# HUMAN HEALTH RISK ASSESSMENT (HHRA) NOTE NUMBER 11
# Southern California Ambient Arsenic Screening Level

**CALIFORNIA DEPARTMENT OF TOXIC SUBSTANCES CONTROL (DTSC)**
**HUMAN AND ECOLOGICAL RISK OFFICE (HERO)**

**RELEASE DATE: December 28, 2020**

## ISSUE

Mitigation or remediation is usually not undertaken to reduce the concentration of contaminants below ambient levels, which comprise both naturally occurring background with added anthropogenic source inputs (i.e., ambient) (US EPA, 2002). Background and ambient concentrations of some inorganic elements can exceed risk-based concentrations. This includes arsenic, where background as well as ambient concentrations exceed the risk-based soil concentration of 0.11 mg/kg (residential soils screening level, DTSC 2019).

## SUMMARY OF AMBIENT ARSENIC SCREENING LEVEL DEVELOPMENT

Background inorganic elements in soil can prove problematic for risk assessment purposes because these elements detected at a site may be comprised of naturally occurring metals, regional anthropogenic contributions or a site-specific release. Arsenic is especially problematic since the risk-based soil concentration is 100-times below typical background and ambient soil concentrations at southern California sites not subject to site-specific releases of arsenic. DTSC collated a data set of soil concentrations from five southern California counties (Los Angeles, Orange, Riverside, San Bernardino and San Diego) and developed an upper-bound estimate of the regional ambient arsenic soil concentration that can be used as a screening tool for sites throughout southern California. Los Angeles County had the largest number of sites (19 school sites under DTSC regulation (Los Angeles Unified School District (LAUSD Data Set)) and the most robust arsenic data set (1097 discrete samples). This evaluation was initiated by the former DTSC Schools Division because the largest number of new schools in the State occurred within the LAUSD. The summary statistics for the LAUSD arsenic data set, excluding outliers, were calculated and the upper-bound arsenic concentration estimated using 1) the 95% Upper Confidence Limit of the 99$^{th}$ percentile ($UL_{0.95}(X_{0.99})$) of the arsenic data set, assuming a lognormal data distribution; and 2) the distribution-free, nonparametric estimate of the theoretical $UL_{0.95}(X_{0.99})$. Both statistical methods resulted in an upper-bound arsenic concentration of approximately 12 mg/kg for the Los Angeles County data set (LAUSD Data Set). The detailed statistical evaluation of the LAUSD arsenic data set can be found at https://dtsc.ca.gov/wp-content/uploads/sites/31/2018/01/Arsenic-Cleanup-Goals-Jan09-A.pdf. The 95% upper confidence limit of the mean (95% UCL) arsenic soil concentration was 2.4 mg/kg, for the Los Angeles County data set (Table 3). The DTSC residential use soil screening level for arsenic in soil is 0.11 mg/kg (DTSC 2019) resulting in a residential cancer risk estimate of 1E-04 for the upper-bound arsenic background screening concentration of 12 mg/kg and a residential cancer risk of 2E-05 for the 95% UCL arsenic concentration of 2.4 mg/kg.

Table 1, below, summarizes the county-based arsenic data sets for southern California.

