Chapter 13

Methods for Data Below the Reporting Limit

To comply with environmental regulations, an industry must show that the daily mean copper concentration in its wastewater discharge does not exceed the legal standard. Yet for many of the wastewater samples taken at two hour intervals, concentrations are below the analytical reporting limit of the laboratory. These "less-thans" make it impossible to compute a simple mean concentration. When the industry substitutes a zero for each less-than, the standard is not exceeded. When the regulatory agency substitutes a value equal to the reporting limit, the standard is exceeded. Which is correct? Has the law been violated?

Ground-water quality is measured both upgradient and downgradient of a waste-disposal site. Comparison of the two groups of data is performed to determine whether contamination of the ground-water system has occurred. Usually t-tests are employed for this purpose, and yet the t-test requires estimates of means and standard deviations which are impossible to obtain unless numerical values are fabricated to replace any less-thans present in the data. By substituting one number, the two groups appear the same. Substituting a second number causes H_0 to be rejected, and the two groups to be declared different. Which is correct?

As trace substances in the world's soils, air and waters are increasingly investigated, concentrations are more frequently being encountered which are less than limits deemed reliable enough to report as numerical values. These less-than values -- values stated only as "<rl", where rl is called the "reporting limit" or "detection limit" or "limit of quantitation" (Keith et al., 1983) -- present a serious interpretation problem for data analysts. Estimates of summary statistics which best represent the entire distribution of data, both below and above the reporting limit, are necessary to accurately analyze environmental conditions. Also needed are hypothesis test and regression procedures that provide valid conclusions and models for such data. These needs must be met using the only information available to the data analyst: concentrations measured above one or more reporting limits, and the observed frequency of data below those limits.

This chapter discusses the most appropriate statistical procedures given that data have been reported as less-thans. It does not consider the alternative of reporting numerical values for all data, including those below reporting limits -- see ASTM (1983), Porter et al. (1988) and Gilliom et al. (1984) for discussion of this alternative.

13.1 Methods for Estimating Summary Statistics

Methods for estimating summary statistics of data which include less-thans (statisticians call these "censored data") can be divided into the three classes discussed below: simple substitution, distributional, and robust methods. Recent papers have documented the relative performance of these methods. Gilliom and Helsel (1986) and Gleit (1985) compared the abilities of several estimation methods in detail over thousands of simulated data sets. Helsel and Gilliom (1986) then applied these methods to numerous water-quality data sets, including those which are not similar to the assumed distributions required by the distributional methods. A single case study was reported by Newman and Dixon (1990). Helsel and Cohn (1988) dealt with censoring at multiple reporting limits. Large differences were found in these methods' abilities to estimate summary statistics for censored data.

Methods may be compared based on their ability to replicate true population statistics. Departures from true values are measured by the root mean squared error (RMSE), which combines both bias and lack of precision. The RMSE of the estimate of the mean $\overline{x}$ in comparison to the true population value μ is shown in equation 13.1. Similar equations would be used for estimation of other summary statistics.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{(\overline{x}-\mu)^2}{\mu} \Big/ N} \qquad [13.1]$$

Methods whose estimates $\overline{x}$ are closer to the true value μ have lower RMSEs, and are considered better.

13.1.1 Simple Substitution Methods

Simple substitution methods (Figure 13.1) substitute a single value such as one-half the reporting limit for each less-than value. Summary statistics are calculated using both these fabricated numbers along with the values above the reporting limit. These methods are widely used, but have no theoretical basis. As Figure 13.1 shows, the distributions resulting from simple substitution methods have large gaps, and do not appear realistic.

