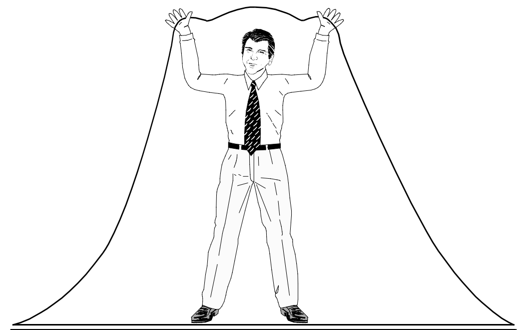
United States Environmental Protection Agency
Environmental Sciences Division
P.O. Box 93478
Las Vegas, NV 89193-3478
November 1999

OFFICE OF RESEARCH AND DEVELOPMENT
TECHNOLOGY SUPPORT PROJECT

EPA

Robust Statistical Intervals for Performance Evaluations



ESD Innovative Technology

Introduction

Environmental samples collected at Superfund sites are routinely analyzed by the various commercial laboratories participating in quality assurance/quality control (QA/QC) programs such as the Contract Laboratory Program (CLP) of the U.S. EPA. The EPA Superfund CLP periodically evaluates, through the performance evaluation (PE) quarterly blind (QB) studies, the competence of participating laboratories in the quantitative analysis of prepared materials. Identically prepared PE samples are sent to participants. PE samples contain amounts of various organic or inorganic compounds known only to the evaluator. Laboratories are expected to report analytical results that are relatively close to the known amount. However, in practice, the recoveries reported by the participants may differ significantly from the “true” spiked amount.

In a PE study, the objective may be to obtain: (1) an interval estimate (LCL, UCL) for the overall mean recovery *(where LCL and UCL represent the lower and upper confidence limits for mean, μ, respectively),* (2) an interval estimate (LSL, USL) within which the majority of the participants are expected to report their analytical results *(where LSL and USL simultaneous limits, respectively),* or (3) an interval estimate (LPL, UPL) for a delayed result, x_0, reported by a participant *(where LPL and UPL represent the lower and upper prediction limits, respectively)*. These intervals are significantly different from each other and care must be exercised to use them appropriately. For example, at a polluted site the objective may be to obtain a threshold value estimating the background level contamination prior to any activity that polluted the site. Here, the upper simultaneous limit, USL, and not the upper confidence limit, UCL, for the population mean may be used. It is inappropriate to compare an individual observation, x_i, with the UCL for the population mean, μ, and expect an adequate coverage for all of the values of x_i, as is sometimes mistakenly done in practice.

There are two main issues that need to be considered. First, an adequate interval estimate should be used for a typical application. The use of the confidence interval (CI) for the mean, μ, or a prediction interval for a single future observation is inappropriate when the objective is to obtain a statistical interval providing simultaneous coverage for the majority of the participants. The test-statistics and the associated critical values change from application to application. Secondly, appropriate statistical methods need to be used to obtain robust and resistant estimates of the population mean, μ, and variance, σ^2. It is important that the degrees of freedom (df) be computed accurately by making the appropriate adjustment for the outliers. All of these measures, when considered collectively, result in more accurate and reliable interval estimates.

Scientists at the National Exposure Research Laboratory’s (NERL’s) Environmental Sciences Division (ESD) in Las Vegas have studied the CLP database extensively and have developed improved methods for assessing some QA measurements. Chief among these improvements is a more robust statistical method, based on simultaneous confidence intervals, for evaluating the performance of the participating laboratories in the QA/QC programs of the U.S. EPA.

The Current Process

Let $x_1, x_2, \ldots, x_n$ represent the recoveries of a certain compound reported by the *n* participants in a typical PE study. The classical maximum likelihood estimates (MLEs) of population mean, μ, and standard deviation (sd), σ, are the sample mean, $\bar{x}$, and sample sd, *s*, respectively.

The U.S. EPA evaluates the analytical results reported by the participants using statistical quality control (SQC) techniques based on the classical MLEs, $\bar{x}$ and *s*. The classical estimates, $\bar{x}$ and s^2, get distorted in the presence of outliers and may result in unreliable and imprecise estimates of the above-mentioned

The Current Process
Continued

intervals. Thus, the outlying observations inherent in most environmental applications can distort the entire estimation process, which in turn can result in incorrect decisions. The robust statistical intervals should be used when outliers are present.

Horn et al. (1988) used the Biweight influence function to obtain a robust prediction interval and recommended its use to assess the performance of a *future* (delayed) result reported by a **single** participant in a PE study of the U.S. EPA. However, in PE studies, one of the main objectives is to obtain adequate acceptance regions within which **most** of the participants are expected to report their analytical results simultaneously. The prediction interval currently used is not appropriate to provide simultaneous coverage for the majority of the participants. Moreover, the Biweight function does not perform well in samples of small sizes ($n \leq 15$). In the current Biweight procedure, no adjustment for the outliers is made in the computation of the df used to obtain the critical values of the associated test-statistics and, consequently, inflated df are used to obtain these critical values.