**Table 1**

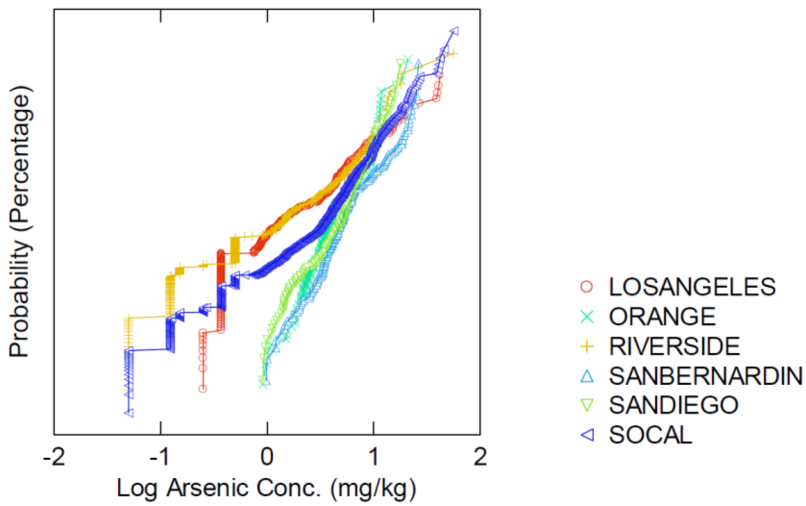
| County | No. of School Sites | No. of Arsenic Samples (after outlier removal) |
|---|---|---|
| **Los Angeles (LAUSD Study)** | 19 | 1086 |
| **Orange County** | 7 | 224 |
| **Riverside County** | 15 | 263 |
| **San Bernardino County** | 6 | 143 |
| **San Diego County** | 3 | 179 |
| **SoCal, Excluding LAUSD[1]** | 31 | 809 |

[1] All data sets were combined, excluding LAUSD, in order to compare them to the benchmark LAUSD data set.

The upper-bound arsenic concentrations were similar to LA County samples for each of the other southern California counties (Orange, Riverside, San Bernardino and San Diego counties), based on the graphical interpretation of the probability plots summarized in Figure 1.

**Figure 1**

**Probability Plots of Arsenic Data Sets by County**



The upper bound arsenic concentration for each data set converges around 12 mg/kg. The differences observed in the lower tails of the data sets reflect the differences and variability in detection limits for non-detected values in data collected over a 10-year period. Figure 2 presents the probability plots for the LAUSD and combined southern California data sets. The data sets show generally good overlap and the combined southern California arsenic data set (i.e., arsenic samples from counties other than Los Angeles) is consistent with the upper bound arsenic concentration of 12 mg/kg derived from the larger LAUSD data set.

**Figure 2**

**Probability Plots for LAUSD and Combined southern California County Data Sets**

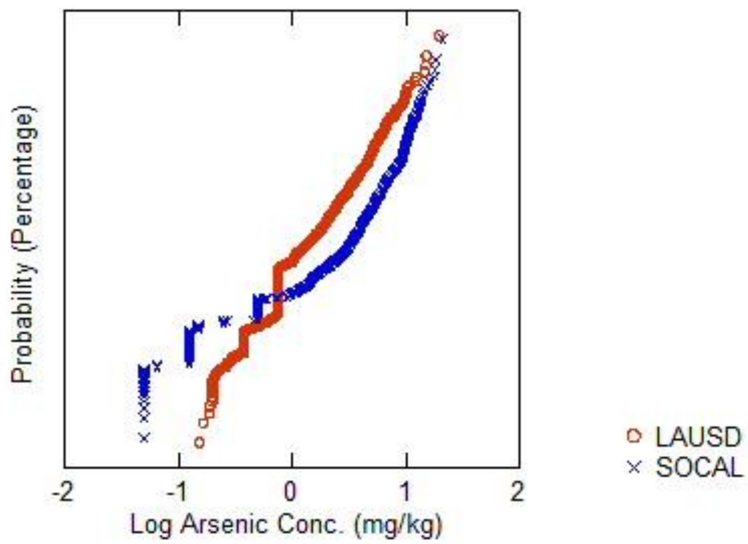


Figure 3 presents the box and whisker plots for the LAUSD Data Set and the southern California individual county data sets providing a comparison of the distributions of arsenic concentrations for each data set. All arsenic data sets were evaluated for outliers using graphical methods (probability plots and Q-Q plots) and outlier tests (Fourth Spread and ProUCL 5.1 Outlier Tests). While the data distributions are not identical, there is overall good overlap, as was shown by the distributions in the cumulative probability plots in Figures 1 and 2. The data sets with larger numbers of sites showed better overlap with the LAUSD Data Set, while data sets with fewer sites showed greater spread and variability.

