\geq 3$ . At many sites, a large amount of valuable discrete onsite and/or offsite background data are already available which cannot be directly compared with the newly collected ISM data. The BISS module of ProUCL can be used on a large existing discrete background data set to simulate background incremental samples (BIS). The availability of a large discrete background data set collected from areas with geological formations and conditions comparable to onsite DU(s) is a requirement for successful application of this module. The values of the simulated BIS data are not directly available/visible to users; however, the BIS data can be accessed by the various modules of ProUCL 5.1 to perform desired statistical evaluations. For example, the simulated BIS data can be merged with the actual ISM field data after comparing the two data sets using a two-sample t-test; the simulated BIS or the merged BIS and ISM data can be used to compute a UCL of the mean, a UPL, or an UTL. It should be pointed out that the BISS module is still under investigation, and will be released for general public in a future version of ProUCL.