The Proposed Process

A more statistically rigorous approach to determine misquantified analytes in PE studies has been discussed by Singh and Nocerino (1993). Comparisons are made with the existing techniques. The “proposed” PROP acceptance interval is a *simultaneous confidence interval* with a built-in outlier detection criterion. The PROP simultaneous confidence intervals: (1) use the robust estimates of population mean, μ, and variance, σ^2, which are not distorted by the presence of multiple outliers (Singh, 1993), (2) use more accurate estimates of df to obtain critical values of the associated test-statistics, and (3) by definition, are better suited for such PE studies and provide adequate simultaneous coverage for the majority of the participants. Some of these intervals are given as below. In the following equations, $\bar{x}^*$, s^*, and wsum refer to the robust estimate of μ, the robust estimate of σ, and the sum of the squared weights, respectively, and are given by:

$$\bar{x}^* = \Sigma w_1(d_i)x_i / \Sigma w_1(d_i) \quad \text{and} \quad s*^2 = \Sigma w_2(d_i)(x_i - \bar{x}^*)^2 / df \qquad \textbf{(1)}$$

where the weights are obtained using the PROP or the Huber (Singh, 1993) influence functions.

The distances, d_i^2, are given by $d_i^2 = (x_i - \bar{x})^2/s^2$ and are identically distributed as a beta distribution:

$$\left((n-1)^2/n\right) \Downarrow (1/2, (n-2)/2)$$

- The (1-α)100% *confidence interval* (LCL, UCL) for the population mean, μ, is given by:

$$P(\bar{x}^* - \frac{t_{df,\alpha/2}s^*}{\sqrt{wsum}} \leq \mu \leq \bar{x}^* + \frac{t_{df,\alpha/2}s^*}{\sqrt{wsum}}) = 1 - \alpha \qquad \textbf{(2)}$$

where $t_{df,\alpha/2}$ represents the critical value from the Student’s t-distribution.

- The (1-α) 100% *simultaneous confidence interval* (LSL, USL) for the majority of the participants is developed as follows. Let $d^2_{m,\alpha}$ represent the α (100%) critical values for the distribution of max (d_i^2). The $d^2_{m,\alpha}$ simultaneous interval with a built-in outlier identification criterion is given by P(max(d_i^2) ≤ = 1 - α, or equivalently, given by the probability statement (Singh and Nocerino, 1993),

$$P(\bar{x}^* - s^* d_{m,\alpha} \leq x_i \leq \bar{x}^* + s^* d_{m,\alpha};\ i = 1, 2, \ldots, n) = 1 - \alpha \qquad \textbf{(3)}$$

- The (1-α) 100% *prediction interval*, (LPL, UPL), *for a future observation*, x_0, is given by:

$$P(\bar{x}^* - t_{df,\alpha/2}s^*\sqrt{\frac{1}{wsum} + 1} \leq x_0 \leq \bar{x}^* + t_{df,\alpha/2}s*\sqrt{\frac{1}{wsum} + 1}) = 1 - \alpha \qquad \textbf{(4)}$$

An Example

The following data set from a QB study illustrates the differences among the above-mentioned interval estimates. Using the analytical results reported by 43 laboratories for the semivolatile chemical, 4-methylphenol, the computations for the various intervals with a confidence coefficient (CC=1-α) of 0.95 are summarized in Table 1. The estimated df obtained using the PROP procedure is df=34.39. This is expected because of the reduced weights assigned to the outlying observations. Using Iglewicz’s (1983) recommendation, one might use a substantially smaller number of df, (0.7)(42) ≈ 29. Notice that the PROP sd is also much smaller, again due to the negligible contribution of the outliers. Figures 1 and 2 show the classical and robust simultaneous intervals. The classical interval in Figure 1 is distorted by the outlying observations (e.g., number 28, circled in the figures). The robust interval estimate of Figure 2 is not influenced by the outliers and provides appropriate simultaneous coverage to the majority of the participants.

An Example
Continued

It should be emphasized here that the outliers do not necessarily represent poor performance laboratories (bad values). In a typical PE study, a high discordant recovery close to the true spiked amount may indicate extremely good performance by the associated participant. However, consistent occurrences of such high values for the same participant in several PE studies may call for an examination regarding the appropriate use of the analytical method. In any case, all of the outliers, low or high, should be down-weighted appropriately so that the resulting estimates will correspond to the estimates of the parameters of the dominant population representing the majority of the participants.