**Figure 3**

**Box Plots of the Southern California Data Sets**

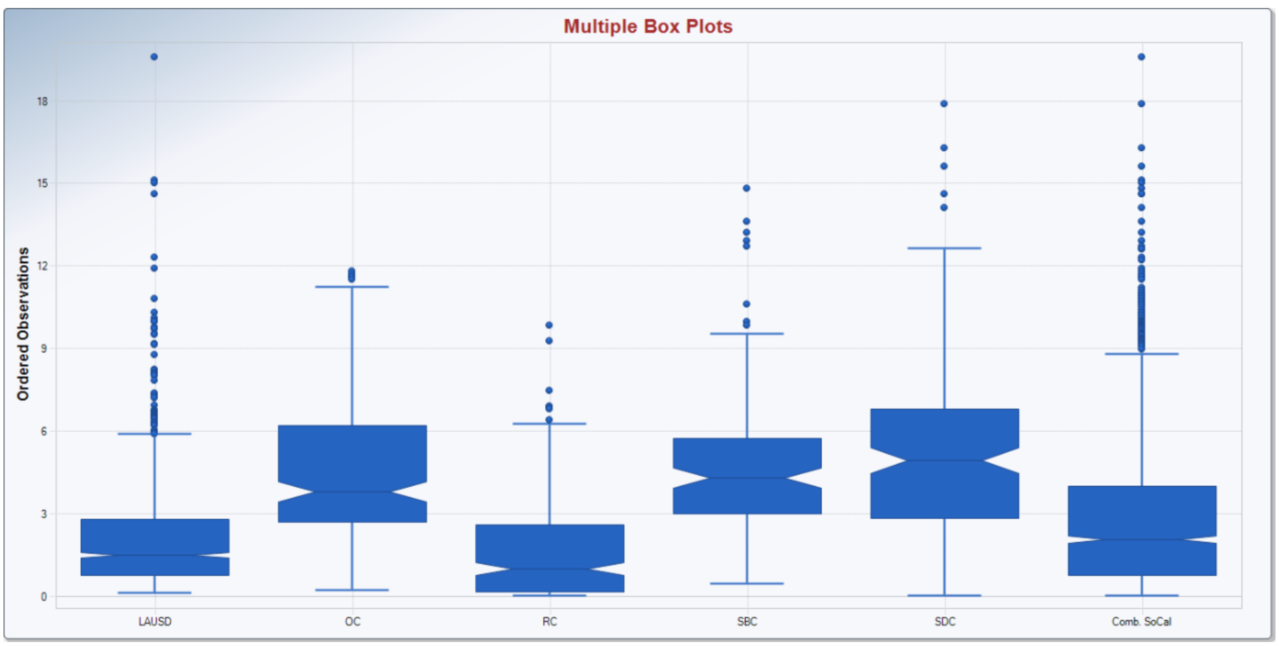
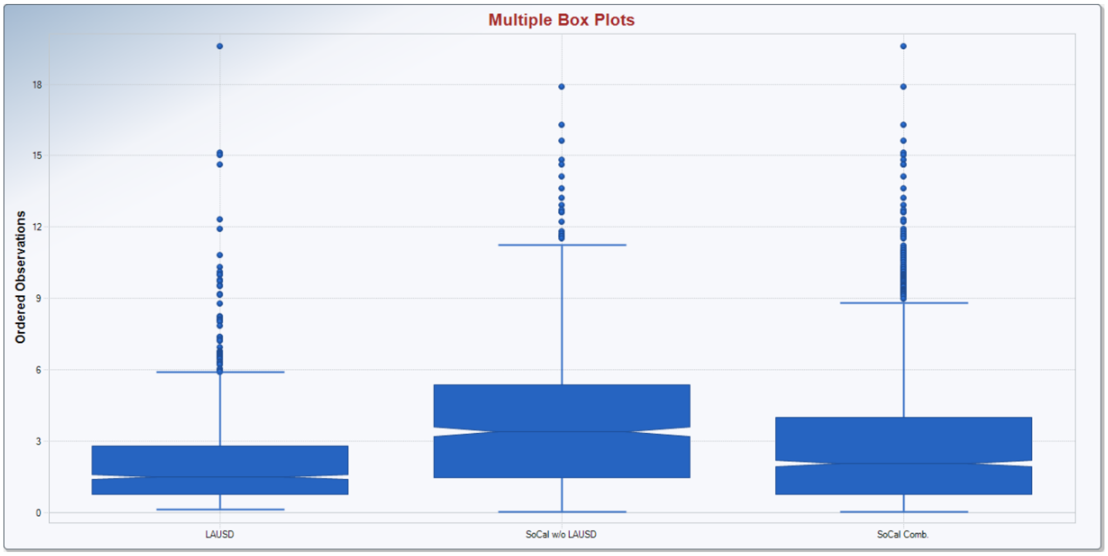


