ESTIMATING CORRELATION WITH MULTIPLY CENSORED DATA

E.A.C. Newton[1], R.A. Rudel[1]

[1]Silent Spring Institute, Newton, USA

Email: *newton@silentspring.org*

Environmental data frequently are left censored due to detection limits of laboratory assay procedures. This presents difficulties in statistical analysis of the data. This paper examines methods for estimating the correlation between variables both of which are censored at multiple levels. Multiple censoring frequently arises due to adjustment of singly censored laboratory results for physical sample size. We discuss maximum likelihood (ML) estimation of ρ and introduce a new method (MLE2) that, instead of using the multiply censored data directly, relies on ML estimates of the covariance of the singly censored laboratory data. We compare the ML methods with several commonly used ad-hoc methods: correlations estimated with non-detects set to DL/2, Kendall's tau-b and correlations estimated using detects only (DET). The methods are compared based on simulations and real data. In the simulations, censoring levels are varied from 0 to 90%, ρ from -0.8 to 0.8 and ν (variance of physical sample size) is set to 0 and 0.5, for a total of 550 parameter combinations with 1000 replications at each combination. We find that all methods except MLE2 are highly biased. With increasing levels of censoring, the ad-hoc methods in general tend toward zero if singly censored and one if multiply censored. MLE2, however, has higher variance than the ad-hoc estimators. Based on RMSE, MLE2 performs the best and DET the worst.