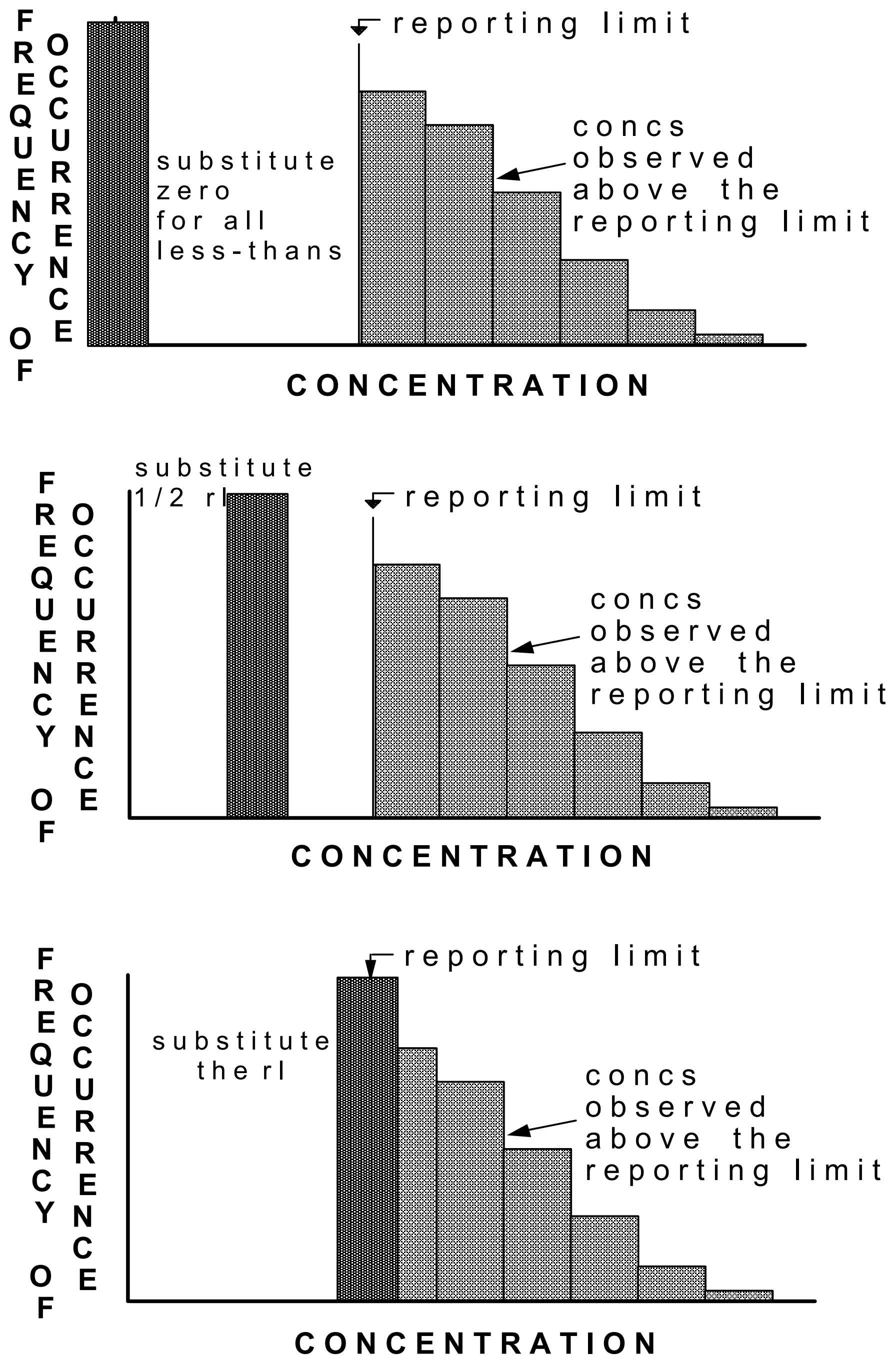


Figure 13.1. Histograms for simple substitution methods.

Studies cited above determined that simple substitution methods performed poorly in comparison to other procedures. Substitution of zero produced estimates of mean and median which were biased low, while substituting the reporting limit resulted in estimates above the true value. Results for the standard deviation and IQR, and for substituting one-half the reporting limit, were also far less desirable than alternative methods. With the advent of powerful desktop computers to perform more complex calculations there appears to be no reason to use simple substitution methods. As the choice of value to be substituted is essentially arbitrary without some knowledge of instrument readings below the reporting limit, and as large differences may occur in the resulting estimates, simple substitution methods are not defensible.

13.1.2 Distributional Methods

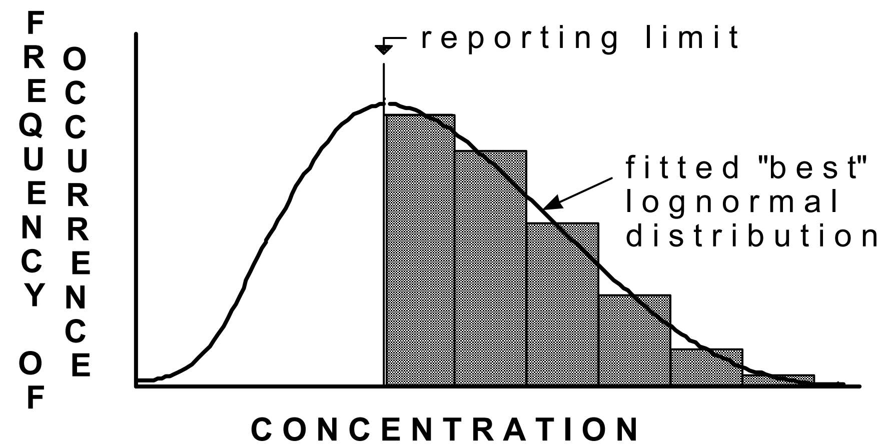
Distributional methods (Figure 13.2) use the characteristics of an assumed distribution to estimate summary statistics. Data both below and above the reporting limit are assumed to follow a distribution such as the lognormal. Given a distribution, estimates of summary statistics are computed which best match the observed concentrations above the reporting limit and the percentage of data below the limit. Estimation methods include maximum-likelihood estimation or MLE (Cohen, 1959), and probability plotting procedures (Travis and Land, 1990). MLE estimates are more precise (lower RMSE) than probability plotting, and both methods are unbiased, when observations fit the assumed distribution exactly and when the sample size is large. However, this is rarely the case in environmental studies. When data do not match the observed distribution, both methods may produce biased and imprecise estimates. Thus the most crucial consideration when using distributional methods is how well the data can be expected to fit the assumed distribution. Even when distributional assumptions are correct, MLEs have been shown to produce estimates with large bias and poor precision for the small sample sizes of n=5, 10, and 15 (Gleit, 1985). MLE methods are commonly used in environmental disciplines such as air quality (Owen and DeRouen, 1980) and geochemistry (Miesch, 1967).

Assuming a lognormal distribution for concentrations, MLEs for larger (n=25, 50) data sets have provided excellent estimates of percentiles (median and IQR) for a variety of data distributions realistic for environmental studies, even those which are not lognormal. However, they have not worked as well for estimating the mean and standard deviation (Gilliom and Helsel, 1986). There are two reasons why this is so.

First, the lognormal distribution is flexible in shape, providing reasonable approximations to data which are nearly symmetric and to some positively-skewed distributions which are not lognormal. Thus the lognormal can mimic the actual shape of the data over much of the distribution, adequately reproducing percentile statistics even though the data were not truly lognormal in shape. However, the moment statistics (mean and standard deviation) are very sensitive to values of the largest observations. Failure of the assumed distribution to fit these observations will result in poor estimates of moments.

Second, there is a transformation bias in lognormal MLE inherent in computing estimates of the mean and standard deviation for any transformation -- including logarithms -- and then retransforming back to original units. Compensating for this bias often requires an assumption about distributional shape. In Chapter 9 transformation bias was discussed in the context of regression. The same phenomena is present for estimates of the mean. Indeed, if no explanatory variables are significant then a regression model simplifies to estimating the mean. Percentiles, however, can be directly transformed between measurement scales without bias.

Maximum Likelihood (MLE) -- fits 'best' lognormal distribution to the data, and then



determines summary statistics of the fitted distribution to represent the data.