Figure 4 presents the box and whisker plots for the LAUSD data set, southern California data set without LAUSD, and the combined southern California data set.

**Figure 4**

**Box Plots, Southern California Data Sets**




To further evaluate the southern California arsenic data sets, ProUCL version 5.1 was used to estimate the Upper Tolerance Limit (UTL) arsenic concentrations, consistent with the original statistical evaluations described previously. The ProUCL UTL Outputs are included in Attachment 1. Table 2 summarizes the arsenic UTL estimates for the original LAUSD Data Set, the southern California Data Set excluding LAUSD and the Combined southern California Data Set. Given the robust number of samples in each data set, the 95% BCA Bootstrap UTL with 99$^{th}$ Percentile was used, consistent with the original statistical evaluations performed and described previously.

**Table 2**

**ProUCL 5.1 Arsenic UTL Estimates, Southern California Arsenic Data**

| Summary Statistic | LA USD Data Set | SoCal Data Set, Excluding LAUSD | Combined SoCal Data Set |
|---|---|---|---|
| **UTL with 99% Coverage** | 9.8 mg/kg | 12.9 | 11.8 |

The ProUCL UTL estimates are consistent with the previous statistical evaluations and support an estimate of 12 mg/kg for the Combined Southern California Arsenic Data Set.

Table 3 summarizes the arsenic 95% UCL concentration estimates for the original LAUSD Data Set, the southern California Data Set excluding LAUSD and the Combined southern California Data Set. ProUCL Outputs are included in Attachment 2.

**Table 3**

**Arsenic 95% UCL Estimates, Southern California Arsenic Data**

| Summary Statistic | LAUSD Data Set | SoCal Data Set, Excluding LAUSD | Combined SoCal Data Set |
|---|---|---|---|
| **95% UCL Conc.** | 2.4 mg/kg | 4.3 mg/kg | 3.1 mg/kg |

## CONCLUSIONS AND RECOMMENDED USE:

The DTSC Site Mitigation and Restoration Program (SMRP) made a risk management decision on the acceptability of estimated risk associated with the upper-bound arsenic screening level for southern California soils.

Southern California site-specific soil concentrations which exceed 12 mg/kg may be indicative of releases of arsenic. Comparison of a site-specific 95% UCL of the mean to the 12 mg/kg upper bound concentration are statistically incorrect. The comparison, or statistical test, of site-specific 95%UCL of the mean should be to the southern California 95%UCL of the mean arsenic soil concentration of 3.1 mg/kg.

Some sites may exhibit higher levels of arsenic due to naturally occurring arsenic associated with certain geologic formations. The appropriate ambient arsenic level for these sites should be determined on a site-specific basis in consultation with the DTSC Geological Services Unit (GSU) Project Geologist and Human and Ecological Risk Office (HERO) Toxicologist. Site-specific soil arsenic concentrations, including site-specific background arsenic levels, will be evaluated on a site-by-site basis, again in consultation with the project Geologist and Toxicologist.

Specific issues which may need to be addressed to determine the applicability of the southern California ambient arsenic data set are:

1. The location of the site be within the geographic five-County distribution of the samples used to set the 12 mg/kg upper bound on the ambient arsenic concentration; or,

2. GSU confirm that the lithology of the site is similar enough to the lithology of the samples used to set the southern California ambient soil arsenic metrics, that a comparison is appropriate; and,

3. GSU confirm there are no outcroppings of rock at the site with elevated arsenic concentrations that would make the southern California ambient soil arsenic metrics comparison inappropriate.