The procedure described here: (1) identifies multiple outliers, (2) uses appropriate test-statistics, (3) computes the adjusted df associated with the test-statistics by assigning reduced weights to the outlying observations, and (4) provides more precise and accurate estimates of the underlying population parameters and the associated intervals. The acceptance intervals based upon the PROP method result in higher probabilities of correctly estimating the performance of a laboratory. Using the PROP method, EPA data analysts can appropriately assess the performance of a member laboratory in a PE study by considering all of the relevant factors that affect bottom line performance. The computations and graphs for these intervals were obtained using the Scout software package developed by Lockheed Martin, Las Vegas, Nevada, for the U.S. EPA.

Table 1. Sampling Statistics and Intervals Obtained Using the Four Estimation Procedures for the PE Analytical Results Data Set Reported by 43 Laboratories Participating in the CLP (CC=0.95).

	df	mean	sd	LCL	UCL	LPL	UPL	LSL	USL
MLE	42.00	27.56	5.38	25.90	29.21	16.57	38.54	11.05	44.06
Huber	40.49	27.83	4.62	26.40	29.26	18.38	37.27	13.72	41.93
PROP	34.39	29.01	2.78	28.08	29.93	23.29	34.73	20.72	37.29
Biwt	42.00	28.38	4.56	26.98	29.78	19.07	37.69	14.40	42.36

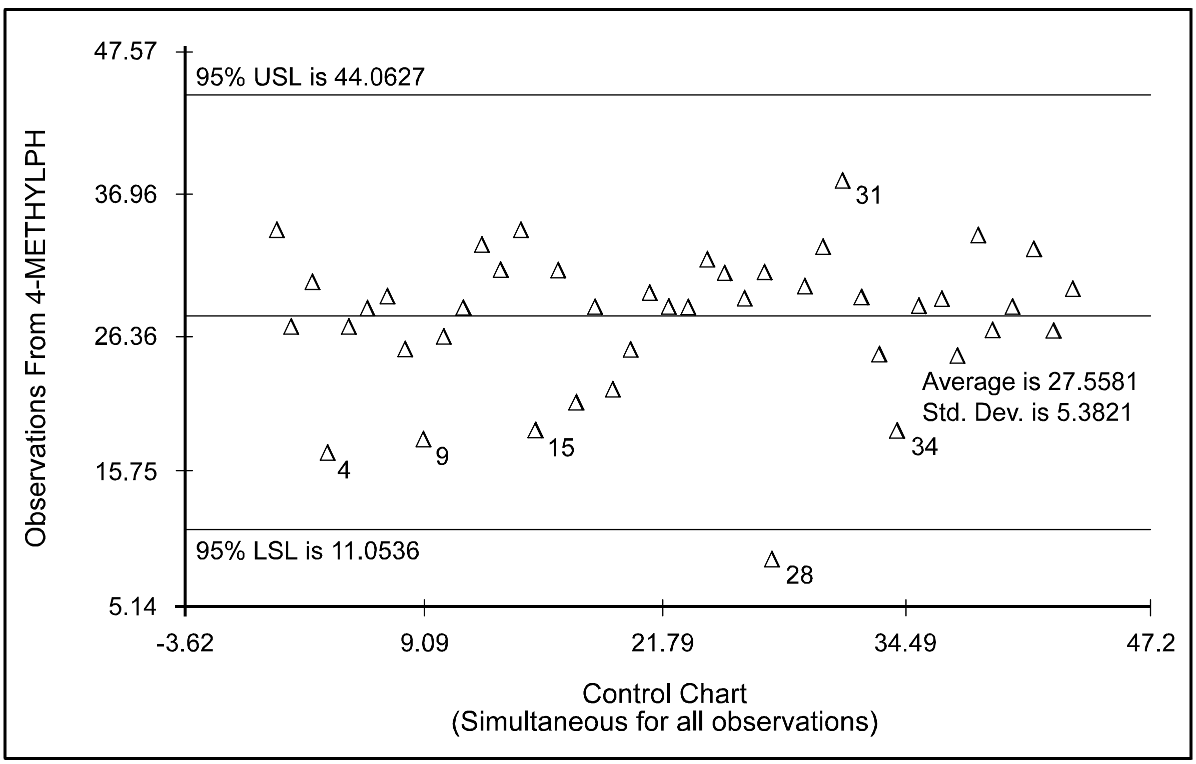


Figure 1. Classical simultaneous interval for 4-methylphenol.