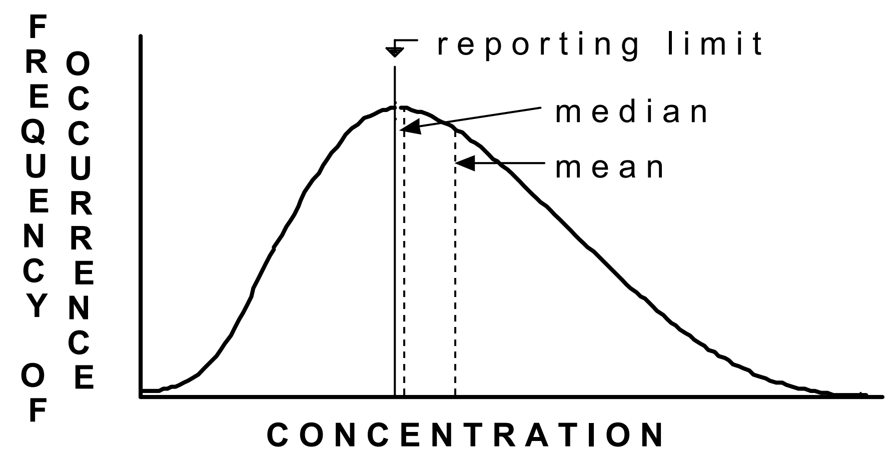


Figure 13.2. Distributional (MLE) method for computing summary statistics.

Two less-frequently used distributional methods are a "fill-in with expected values" MLE technique (Gleit, 1985) and a probability plot method which estimates the mean and standard deviation by the intercept and slope, respectively, of a line fit to data above the reporting limit (Travis and Land, 1990). Probability plot methods are easy to compute with standard statistics software, an advantage for practitioners. Both methods suffer from transformation bias when estimates are computed in one scale and then retransformed back into original units. Thus

Travis and Land (1990) recommended the probability plot for estimating the geometric mean. Its use for estimating the mean in original units would have to take transformation bias into consideration. Both methods should be somewhat less precise than MLEs.

13.1.3 Robust Methods

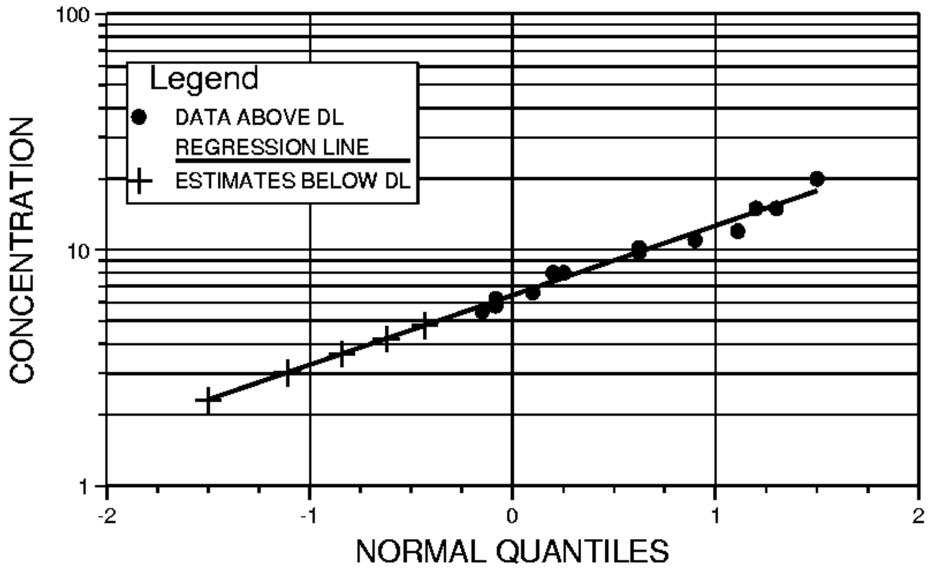
Robust methods (Figure 13.3) combine observed data above the reporting limit with below-limit values extrapolated assuming a distributional shape, in order to compute estimates of summary statistics. A distribution is fit to the data above the reporting limit by either MLE or probability plot procedures, but the fitted distribution is used only to extrapolate a collection of values below the reporting limit. These extrapolated values are <u>not</u> considered as estimates for specific samples, but only used collectively to estimate summary statistics. The robustness of these methods result primarily from their use of observed data rather than a fitted distribution above the reporting limit. They also avoid transformation bias by performing all computations of summary statistics in original units.

Robust methods have produced consistently small errors for all four summary statistics in simulation studies (Gilliom and Helsel, 1986), as well as when applied to actual data (Helsel and Gilliom, 1986). Robust methods have at least two advantages over distributional methods for computation of means and standard deviations. First, they are not as sensitive to the fit of a distribution for the largest observations because actual observed data are used rather than a fitted distribution above the reporting limit. Second, estimates of extrapolated values can be directly retransformed and summary statistics computed in the original units, avoiding transformation bias.

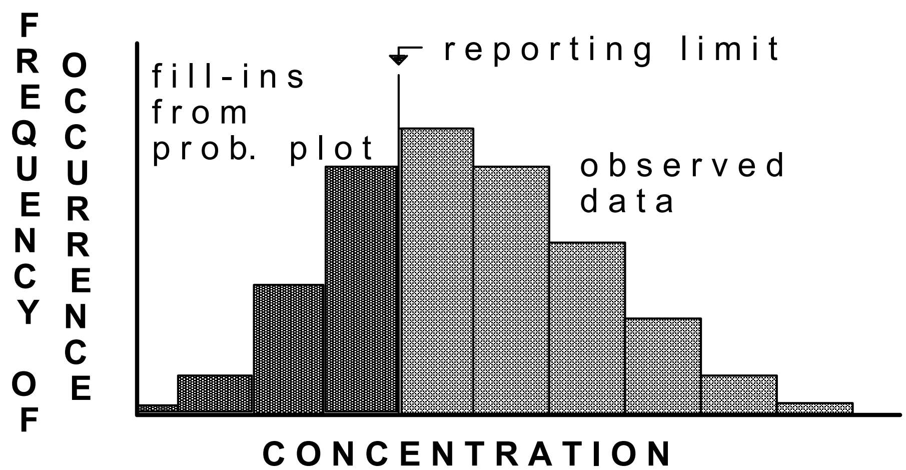
13.1.4 Recommendations

In practice, the distributions of environmental data are rarely if ever known, and may vary between constituents, time periods, and locations. Robust methods should therefore be used to protect against the possibly large errors of distributional methods when estimating the mean and standard deviation. Either robust probability plot or distributional MLE procedures have been shown to perform well for estimating the median and IQR. Use of these methods rather than simple substitution methods for environmental data should substantially lower estimation errors for summary statistics.

