# Quality Control Data Analysis

Dave Lorenz and Jeffrey Martin

July 29, 2015

These examples demonstrate some of the functions and statistical methods for the analysis of water-quality control data that are available in the **smwrQW** package.

```
> # Load the smwrQW package
> library(smwrQW)
```

# 1  Statistical Concepts

It is not possible, physically or financially, to measure all occurrences of every characteristic of interest in environmental studies. For some characteristics, any direct measurement is impossible. Thus, statistical methods are necessary to make estimates of these characteristics. Such estimates can be less than satisfying, and even the subject of disbelief or derision.

## 1.1  Confidence Intervals

The uncertainty of an inferential statistic often is indicated by reporting a range of values, referred to as a "confidence interval." Confidence intervals are constructed to contain an unknown characteristic of the population, such as the mean, median, standard deviation, or a percentile, with a specified probability. The width of the confidence interval is the uncertainty due to estimation of a population characteristic based on sample data.

Confidence interval for the mean (include uncensored and censored in all of these)

Confidence interval for the median

UCL for a percentile

CLs for a proportion: (Use binom.test–binom.test(5,20, p=.25, conf.level=.9))

# 2    Analysis of Blanks

Blanks are used to estimate the positive bias that can be caused by extraneous contamination introduced into environmental samples during collection, processing, shipment, and laboratory analysis. Evaluation of data from field blanks depends on the inference space represented by the blanks. In general, there are two possibilities: (1) a single blank is prepared to represent potential sources of contamination that affect a specific, small set of environmental samples, or (2) multiple blanks are prepared periodically over time and space to represent potential sources of contamination that might affect a much larger set of environmental samples.

# 3 Analysis of Spikes

Spikes are used to estimate the positive or negative bias that can affect the measured results for environmental samples because of analyte degradation or problems with the analytical methods. This bias is estimated by determining the recovery of known concentrations of the analytes in the spiked sample. Calculation of recovery for matrix spikes requires a separate environmental sample to determine the background concentration of the analyte in the unspiked matrix. Recovery in the spiked matrix samples can be compared to some criteria or to typical recovery for the analytical method based on laboratory reagent spikes.

# 4    Analysis of Replicates

Replicates are used to measure variability, which is defined as the random error in independent measurements as the result of repeated application of the measurement process under identical conditions. Statistical evaluation of replicate variability is based on the standard deviation of measured values in the primary environmental sample and the replicate sample or samples. If only one set of a large number of replicates was collected, the standard deviation could be calculated directly; however, the general practice is to collect many sets of a small number of replicates under different conditions.

# References

[1] Helsel, D.R. 2012, Statistics for Censored Environmental Data Using Minitab and R: New York, Wiley, 324 p.

[2] Lorenz, D.L., 2016, smwrQW–an R package for managing and analyzing water-quality data, version 1.0.0: U.S. Geological Survey Open File Report 2016-XXXX.

[3] Mueller, D.K., Schertz, T.L., Martin, J.D., and Sandstrom, M.W., 2015, Design, analysis, and interpretation of field quality-control data for water-sampling projects: U.S. Geological Survey Techniques and Methods book 4, chap. C4, 54 p.