# Analysis of Transformed Environmental Data with Detection Limits

Jade Freeman
*The George Washington University, Department of Statistics*
*Washington, DC 20052, USA*
*lee-freeman.jade@epa.gov*

Reza Modarres
*The George Washington University, Department of Statistics*
*Washington, DC 20052, USA*
*reza@gwu.edu*

## 1. Introduction

There have been growing concerns about the potential health and ecological effects of various environmental contaminants. Increasing research resources have been committed to scientific studies of environmental issues and problems. Several government agencies are charged with setting and monitoring environmental standards for potentially harmful contaminants in air, water, soil, and food. Often, interest center on the distribution of the mean concentration of such contaminants. For example, risk is defined as the product of dose and the probability of an adverse effect per unit dose. Here, dose is the amount of toxicant or the number of microbes an individual consumes. The mean concentration of contaminants in an environment is often used as a direct measure of dose. For example, USEPA sets the limits for the amount of pollutants in the industrial wastewater. Such limits are based on the mean concentrations of the pollutants in the effluent from a model wastewater treatment process.

Clearly, a quantitative study of such toxicants and microbes requires statistical analysis. Unfortunately, most environmental data are not well suited for routine statistical analysis. Often environmental data include extreme observations at both low and high levels of concentrations. To complicate the problem further, the observations at low levels are often below the range of the measuring device. Such observations are called "censored" and only reported as less than the "detection limit". USEPA (1992) defines detection limit as the lowest concentration level, which can be determined to be different from a blank.

## 2. The Use of Transformation and E-M Algorithm

Transformations have been a prominent tool in analysis of non-normal data. Existing statistical procedures have been relatively successful in dealing with censored observations. However, most such procedures are model-based. That is, parameter estimation and statistical inference are carried out after an underlying model for the data is stipulated. Shumway, Azari and Jonhson (1989) used Box-Cox (BC) transformations to estimate the mean concentration of environmental contaminants. Here, we would like to estimate the mean in the untransformed scale. The strategy is to estimate the mean and variance after a Box-Cox transformation to normality and estimate the mean concentration after a back transformation to the original scale. The E-M algorithm is used to maximize the likelihood of censored observations under normality. The invariance property of the maximum likelihood estimates is used to obtain estimates in the original scale. Confidence interval based for the mean concentration can be obtained by bootstrap or the delta method.

We are interested in analyzing bivariate samples where one or more variables may be censored. We consider a bivariate data set in which all N values of variable $X_1$ are observed but only n$\leq$ N values of variable $X_2$ are observed. We consider left singly censored samples with Type I censoring, where the censoring point T is fixed. More complicated examples include the screened-samples (Cohen, 1955), where an observation on $X_2$ is available whenever $X_1$ is above a screening level. Here, the censoring point on $X_2$ is random and depends on whether $X_1$ is above T.

We use the E-M algorithm to provide maximum likelihood estimates of the parameters of the joint distribution after a bivariate BC transformation to normality. We investigate the selection of the BC transformation that conforms best to the bivariate normal distribution. We obtain the asymptotic covariance matrix of the maximum likelihood estimates and use it to construct bivariate confidence regions for the joint distribution of the means. Small-sample performances of the estimators are investigated by Monte Carlo simulation.

## REFERENCES


Cohen, A. C. (1955). Restriction and Selection In Samples From Bivariate Normal Distribution. *American Statistical Association Journal*, September, 884-893.

Shumway, R. H. Azari, A. S. and Johnson, P. (1989). Estimating Mean Concentration Under Transformation for Environmental Data With Detection Limits. *Technometrics*, August, Vol. 31, No. 3, 347-356.

USEPA (1992). Methods for Evaluating the Attainment of Cleanup Standards, Volume 2: Groundwater. EPA/230-R-92-014. Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency, Washington, D.C.