As an alternative to estimating percentiles, sample values can sometimes be used. When less than 50% of the data are below the reporting limit, the sample median is known. Similarly, when less than 25% of the data are censored, the sample IQR is known. Some information is available about percentiles when even larger amounts of data lie below the reporting threshold, as shown in the examples below. Unfortunately no similar process is available for sample estimates of mean and standard deviation.



A.



B.

Figure 13.3. Robust (probability plot) method of estimating summary statistics

A) regression of log of concentration versus normal score is used to extrapolate "fill-in" values below the reporting limit.

B) these "fill-ins" are retransformed back to original units, and combined with data above the reporting limit to compute estimates of summary statistics

Example 1

<1 <1 <1 <10 <10 <10 <50.

The mean and std deviation cannot be estimated by any method, as there are no data above the reporting limit. For the median and IQR, a great deal of information is present. To compute a median where all data are below one or more reporting limits and the sample size is odd, remove the < sign, compute the sample median, and then restore the < sign. Therefore the median is <10. The IQR must equal the sample 75th percentile, as the 25th percentile could equal zero. Here the IQR is <10.

Example 2:

<1 <1 <1 <10 <20 <20

When all data are below one or more reporting limits and n is even, again remove the < signs. The larger of the two center observations (the ([n/2]+1)th observation) is used as the median, rather than the average of the two center observations as for uncensored data. Restore the < sign. The median of the above 6 points is <10. The IQR is computed as in example 1, and here would be <20.

Example 3:

<1 <1 <1 5 7 8 12 16 25

For data above and below one reporting limit, the sample median is known to be 7, as less than 50% of the data are censored. Because more than 25% of the data are censored, the sample IQR must be computed as a range. If all the <1's are actually 0, the IQR = 14 – 0 = 14. If all <1's are very close to 1, the IQR = 13. So the sample IQR could be reported as "13 to 14" if that were of sufficient precision. Otherwise, the probability plot and maximum likelihood methods must be used to estimate the moment and percentile statistics.

13.1.5 Multiple Reporting Limits

Data sets may contain values censored at more than one reporting limit. This commonly occurs as limits are lowered over time at a single lab, or when data having different reporting limits are combined from multiple laboratories. Estimation methods belonging to the above three classes are available for this situation. A comparison of these methods (Helsel and Cohn, 1988) concluded that robust methods again provide the best estimates of mean and standard deviation, and MLEs for percentiles. For example, in Figure 13.4 the error rates for six estimation methods are compared to the error that would occur had all data been above the reporting limit (shown as the 100% line). Figure 13.5 shows the same information when the data differ markedly from a lognormal distribution. The simple substitution methods (ZE, HA and DL: substitution of zero, one-half and one times the reporting limit, respectively) have more error in most cases than does the robust probability plot method MR. Where the substitution methods have lower RMSE, it is an artifact of constant, strongly biased estimates, also not a desirable result. The maximum likelihood procedure MM and the MLE adjusted for transformation bias AM show themselves to be excellent estimation methods for percentiles, but suffer from large errors when estimating the mean and standard deviation.

In summary, use of MLE for estimation of percentiles, and the robust probability plot method for estimating the mean and standard deviation, should greatly decrease errors as compared to simple substitution methods for data with multiple reporting limits.

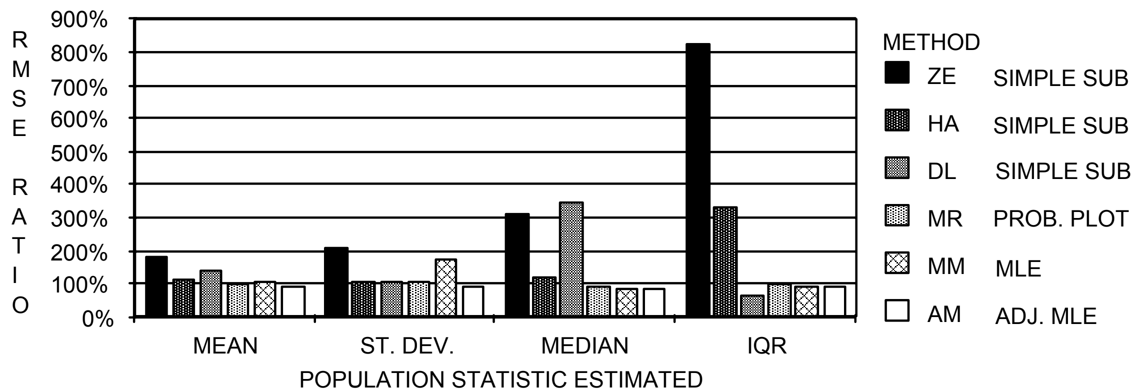


Figure 13.4. Error rates (RMSE -- root mean square error) of six multiple-detection methods divided by error rates for uncensored data estimates, in percent, for data similar to a lognormal distribution (from Helsel and Cohn, 1988)

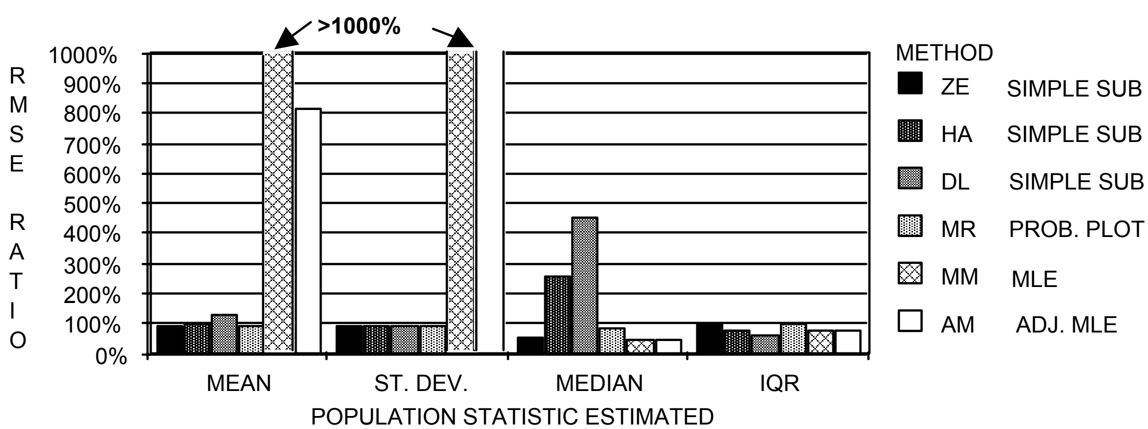


Figure 13.5. Error rates (RMSE -- root mean square error) of six multiple-detection methods divided by error rates for uncensored data estimates, in percent, for data <u>not</u> similar to a lognormal distribution. (from Helsel and Cohn, 1988)