If you have any questions on this HHRA Note, please contact William Bosan, Ph.D., HERO Senior Toxicologist at (714) 484-5399, or via William.Bosan@dtsc.ca.gov.

## REFERENCES

US EPA, 2002. Role of Background in the CERCLA Cleanup Program. Office of Solid Waste and Emergency Response, Office of Emergency and Remedial Response, OSWER 9285.6-07P. August 26, 2002. [https://www.epa.gov/sites/production/files/2015-11/documents/bkgpol_jan01.pdf]

DTSC, 2019. Human Health Risk Assessment (HHRA) Note Number 3, DTSC-Modified Screening Levels (DTSC-SLs), April 2019. [https://dtsc.ca.gov/wp-content/uploads/sites/31/2019/04/HHRA-Note-3-2019-04.pdf]

# Attachment 1

# ProUCL 5.1 UTL Outputs

**Nonparametric Background Statistics for Uncensored Full Data Sets**

**User Selected Options**

| | |
|---|---|
| Date/Time of Computation | ProUCL 5.16/18/2019 1:02:12 PM |
| From File | WorkSheet.xls |
| Full Precision | OFF |
| Confidence Coefficient | 95% |
| Coverage | 95% |
| Number of Bootstrap Operations | 2000 |

**LAUSD**

**General Statistics**

| | | | |
|---|---|---|---|
| Total Number of Observations | 1086 | Number of Distinct Observations | 459 |
| Minimum | 0.154 | First Quartile | 0.75 |
| Second Largest | 15.1 | Median | 1.5 |
| Maximum | 19.6 | Third Quartile | 2.8 |
| Mean | 2.14 | SD | 2.016 |
| Coefficient of Variation | 0.942 | Skewness | 2.66 |
| Mean of logged Data | 0.412 | SD of logged Data | 0.84 |

**Critical Values for Background Threshold Values (BTVs)**

| | | | |
|---|---|---|---|
| Tolerance Factor K (For UTL) | 1.724 | d2max (for USL) | 3.898 |

**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 | 1043 | 95% UTL with 95% Coverage | 6.34 |
| Approx, f used to compute achieved CC | 1.248 | Approximate Actual Confidence Coefficient achieved by UTL | 0.937 |
| | | Approximate Sample Size needed to achieve specified CC | 1103 |
| 95% Percentile Bootstrap UTL with 95% Coverage | 6.33 | 95% BCA Bootstrap UTL with 95% Coverage | 6.3 |
| 95% UPL | 5.77 | 90% Percentile | 4.695 |
| 90% Chebyshev UPL | 8.19 | 95% Percentile | 5.743 |
| 95% Chebyshev UPL | 10.93 | 99% Percentile | 9.82 |
| 95% USL | 19.6 | | |

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

**SoCal w/o LAUSD**

**General Statistics**

| | | | |
|---|---|---|---|
| Total Number of Observations | 809 | Number of Distinct Observations | 400 |
| Minimum | 0.05 | First Quartile | 1.46 |
| Second Largest | 16.3 | Median | 3.4 |
| Maximum | 17.9 | Third Quartile | 5.37 |
| Mean | 3.832 | SD | 3.067 |
| Coefficient of Variation | 0.8 | Skewness | 1.099 |
| Mean of logged Data | 0.834 | SD of logged Data | 1.27 |

**Critical Values for Background Threshold Values (BTVs)**

| | | | |
|---|---|---|---|
| Tolerance Factor K (For UTL) | 1.736 | d2max (for USL) | 3.823 |