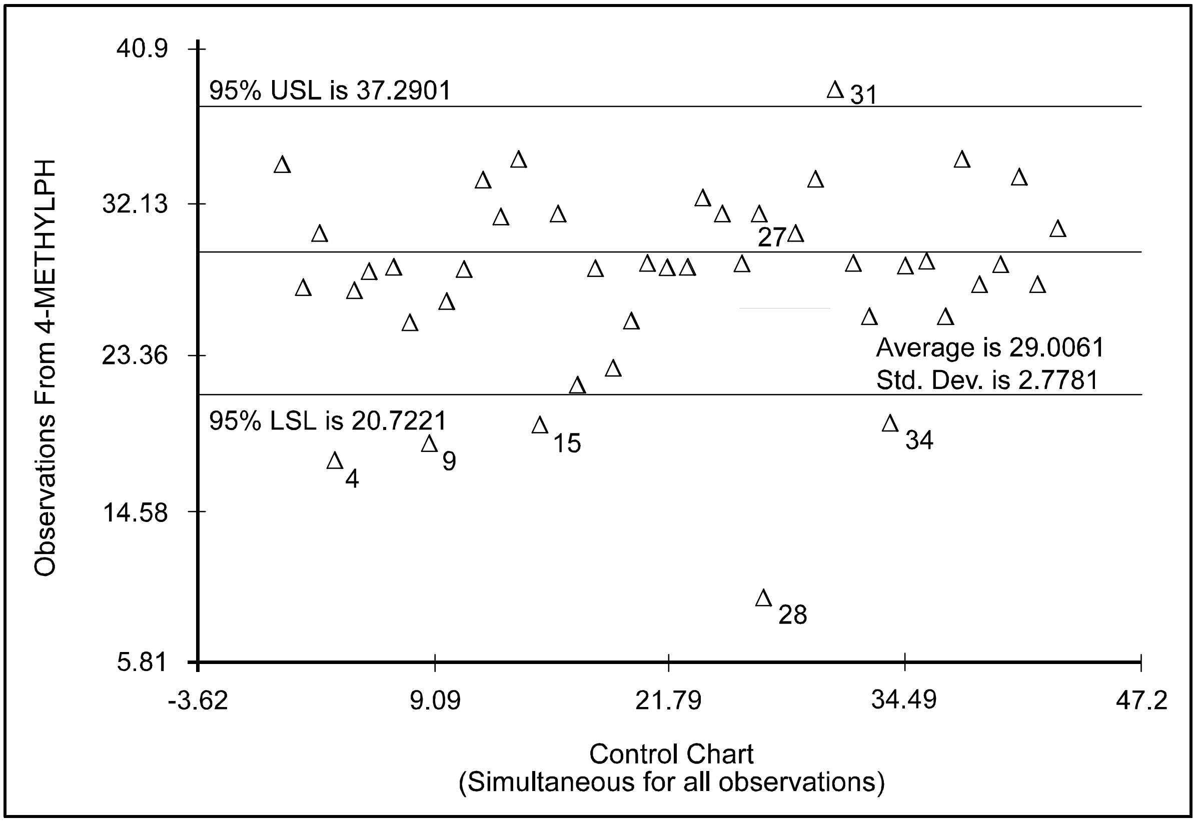


Figure 2. Robust simultaneous interval for 4-methylphenol.

References

1. Horn, P. S., Britton, P. W., and Lewis, D. F., "On the Prediction of a Single Future Observation from a Possibly Noisy Sample," *The Statistician*, 37, 165-172, 1988.

2. Iglewicz, B., "Robust Scale Estimators and Confidence Intervals for Location," in *Understanding Robust and Exploratory Data Analysis*, Hoaglin, D. C., Mosteller, F., and Tukey, J. W., *eds.*, New York, John Wiley, 1983.

3. Singh, A., "Omnibus Robust Procedures for Assessment of Multivariate Normality and Detection of Multivariate Outliers," in *Multivariate Environmental Statistics*, 445-488. Patil, G. P., and Rao, C. R., *eds.*, North Holland/Elsevier Science Publishers, 1993.

4. Singh, A., Nocerino, J. M., "Robust Intervals For Some Environmental Applications," *Chemometrics and Intelligent Laboratory Systems*, 37, 1997, 55-69.

For Further Information

For information about the Technology Center at the ESD, contact:

Mr. J. Gareth Pearson, Director
Technology Support Center
U.S. Environmental Protection Agency
Environmental Sciences Division
P.O. Box 93478
Las Vegas, NV 89193-3478

Tel.: (702) 798-2270

For more detailed information about the correct use of robust statistical intervals, contact:

Mr. John Nocerino
U.S. Environmental Protection Agency
Environmental Sciences Division
P.O. Box 93478
Las Vegas, NV 89193-3478

Tel.: (702) 798-2110

COMMUNICATION • TRAINING • TECHNOLOGY SUPPORT •
Technology Support Project

The Technology Support Center fact sheet series is developed and written by Clare L. Gerlach and Anita Singh, Lockheed Martin, Las Vegas.