Example 4:

For data above and below multiple reporting limits, such as

<1 <1 <1 5 7 8 <10 <10 <10 12 16 25

it is unclear whether the <10's are below or above 1, or 5, etc. Therefore ordering the data from smallest to largest is impossible. Instead, the probability plot method is used to compute the mean and standard deviation, and the maximum likelihood method for the median and IQR (Helsel and Cohn, 1988). These give the following:

mean	=	7.8	median	=	2.8
std dev	=	6.9	IQR	=	7.5.

13.2 Methods for Hypothesis Testing

Methods for hypothesis testing of censored data can also be classified into the three types of procedures: simple substitution, distributional or parametric, and robust or nonparametric. The advantages and disadvantages of each are summarized below.

13.2.1 Simple Substitution Methods

When censoring is present, values are often fabricated in order to perform parametric tests such as t-tests. Problems caused by such substitution methods are illustrated below. Investigators have also deleted censored data prior to hypothesis testing. This latter approach is the worst procedure, as it causes a large and variable bias in the parameter estimates for each group. After deletion, comparisons made are between the upper X% of one group versus the upper Y% of another, where X and Y may be very different. Such tests have little or no meaning.

Example 5

As an example of hypothesis test methods for censored data, tests will be performed to determine whether or not means or medians significantly differ between two groups. Two data sets were generated from lognormal distributions having the same variance, but with differing mean values. Sample statistics for the two data sets before and after censoring are given in Table 13.1. Prior to any censoring, group means were found to be different by a t-test ($p=0.04$, Table 13.2). The data were then censored at a reporting limit of 1 μg/L, so that all data below 1.0 were recorded as <1. This produced 14 less-than values (70%) in group A, and 5 less-than values (23%) in group B.

The simple substitution method for comparing two groups of censored data is to fabricate data for all less-than values, and include these "data" with detected observations when performing a t-test. No *a priori* arguments for fabrication of any particular value between 0 and the reporting limit can be made. Substituting zero for all less-than values, the means are declared significantly different ($p = 0.01$). Yet when the reporting limit of 1.0 is substituted, the means are not found to be different ($p = 0.19$). The conclusion is thus strongly dependent on the value substituted!

Fabrication of data followed by a t-test is an arbitrary process leading to ambiguous conclusions. It should be avoided.

13.2.2 Distributional Test Procedures

Parametric tests are also available which do not require substitutions for less-thans. Instead, maximum likelihood methods are used to solve the relevant equations. Where the distributional assumptions are appropriate, these relatively unknown tests have great utility.

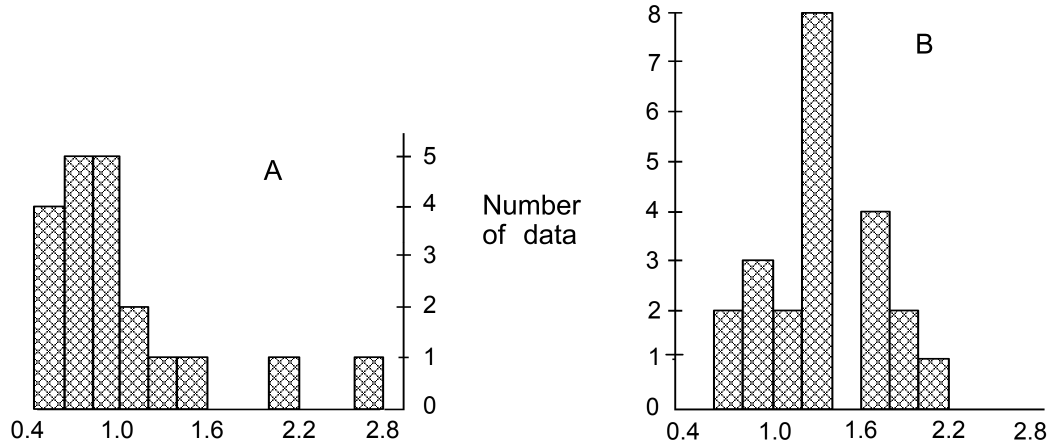
The distributional method for a t-test situation is performed using a regression procedure for censored data known as tobit regression (Judge et al, 1985). Tobit regression uses both the data values above the reporting limit, and the proportion of data below the reporting limit, to compute a slope coefficient by maximum likelihood. For a two-group test, the explanatory variable in the regression equation is the binary variable of group number, so that data in one group have a value of 0, and in the other group a value of 1. The regression slope then equals the difference between the two group means, and the t-test for whether this slope differs from zero is also a test of whether the group means differ. Tobit regression is discussed further in section 13.3. One advantage to Tobit regression for hypothesis testing is that multiple reporting limits may easily be incorporated. The caution for its use is that proper application does require the data in both groups to be normally distributed around their group mean, and for the variance in each group to be equal. For large amounts of censoring these restrictions are difficult to verify.

13.2.3 Nonparametric Tests

With nonparametric tests, no fabrication of data values is required. All censored data are represented by ranks which are tied at values lower than the lowest number above the reporting limit. The rank-sum test compares the medians of two independent data groups (Chapter 5).

Prior to censoring, a rank-sum test on the example 5 data produced a much lower p-value (p=0.003) than did the t-test (Table 13.2). This lower p-value is consistent with the proven greater power of the nonparametric test to detect differences between groups of skewed data, as compared to the t-test. To compute the rank-sum test on censored data, the 19 less-than values are considered tied at the lowest value, with each assigned a rank of 10 (the mean of ranks 1-19). The next highest value, the data point just above the reporting limit, obtains a rank of 20. All data above the reporting limit will have ranks identical to those which would have been obtained had no censoring been present. The resulting p-value is 0.002, essentially the same as for the uncensored data, and the two groups are easily declared different. Thus in this example the nonparametric method makes very efficient use of the information contained in the less-than values, avoids arbitrary assignment of fabricated values, and accurately represents the lack of knowledge below the reporting limit. Results do not depend on any distributional assumption.

