ELSEVIER

Chemosphere 65 (2006) 2434–2439

CHEMOSPHERE

www.elsevier.com/locate/chemosphere

Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it

Dennis R. Helsel *

U.S. Geological Survey, P.O. Box 25046, MS 964, Lakewood, CO 80225, USA

Received 3 January 2006; received in revised form 17 April 2006; accepted 18 April 2006
Available online 5 June 2006

Abstract

The most commonly used method in environmental chemistry to deal with values below detection limits is to substitute a fraction of the detection limit for each nondetect. Two decades of research has shown that this fabrication of values produces poor estimates of statistics, and commonly obscures patterns and trends in the data. Papers using substitution may conclude that significant differences, correlations, and regression relationships do not exist, when in fact they do. The reverse may also be true. Fortunately, good alternative methods for dealing with nondetects already exist, and are summarized here with references to original sources. Substituting values for nondetects should be used rarely, and should generally be considered unacceptable in scientific research. There are better ways.
Published by Elsevier Ltd.

Keywords: Nondetect; Detection limit; Censored data; Statistics

1. Introduction

In his satire ''Hitchhiker's Guide To The Galaxy'', Douglas Adams wrote of his characters' search through space to find the answer to ''the question of Life, The Universe and Everything''. In what is undoubtedly a commentary on the inability of science to answer such questions, the computer built to process it determines that the answer is 42. There is beauty in a precise answer – a totally arbitrary, but precise, answer.

Environmental scientists often provide a similar answer to a different question – what to do with ''nondetect'' data? Nondetects are low-level concentrations of organic or inorganic chemicals with values known only to be somewhere between zero and the laboratory's detection/reporting limits. Measurements are considered too imprecise to report as a single number, so the value is commonly reported as being less than an analytical threshold, for example ''<1''. Long considered second class data, nondetects complicate the familiar computations of descriptive statistics, of testing differences among groups, and of correlation coefficients and regression equations.

The worst practice when dealing with nondetects is to exclude or delete them. This produces a strong upward bias in all subsequent measures of location such as means and medians. After exclusion, comparisons are being made between the mean of the top 20% of concentrations in one group versus the top 50% of another group, for example. This provides little insight into the original data. Excluding nondetects removes the primary signal that should be sent to hypothesis tests – the proportion of data in each group that lies above the reporting limit(s), the shift producing the difference between 20% and 50% detects.

The most common procedure within environmental chemistry to deal with nondetects continues to be substitution of some fraction of the detection limit. This method is better labeled as ''fabrication'', as it reports and uses a single value for concentration data where a single value is unknown. Within the field of water chemistry, one-half is the most commonly used fraction, so that 0.5 is used as if it had been measured whenever a <1 (detection limit

* Tel.: +1 303 2365340; fax: +1 303 2361425.
E-mail address: dhelsel@usgs.gov

0045-6535/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.chemosphere.2006.04.051

of 1) occurs. For air chemistry, one over the square root of two, or about 0.7 times the detection limit, is commonly used. Douglas Adams might have chosen 0.42. Studies 20 years ago found substitution to be a poor method for computing descriptive statistics (Gilliom and Helsel, 1986). Subsequent justifications for using one-half the reporting limit when data follow a uniform distribution (Hornung and Reed, 1990) only considered estimation of the mean. Any substitution of a constant fraction of reporting limits will distort estimates of the standard deviation, and therefore all (parametric) hypothesis tests using that statistic. This is illustrated later using simulations. Also, justifications such as these have never considered errors due to changing reporting limits arising from changing interferences between samples or similar causes. Substituting values tied to those changing limits introduces a signal into the data that was not present in the media sampled. Substituted values using a fraction anywhere between 0 and 0.99 times the detection limit are equivalently arbitrary, equivalently precise, equivalently wrong.

Examples of substitution of fractions of the detection limit for nondetects abound in the scientific literature. McCarthy et al. (1997) computed descriptive statistics of organic compounds in relatively uncontaminated areas. They employed substitution of a 'sliding scale' fraction of the detection limit, setting the fraction to be a function of the proportion of nondetects in the data set. The accuracy and value of their resulting statistics is unknowable. Another scientist using different fractions to provide values for nondetect data would get different results. Similarly, Tajimi et al. (2005) computed correlation coefficients after substituting one-half the detection limit for all nondetects. They found no correlations between dioxin concentrations and the factors they investigated. Was this because there were none, or was it the result of their data substitutions? Barringer et al. (2005) tested for differences in mercury concentrations of groundwater in areas of differing land use. Were their results due to concentrations actually found in the aquifer, or to the fact that one-half the detection limit was substituted for some nondetects, while other nondetects were simply deleted? Finally, Rocque and Winker (2004) substituted random values between zero and the detection limits in order to compute sums and test hypotheses. How would those results have changed if different random values had been assigned?

Statisticians use the term "censored data" for data sets where specific values for some observations are not quantified, but are known to exceed or to be less than a threshold value. Techniques for computing statistics for censored data have long been employed in medical and industrial studies, where the length of time is measured until an event occurs such as the recurrence of a disease or failure of a manufactured part. For some observations the event may not have occurred by the time the experiment ends. For these, the time is known only to be greater than the experiment's length, a censored "greater-than" value. Methods for computing descriptive statistics, testing hypotheses, and performing correlation and regression are all commonly used in medical and industrial statistics, without substituting arbitrary values. These methods go by the names of "survival analysis" and "reliability analysis". There is no reason why these same methods could not also be used in the environmental sciences, but to date, their use is rare.

Two early examples using methods for censored data in environmental applications are Millard and Deverel (1988) and She (1997). Millard and Deverel (1988) pioneered the use of two-group survival analysis methods in environmental work, testing for differences in metals concentrations in the groundwaters of two aquifers. Many nondetected values were present, at multiple detection limits. They found differences in zinc concentrations between the two aquifers using a survival analysis method called a score test. Had they substituted one-half the detection limit for zinc concentrations and run a *t*-test, they would not have found those differences (Helsel, 2005b). She (1997) computed descriptive statistics of organics concentrations in sediments using a survival analysis method called Kaplan-Meier, the standard procedure in medical statistics. Means, medians and other statistics were computed without substitutions, even though the data contained 20% nondetects censored at eight different detection limits. Substitution would have given very different results. More recently, Baccarelli et al. (2005) reviewed a variety of methods for handling nondetects in a study of dioxin exposure. They found that imputation methods designed for censored data far outperformed substitution of values such as one-half the detection limit. Other examples of the use of survival analysis methods for environmental data can be found in Helsel (2005b).

The goal of this paper is to clearly illustrate the problems with substitution of arbitrary values for nondetects. Methods designed expressly for censored data are directly compared to results using arbitrary substitution of values for nondetects when computing summary statistics, regression equations, and hypothesis tests.

2. Methods

Statisticians generate simulated data for much the same reasons as chemists prepare standard solutions – so that the conditions are exactly known. Statistical methods are then applied to the data, and the similarity of their results to the known, correct values provides a measure of the quality of each method. Fifty X, Y pairs of data were generated for this study with X values uniformly distributed from 0 to 100. The Y values were computed from a regression equation with slope = 1.5 and intercept = 120. Noise was then randomly added to each Y value so that points did not fall exactly on the straight line. The result is data having a strong linear relation between Y and X with a moderate amount of noise in comparison to that linear signal.

The noise applied to the data represented a "mixed normal" distribution, two normal distributions where the second had a larger standard deviation than the first. All