**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 | 778 | 95% UTL with 95% Coverage | 10.4 |
| Approx, f used to compute achieved CC | 1.28 | Approximate Actual Confidence Coefficient achieved by UTL | 0.93 |
| | | Approximate Sample Size needed to achieve specified CC | 831 |
| 95% Percentile Bootstrap UTL with 95% Coverage | 10.36 | 95% BCA Bootstrap UTL with 95% Coverage | 10.4 |
| 95% UPL | 9.975 | 90% Percentile | 8.286 |
| 90% Chebyshev UPL | 13.04 | 95% Percentile | 9.958 |
| 95% Chebyshev UPL | 17.21 | 99% Percentile | 12.88 |
| 95% USL | 17.9 | | |

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

**Comb. SoCal**

**General Statistics**

| | | | |
|---|---|---|---|
| Total Number of Observations | 1895 | Number of Distinct Observations | 677 |
| Minimum | 0.05 | First Quartile | 0.75 |
| Second Largest | 17.9 | Median | 2.05 |
| Maximum | 19.6 | Third Quartile | 4 |
| Mean | 2.862 | SD | 2.653 |
| Coefficient of Variation | 0.927 | Skewness | 1.755 |
| Mean of logged Data | 0.592 | SD of logged Data | 1.066 |

**Critical Values for Background Threshold Values (BTVs)**

| | | | |
|---|---|---|---|
| Tolerance Factor K (For UTL) | 1.704 | d2max (for USL) | 4.035 |

**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 | 1815 | 95% UTL with 95% Coverage | 9.3 |
| Approx, f used to compute achieved CC | 1.179 | Approximate Actual Confidence Coefficient achieved by UTL | 0.936 |
| | | Approximate Sample Size needed to achieve specified CC | 1919 |
| 95% Percentile Bootstrap UTL with 95% Coverage | 9.3 | 95% BCA Bootstrap UTL with 95% Coverage | 9.3 |
| 95% UPL | 8.522 | 90% Percentile | 6.212 |
| 90% Chebyshev UPL | 10.82 | 95% Percentile | 8.486 |
| 95% Chebyshev UPL | 14.43 | 99% Percentile | 11.81 |
| 95% USL | 19.6 | | |

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

# Attachment 2

# ProUCL 5.1 95% UCL Outputs

### Nonparametric UCL Statistics for Uncensored Full Data Sets

| | |
|---|---|
| User Selected Options | |
| Date/Time of Computation | ProUCL 5.16/18/2019 1:21:47 PM |
| From File | WorkSheet.xls |
| Full Precision | OFF |
| Confidence Coefficient | 95% |
| Number of Bootstrap Operations | 2000 |

**LAUSD**

**General Statistics**

| | | | |
|---|---|---|---|
| Total Number of Observations | 1086 | Number of Distinct Observations | 459 |
| | | Number of Missing Observations | 0 |
| Minimum | 0.154 | Mean | 2.14 |
| Maximum | 19.6 | Median | 1.5 |
| SD | 2.016 | Std. Error of Mean | 0.0612 |
| Coefficient of Variation | 0.942 | Skewness | 2.66 |
| Mean of logged Data | 0.412 | SD of logged Data | 0.84 |

**Nonparametric Distribution Free UCL Statistics**

Data do not follow a Discernible Distribution (0.05)

**Assuming Normal Distribution**

| **95% Normal UCL** | | **95% UCLs (Adjusted for Skewness)** | |
|---|---|---|---|
| 95% Student's-t UCL | 2.241 | 95% Adjusted-CLT UCL (Chen-1995) | 2.246 |
| | | 95% Modified-t UCL (Johnson-1978) | 2.241 |

**Nonparametric Distribution Free UCLs**

| | | | |
|---|---|---|---|
| 95% CLT UCL | 2.241 | 95% Jackknife UCL | 2.241 |
| 95% Standard Bootstrap UCL | 2.24 | 95% Bootstrap-t UCL | 2.248 |
| 95% Hall's Bootstrap UCL | 2.248 | 95% Percentile Bootstrap UCL | 2.242 |
| 95% BCA Bootstrap UCL | 2.249 | | |
| 90% Chebyshev(Mean, Sd) UCL | 2.323 | 95% Chebyshev(Mean, Sd) UCL | 2.407 |
| 97.5% Chebyshev(Mean, Sd) UCL | 2.522 | 99% Chebyshev(Mean, Sd) UCL | 2.749 |

**Suggested UCL to Use**

| | |
|---|---|
| 95% Chebyshev (Mean, Sd) UCL | 2.407 |

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.