TABLE 13.1--Characteristics of Two Lognormal Data Groups (A and B)



A		B
20	no. observations	22
1.00	mean	1.32
0.57	std. deviation	0.39
0.65	25th percentile	1.07
0.85	median	1.25
1.07	75th percentile	1.66
14	no. <rl	5

TABLE 13.2-- Significance Tests Between Groups A and B

Hypothesis test used	test statistic	p
Uncensored data		
t-test (Satterthwaite approx.)	−2.13	0.04
regression with binary variable	−2.17	0.04
rank-sum test		
	−2.92	0.003
After imposing artificial reporting limit		
t-test		
less-thans = 0.0	−2.68	0.01
less-thans = 0.5	−2.28	0.03
less-thans = 1.0	−1.34	0.19
tobit regression with binary variable	−2.28	0.03
rank-sum test	−3.07	0.002

When severe censoring (near 50% or more) occurs, all of the above tests will have little power to detect differences in central values. The investigator will be inhibited in stating conclusions about the relative magnitudes of central values, and other characteristics must be compared. For example, contingency tables (Chapter 14) can test for a difference in the proportion of data above the reporting limit in each group. The test can be used when the data are reported only as detected or not detected. It may also be used when response data can be categorized into three or more groups, such as: below detection, detected but below some health standard, and exceeding standards. The test determines whether the proportion of data falling into each response category differs as a function of different explanatory groups, such as different sites, land use categories, etc.

13.2.4 Hypothesis Testing With Multiple Reporting Limits

More than one reporting limit is often present in environmental data. When this occurs, hypothesis tests such as comparisons between data groups are greatly complicated. It can be safely said that fabrication of data followed by computation of t-tests or similar parametric procedures is at least as arbitrary with multiple reporting limits as with one reporting limit, and should be avoided. The deletion of data below all reporting limits prior to testing should also be completely avoided.

Tobit regression can be utilized with multiple reporting limits. Data should have a normal distribution around all group means and equal group variances to use the test. These assumptions are difficult to verify with censored data, especially for small data sets.

One robust method which can always be performed is to censor all data at the highest reporting limit, and then perform the appropriate nonparametric test. Thus the data set

<1 <1 <1 5 7 8 <10 <10 <10 12 16 25

would become <10 <10 <10 <10 <10 <10 <10 <10 <10 12 16 25.

and a rank-sum test performed to compare this with another data set. Clearly this produces a loss of information which may be severe enough to obscure actual differences between groups (a loss of power). However, for some situations this is the best that can be done.

Alternatively, nonparametric score tests common in the medical "survival analysis" literature can sometimes be applied to the case of multiple reporting limits (Millard and Deverel, 1988). These tests modify uncensored rank test statistics to compare groups of data. The modifications allow for the presence of multiple reporting limits. In the most comprehensive review of these score tests, Latta (1981) found most of them to be inappropriate for the case of unequal sample sizes. The Peto-Prentice test with asymptotic variance estimate was found by Latta (1981) to be the least sensitive to unequal sample sizes. Another crucial assumption of score tests is that the censoring mechanism must be independent of the effect under investigation (see box). Unfortunately, this is often not the case with environmental data.

Examples when a score test would be inappropriate.

Score tests are inappropriate when the censoring mechanism differs for the two groups. That is, the probability of obtaining a value below a given reporting limit differs for the two groups when the null hypothesis that the groups are identical is true.

1. Suppose a trend over time was being investigated. The first five years of data were produced with a method having a reporting limit of 10 mg/L; the second five years used an improved method with 1 mg/L as its reporting limit. A score test of the first half of the data versus the second would not be valid, as the censoring mechanism itself varied as a direct function of time.
2. Two groups of data are compared as in a rank-sum test, but most of the data from group A were measured with a chemical method having 1 as its reporting limit, while most of group B were measured with a method having 10 as its reporting limit. A score test would not yield valid results, as the censoring mechanism varies as a function of what is being investigated (the two groups).

Examples when a score test would be appropriate.

A score test yields valid results when the change in censoring mechanism is not related to the effect being measured. Stated another way, the probability of obtaining data below each reporting limit is the same for all groups, assuming the null hypothesis of no trend or no difference is true. Here a score test provides much greater power than artificially censoring all data below the highest reporting limit before using the rank-sum test.

1. Comparisons were made between two groups of data collected at roughly the same times, and analyzed by the same methods, even though those methods and reporting limits changed over time. Score tests are valid here.

2. Differing reporting limits resulted from analyses at different laboratories, but the labs were assigned at random to each sample. Censoring is thus not a function of what is being tested, but is a random effect, and score tests would be valid.

13.2.5 Recommendations

Robust hypothesis tests have several advantages over their distributional counterparts when applied to censored data. These advantages include: (1) the ability to disregard whether data adhere to a normal distribution. Verifying normality is difficult to do with censored data; (2) greater power for the skewed distributions common to environmental data; and (3) data below the reporting limit are incorporated without fabrication of values or bias. Information contained in less-than values is accurately used, not misrepresenting the state of that information.

Tests incorporating multiple reporting limits are more problematic, and should be an area of future research. When adherence to a normal distribution can be documented, Tobit regression offers the ability to incorporate multiple reporting limits in a distributional test regardless of a change in censoring mechanism. Nonparametric score tests require consistency in censoring mechanism with respect to the effect being tested.

13.3 Methods For Regression With Censored Data

With censored data the use of ordinary least squares (OLS) for regression is prohibited. Coefficients for slopes and intercept cannot be computed without values for the censored observations, and substituting fabricated values may produce coefficients strongly dependent on the values substituted. Four alternative methods capable of incorporating censored observations are described below.

The choice of method depends on the amount of censoring present, as well as on the purpose of the analysis. For small amounts of censoring (below 20%), either Kendall's line or the tobit line may be used. Kendall's would be preferred if the residuals were not normally distributed, or when outliers are present. For moderate censoring (20-50%), Tobit or logistic regression must be used. With large amounts of censoring, inferences about concentrations themselves must be abandoned, and logistic regression employed. When both the explanatory and response variables are censored, tobit regression is applicable for small amounts of censoring. For larger amounts of censoring, contingency tables or rank correlation coefficients are the only option.

