Division of Materials and Waste Management
May 2019

Calculating 95% Upper Confidence Level (UCL)

Example Calculation of the 95% UCL for a Normal Mean

Ten samples of material are taken to demonstrate that the material meets the beneficial use standards in Table 1 in the Guidance Document. The samples are obtained using a simple random sampling design. For each constituent tested, you must obtain 95% UCL. Analysis of the samples for lead generated the following results: 160, 175, 210, 220, 230, 240, 245, 270, 310, and 380 ppm. The limit for lead is 300 ppm.

Step 1: Mean and Standard Deviation Calculation

Assuming a normal model is acceptable (using the Shapiro-Wilk test), the mean and standard deviation should be calculated. The mean and standard deviation can be obtained using statistical software or by hand, using the following equations.

$$\boldsymbol{Mean} = \frac{Sum\ of\ Data}{Total\ Number\ of\ Data\ Points}$$

$$\boldsymbol{Mean} = \frac{(160 + 175 + 210 + 220 + 230 + 240 + 245 + 270 + 310 + 380)}{10} = \boldsymbol{244\ ppm}$$

Standard deviation for a sample:

1) Subtract the Mean from each data point and square the result:

Data Point 1: $(160 - 244)^2 = 7056$
Data Point 2: $(175 - 244)^2 = 4761$
Data Point 3: $(210 - 244)^2 = 1156$
Data Point 4: $(220 - 244)^2 = 574$
Data Point 5: $(230 - 244)^2 = 196$
Data Point 6: $(240 - 244)^2 = 16$
Data Point 7: $(245 - 244)^2 = 1$

Data Point 8: $(270 - 244)^2 = 676$

Data Point 9: $(310 - 244)^2 = 4356$

Data Point 10: $(380 - 244)^2 = 18496$

2) Divide the sum of those squared differences by the number of data points minus 1:

$$\frac{7056 + 4761 + 1156 + 574 + 196 + 16 + 1 + 676 + 4356 + 18496}{(10 - 1)} = 4143.1$$

3) Take the square root of that number to obtain the standard deviation:

$$\boldsymbol{Standard\ Deviation} = \sqrt{4143.1} = \mathbf{64.4}\ \boldsymbol{ppm}$$

Step 2: T-Value and 95% UCL Calculation

1) Find your T-Value from Table 1 for the number of samples taken:

# of samples minus 1 (n-1)	T Value for 95% UCL	# of samples minus 1 (n-1)	T Value for 95% UCL
1	6.314	11	1.796
2	2.920	12	1.782
3	2.353	13	1.771
4	2.132	14	1.761
5	2.015	15	1.753
6	1.943	16	1.746
7	1.895	17	1.740
8	1.860	18	1.734
9	1.833	19	1.729
10	1.812	20	1.725

Table 1: T Values

2) The UCL is calculated as follows:

$$\boldsymbol{UCL} = 244 + \left(1.833 \times \frac{64.4}{\sqrt{10}}\right) \approx \mathbf{281}\ \boldsymbol{ppm}$$

Step 3: Compare 95% UCL to the Table of Constituents

Compare the 95% UCL calculated in Step 2 to the limit in the table of constituents for that specific constituent. By showing that the 95% UCL for the sample, which is 281 ppm, is less than the 300-ppm limit for lead, we can conclude with at least 95% confidence that the mean concentration of the constituent in the material is less than 300 ppm.

What Should I do with My Nondetect Results?

U.S. EPA's ProUCL Statistical Software for Environmental Applications for Data Sets with and without Nondetect (ND) Observations was specifically designed to evaluate environmental data sets with limited sample numbers, ND results, and skewed data. It can be run on environmental data sets with and without ND data samples. ProUCL requires no formal background in statistics, but some statistical training is helpful to understand the assumptions and input requirements for statistical tests used in decision making. Input data sets are straightforward, requiring columns of detected values for contaminants and whether each value is a detect or a ND at the quantitation limit. Desired statistical tests can be selected from drop-down menus, and relevant options from subsequent menus. Data can be evaluated for fit to normal, lognormal, or gamma distributions. Outputs include recommendations, cautions, and cited references.

Resources for Statistical Analysis

- *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*
- *ProUCL Version Technical Guide*.