**SoCal w/o LAUSD**

**General Statistics**

| | | | |
|---|---|---|---|
| Total Number of Observations | 809 | Number of Distinct Observations | 400 |
| | | Number of Missing Observations | 0 |
| Minimum | 0.05 | Mean | 3.832 |
| Maximum | 17.9 | Median | 3.4 |
| SD | 3.067 | Std. Error of Mean | 0.108 |
| Coefficient of Variation | 0.8 | Skewness | 1.099 |
| Mean of logged Data | 0.834 | SD of logged Data | 1.27 |

**Nonparametric Distribution Free UCL Statistics**

**Data do not follow a Discernible Distribution (0.05)**

**Assuming Normal Distribution**

| **95% Normal UCL** | | **95% UCLs (Adjusted for Skewness)** | |
|---|---|---|---|
| 95% Student's-t UCL | 4.009 | 95% Adjusted-CLT UCL (Chen-1995) | 4.013 |
| | | 95% Modified-t UCL (Johnson-1978) | 4.01 |

**Nonparametric Distribution Free UCLs**

| | | | |
|---|---|---|---|
| 95% CLT UCL | 4.009 | 95% Jackknife UCL | 4.009 |
| 95% Standard Bootstrap UCL | 4.01 | 95% Bootstrap-t UCL | 4.023 |
| 95% Hall's Bootstrap UCL | 4.012 | 95% Percentile Bootstrap UCL | 4.009 |
| 95% BCA Bootstrap UCL | 4.026 | | |
| 90% Chebyshev(Mean, Sd) UCL | 4.155 | 95% Chebyshev(Mean, Sd) UCL | 4.302 |
| 97.5% Chebyshev(Mean, Sd) UCL | 4.505 | 99% Chebyshev(Mean, Sd) UCL | 4.904 |

**Suggested UCL to Use**

| | |
|---|---|
| 95% Chebyshev (Mean, Sd) UCL | 4.302 |

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.

**Comb. SoCal**

### General Statistics

| | | | |
|---|---|---|---|
| Total Number of Observations | 1895 | Number of Distinct Observations | 677 |
| | | Number of Missing Observations | 0 |
| Minimum | 0.05 | Mean | 2.862 |
| Maximum | 19.6 | Median | 2.05 |
| SD | 2.653 | Std. Error of Mean | 0.061 |
| Coefficient of Variation | 0.927 | Skewness | 1.755 |
| Mean of logged Data | 0.592 | SD of logged Data | 1.066 |

### Nonparametric Distribution Free UCL Statistics

Data do not follow a Discernible Distribution (0.05)

### Assuming Normal Distribution

| 95% Normal UCL | | 95% UCLs (Adjusted for Skewness) | |
|---|---|---|---|
| 95% Student's-t UCL | 2.962 | 95% Adjusted-CLT UCL (Chen-1995) | 2.965 |
| | | 95% Modified-t UCL (Johnson-1978) | 2.963 |

### Nonparametric Distribution Free UCLs

| | | | |
|---|---|---|---|
| 95% CLT UCL | 2.962 | 95% Jackknife UCL | 2.962 |
| 95% Standard Bootstrap UCL | 2.962 | 95% Bootstrap-t UCL | 2.966 |
| 95% Hall's Bootstrap UCL | 2.97 | 95% Percentile Bootstrap UCL | 2.966 |
| 95% BCA Bootstrap UCL | 2.961 | | |
| 90% Chebyshev(Mean, Sd) UCL | 3.045 | 95% Chebyshev(Mean, Sd) UCL | 3.128 |
| 97.5% Chebyshev(Mean, Sd) UCL | 3.243 | 99% Chebyshev(Mean, Sd) UCL | 3.469 |

### Suggested UCL to Use

| | |
|---|---|
| 95% Chebyshev (Mean, Sd) UCL | 3.128 |

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.