13.3.1 Kendall's Robust Line Fit

When one censoring level is present, Kendall's rank-based procedure for fitting a straight line to data can test the significance of the relationship between a response and explanatory variable (Chapter 10). An equation for the line, including an estimate of the slope, is usually also desirable. This can be computed when the amount of censoring is small. Kendall's estimate of slope is the median of all possible pairwise slopes of the data. To compute the slope with censoring, twice compute the median of all possible slopes, once with zero substituted for all less-thans, and once with the reporting limit substituted. For small amounts of censoring the resulting slope will change very little, or not at all, and can be reported as a range if necessary. If the slope value change is of an unacceptable magnitude, tobit or logistic regression must be performed.

Research is underway on methods based on scores similar to those for hypothesis tests with multiply-censored data that may allow robust regression fits to data with multiple reporting limits (McKean and Sievers, 1989).

13.3.2 Tobit Regression

Censored response data can be incorporated along with uncensored observations into a procedure called tobit regression (Judge et al., 1985). It is similar to OLS except that the coefficients are fit by maximum-likelihood estimation. MLE estimates of slope and intercept are based on the assumption that the residuals are normally distributed around the tobit line, with constant variance across the range of predicted values. Again, it is difficult to check these assumptions with censored data. Should the data include outliers, these can have a strong influence on the location of the line and on significance tests (Figure 13.6), as is true with uncensored OLS. Verification of linearity and constant variance assumptions should be attempted when only small amounts of data are censored using residuals plots. Residuals for uncensored observations would be plotted versus predicted values. For larger percentages of less-thans, decisions whether to transform the response variable must often be made based on previous knowledge ("metals always need to be log-transformed", etc.). Tobit regression is also applicable when both the response and explanatory variables are censored, such as a regression relationship between two chemical constituents. However, the amount of censoring must be sufficiently small that the linearity, constant variance, and normality assumptions of the procedure can be checked. Finally, Cohn (1988) as well as others have proven that the tobit estimates are slightly biased, and have derived bias corrections for the method.

13.3.3 Logistic Regression

Here the response variable is categorical. No longer is a concentration being predicted, but a probability of being in discrete binary categories such as above or below the reporting limit. One response (above, for example) is assigned a value of 1, and the second response a 0. The probability of being in one category versus the second is tested to see if it differs as a function of continuous explanatory variable(s). Examples include predicting the probability of detecting

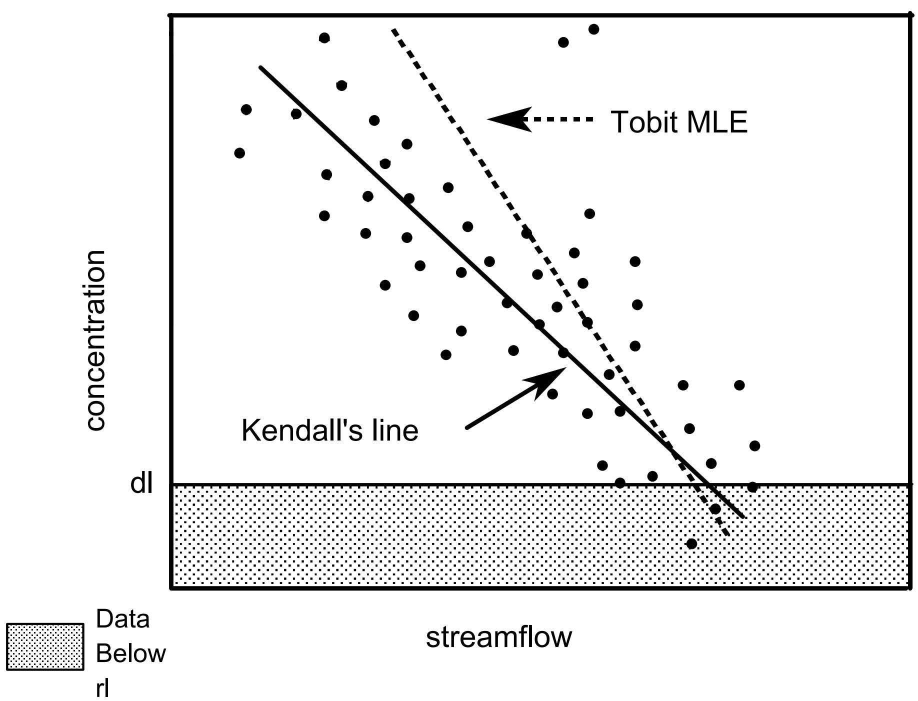


Figure 13.6. Kendall's and tobit MLE lines for censored data with outliers. Note the tobit line is strongly influenced by outliers.

concentrations of some organic contaminant from continuous variables such as nitrate concentrations, population density, percent of some appropriate land use variable, or of irrigation intensity. Predictions from this regression-type relationship will fall between 0 and 1, and are interpreted as the probability [p] of observing a response of 1. Therefore [1−p] is the probability of a 0 response.

Logistic regression may be used to predict the probabilities of more than 2 response categories. When there are m>2 ordinal (may be placed in an order) responses possible, (m−1) equations must be derived from the data. For example, if 3 responses are possible (concentrations below rl =0, above rl but below health standards =1, and above health standards =2), two logistic regressions must be computed. First, an equation must be written for the probability of being nonzero (the probability of being above the rl). Next the probability of a 2 (probability of being above the health standard) is also modelled. Together, these two equations completely define the three probabilities p(y=0), p(y=1), and p(y=2) as a function of the explanatory variables. See Chapter 15 for more detail.

13.3.4 Contingency Tables

Contingency tables are useful in the regression context if both explanatory and response variables contain censoring. For example, suppose the relationship between two trace metals in soils (such as arsenic and aluminum) is to be described. The worst procedure would again be to throw away the data below the reporting limits, and perform a regression. Figure 13.7 shows that a true linear relationship with negative slope could be completely obscured if censored data were ignored, and only data in the upper right quadrant investigated. Contingency tables provide a measure of the strength of the relationship between censored variables -- the phi statistic ϕ (Chapter 14), a type of correlation coefficient. An equation which describes this relationship, as per regression, is not available. Instead, the probability of y being in one category can be stated as a function of the category of x. For the Figure 13.7 data, the probability of arsenic being above the reporting limit is 21/36 = 0.58 when aluminum is above reporting limit, and 17/18 = 0.94 when aluminum is below the reporting limit.