For data sets containing NDs with multiple detection limits (DLs) or reporting limits (RLs), 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). In addition to computing general statistics, ProUCL has goodness-of-fit (GOF) tests for normal, lognormal, and gamma distributions, and parametric and nonparametric methods including bootstrap methods for skewed data sets for computation of decision-making statistics. Ohio EPA highly recommends using ProUCL for all statistical computations. U.S. EPA offers tutorials on ProUCL. The following information was taken from the ProUCL 5.1 user's guide:

The user informs the program about the status of a variable consisting of NDs. No qualifiers or flags (e.g., J, B, U, UJ, X, or <DL or RL) should be entered in data files with ND observations. You can copy and paste your data from an EXCEL spreadsheet into a Work Sheet xls in ProUCL.

Enter the data for variables with ND values in two columns. One column should consist 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. See the example below, 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. If this format is not followed, the program will not recognize that the data set has NDs.*

The General Statistics/With NDs option also provides simple statistics (e.g., % NDs, Max detect, Min detect, Mean) 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.

Antimony	D_Antimony	Arsenic	D_Arsenic	Barium	Cadmium	D_Cadmium	Chromium	D_Chromium
1.2	1	0	0	19	0.846	1	24	1
3.1	1	0.19	1	19.8	0.904	1	0	0
1.35	1	0	0	20.5	0.669	1	0	0
1.26	1	0	0	28.4	0.234	1	20.6	1
2.24	1	0.701	1	22.5	0.487	1	48	1
4.35	1	0	0	25.3	0.642	1	3.71	1
0.69	1	0	0	29.8	0	0	0.863	1
2.11	1	0.735	1	22.8	1.12	1	3.06	1
0.9	1			5.76	0	0	1.3	1
0	0			9.85	0	0	0.7	1

Figure A: ProUCL Example Data

For most beneficial use permits, the permittee is required to determine that the 95% UCL of the mean for each constituent in the beneficial use byproduct does not exceed its constituent concentration limit specified in the permit. DMWM highly recommends using ProUCL to calculate the 95% UCL, especially when you have NDs in your sample results. Below is an example of the output from selecting UCLs/EPCs and “All” from the drop-down menu.

Arsenic			
General Statistics			
Total Number of Observations	8	Number of Distinct Observations	4
		Number of Missing Observations	1
Minimum	0	Mean	0.203
Maximum	0.735	Median	0
SD	0.325	Std. Error of Mean	0.115
Coefficient of Variation	1.597	Skewness	1.291
Note: Sample size is small (e.g., <10), if data are collected using ISM approach, you should use			
guidance provided in ITRC Tech Reg Guide on ISM (ITRC, 2012) to compute statistics of interest.			
For example, you may want to use Chebyshev UCL to estimate EPC (ITRC, 2012).			
Chebyshev UCL can be computed using the Nonparametric and All UCL Options of ProUCL 5.1			
Normal GOF Test			
Shapiro Wilk Test Statistic	0.664	Shapiro Wilk GOF Test	
5% Shapiro Wilk Critical Value	0.818	Data Not Normal at 5% Significance Level	
Lilliefors Test Statistic	0.359	Lilliefors GOF Test	
5% Lilliefors Critical Value	0.283	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	0.421	95% Adjusted-CLT UCL (Chen-1995)	0.448
		95% Modified-t UCL (Johnson-1978)	0.429
Gamma Statistics Not Available			
Lognormal Statistics Not Available			
Nonparametric Distribution Free UCL Statistics			
Data do not follow a Discernible Distribution (0.05)			
Nonparametric Distribution Free UCLs			
95% CLT UCL	0.392	95% Jackknife UCL	0.421
95% Standard Bootstrap UCL	N/A	95% Bootstrap-t UCL	N/A
95% Hall's Bootstrap UCL	N/A	95% Percentile Bootstrap UCL	N/A
95% BCA Bootstrap UCL	N/A		
90% Chebyshev(Mean, Sd) UCL	0.547	95% Chebyshev(Mean, Sd) UCL	0.703
97.5% Chebyshev(Mean, Sd) UCL	0.92	99% Chebyshev(Mean, Sd) UCL	1.345
Suggested UCL to Use			
95% Chebyshev (Mean, Sd) UCL	0.703		
Note: Suggestions regarding the selection of a 95% UCL are provided to help the user to select the most appropriate 95% UCL.			
Recommendations are based upon data size, data distribution, and skewness.			
These recommendations are based upon the results of the simulation studies summarized in Singh, Maichle, and Lee (2006).			
However, simulations results will not cover all Real World data sets; for additional insight the user may want to consult a statistician.			