13.3.5 Rank Correlation Coefficients

The robust correlation coefficients Kendall's tau or Spearman's rho (Chapter 8) could also be computed when both variables are censored. All values below the reporting limit for a single variable are assigned tied ranks. Rank correlations do not provide estimates of the probability of exceeding the reporting limit as does a contingency table. So they are not applicable in a regression context, but would be more applicable than contingency tables in a correlation context. One such context would be in "chemometrics" (Breen and Robinson, 1985): the computation of correlation coefficients for censored data as inputs to a principal components or factor analysis.

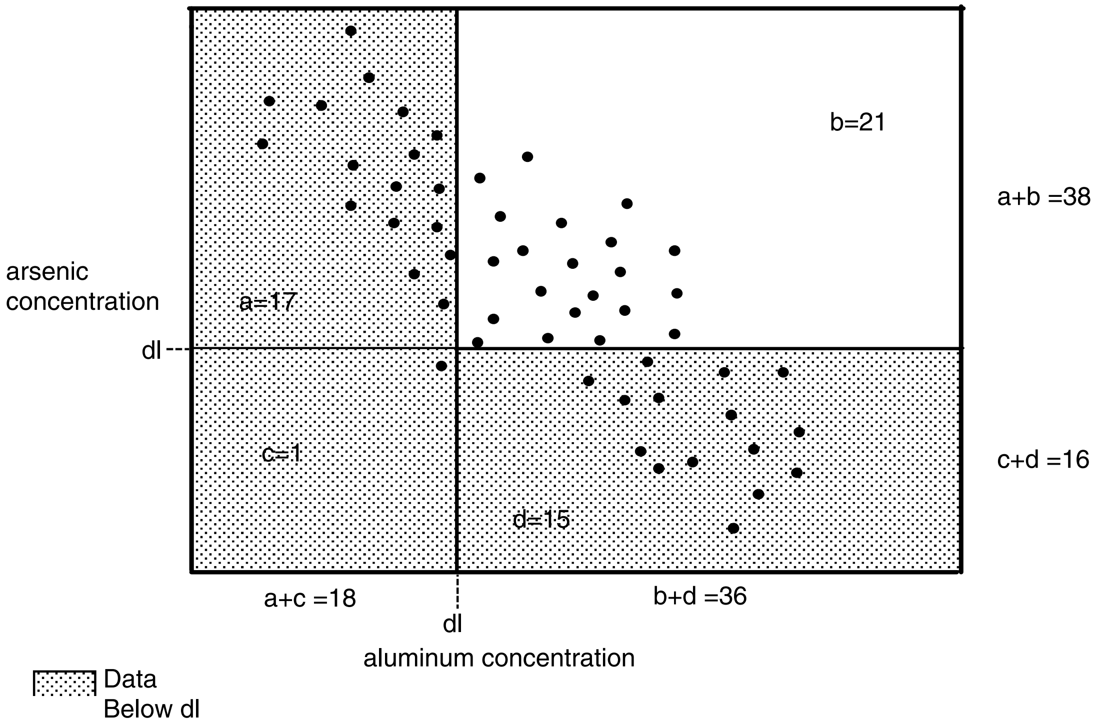


Figure 13.7. Contingency table relationship between two censored variables. (Ignoring censored data would produce the misleading conclusion that no relationship exists between the two variables)

13.3.6 Recommendations

Relationships between variables having data below reporting limits can be investigated in a manner similar to regression. Values should not be fabricated for less-thans prior to regression. Instead, Table 13.3 summarizes alternative methods appropriate for censored data. For small amounts of censoring and one reporting limit, Kendall's robust line can be fit to the data. For moderate censoring and/or multiple reporting limits, tobit regression can be performed. For more severe censoring of the dependent variable, logistic regression is appropriate. When both response and explanatory variables contain severe censoring, contingency tables and rank correlation coefficients can be performed.

Estimation of Summary Statistics		
Mean and Standard Deviation		Percentiles
Robust Probability Plot		Robust Probability Plot or MLE
	Hypothesis Tests	
	One Reporting Limit	Several Reporting Limits
Compare 2 groups:	rank-sum test	tobit regression
Compare >2 groups:	Kruskal-Wallis test	tobit regression
Severe Censoring (>50%):	above tests, or contingency tables	–
	Regression	
Small % censoring	Moderate % censoring	Large % censoring
Kendall's robust line	tobit regression	logistic regression
tobit regression	logistic regression	contingency tables

Table 13.3 Recommended Techniques for Interpretation of Censored Data

Exercises

13.1 Below are concentrations of triphenyltin (TPT) measured in a sediment core by Fent and Hunn (1991). Estimate the mean, standard deviaion, median and interquartile range for these data.

Concentrations of TPT, in □ g/kg dry weight

51	29	71	69	34	56	83	<2	<2	107	35	<2
26	4	10	<2	2	<2						

13.2 Below are depths (bottom of segment) for the 18 TPT concentrations of exercise 13.1. Compute the significance of the relationship between concentration and depth for this core.

Depth (cm) of bottom of sediment core

0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
7	8	9	10	11	12						

13.3 Silver concentrations in standard solutions were reported by several laboratories in an inter-lab comparison [Janzer, 1986]. The 56 analyses included 36 values below one of 12 detection limits. One large outlier (a "far outside" value on a boxplot) of 560 µg/L was also reported. Compute the mean, standard deviation, median and interquartile range for the data presented below:

0.8	<25	<5	<0.2	<0.5	5.0	<0.3	<0.2	0.1
2.7	<0.1	<20	1.42.0	<2.5	2.0	2.0	<1	
<10	<1	<0.2	1.0<10	<0.2	0.2	1.2	<1	
1.0	<6	<1	0.7<1	<53.2	2.0	10.0		
1.0	4.4	<1	<1<1	<20	<5	<10	<10	
90	1.5	<1	<2<10	560<5	0.1	<20		
<1	<0.1							