Methods of Dealing with Values Below the Limit of Detection using SAS

Carry W. Croghan, US-EPA, Research Triangle Park, NC, and Peter P. Egeghy US-EPA, Las Vegas, NV

ABSTRACT

Due to limitations of chemical analysis procedures, small concentrations cannot be precisely measured. These concentrations are said to be below the limit of detection (LOD). In statistical analyses, these values are often censored and substituted with a constant value, such as half the LOD, the LOD divided by the square root of 2, or zero. These methods for handling below-detection values results in two distributions, a uniform distribution for those values below the LOD, and the true distribution. As a result, this can produce questionable descriptive statistics depending upon the percentage of values below the LOD. An alternative method uses the characteristics of the distribution of the values above the LOD to estimate the values below the LOD. This can be done with an extrapolation technique or maximum likelihood estimation. An example program using the same data is presented calculating the mean, standard deviation, t-test, and relative difference in the means for various methods and compares the results. The extrapolation and maximum likelihood estimate techniques have smaller error rates than all the standard replacement techniques. Although more computational, these methods produce more reliable descriptive statistics.

INTRODUCTION

A common problem faced by researchers analyzing data in the environmental and occupational health fields is how to appropriately handle observations reported to have nondetectable levels of a contaminant (Hornung and Reed, 1990). Even with technical advances in the areas of field sampling, sample processing protocols, and laboratory instrumentation, the desire to identify and measure contaminants down to extraordinarily low levels means that there will always be a threshold below which a value cannot be accurately quantified (Lambert et al., 1991). As a value approaches zero there is a point where any signal due to the contaminant cannot be differentiated from the background noise, i.e. the precision of the instrument/method is not sufficient to discern the presence of the compound and quantify it to an acceptable level of accuracy. While one can differentiate between an "instrument limit of detection" and "method limit of detection," it is the method limit of detection that is relevant to this discussion.

By strict definition, the limit of detection (LOD) is the level at which a measurement has a 95% probability of being different than zero (Taylor, 1987). There is no universal procedure for determining a method LOD, but it is broadly defined as the concentration corresponding to the mean blank response (that is, the mean response produced by blank samples) plus three standard deviations of the blank response. Various identification criteria (for example, retention time on a specific analytical column, proper ratio of ions measured by a mass spectrometer, etc.) must also be met to confirm the presence of the compound of interest. While it is almost always possible to calculate values below the method LOD, these values are not considered to be statistically different from zero, and are generally reported as "below detection limit" or "nondetect."

Although statistically different from zero, values near the LOD are generally less accurate and precise (that is, less reliable) than values that are much larger than the LOD. The limit of quantitation (LOQ) is the term often used by laboratories to indicate the smallest amount that they consider to be reliably quantifiable. There is often considerable confusion concerning the distinction between the two, but the LOQ is generally defined as some multiple of the LOD, usually three times the LOD. Estimates of concentrations above the LOD but below the LOQ are potentially problematic due to their higher relative uncertainty (compared to values above the LOQ); these values are rarely treated any differently from values above the LOQ.

Unfortunately, it is not as easy to use below-LOD values in data analysis compared to below-LOQ values. First, there is the standard laboratory practice of reporting "nondetect" instead of numbers for values below the LOD (often referred to as "censoring"). Second, even if numbers were reported, they may be highly unreliable (though many researchers might prefer highly unreliable over nothing). Thus, a researcher faced with data containing nondetects must decide how to appropriately use them in statistical analyses. One might reason that since the actual values of these observations are extraordinarily small, they are unimportant. However, these observations can still have a large effect on the parameters of the distribution of the entire set of observations. Treating them incorrectly may introduce severe bias when estimating the mean and variance of the distribution (Lyles et al., 2001), which may consequently distort regression coefficients and their standard errors and reduce power in hypothesis tests (Hughes, 2000). Depending upon the severity of the censoring, the distributional statistics may be highly questionable.

The problem of estimating the parameters of distributions with censored data has been extensively studied (Cohen, 1959; Gleit, 1985; Helsel and Guilliom, 1986; Helsel, 1990; Hornung and Reed, 1990; Lambert, 1991; Özkaynak et al., 1991; Finkelstein and Verma, 2001). The main approaches to handling left censored data are: 1) simple replacement, 2) extrapolation, and 3) maximum likelihood estimation. The most common, and easiest, strategy is simple replacement, where censored values are replaced with zero, with some fraction of the detection limit (usually either 1/2 or $1/\sqrt{2}$), or with the detection limit itself. Extrapolation strategies use regression or probability plotting techniques to calculate the mean and standard deviation based on the regression line of the observed, above-LOD values and their rank scores. A third strategy, maximum likelihood estimation (MLE), is also based on the statistical properties of the noncensored portion of the data, but uses an iterative process to solve for the means and variance. Although the latter two methods do not generate replacement data for the censored observations, they can be modified to do so if needed.

The SAS language allows for both the statistical analyses and data manipulations required to perform these techniques. Using SAS procedures and options for the statistical computations will reduce the necessary coding that other languages would require. The flexibility in outputting the information generated by the procedures, allows for the additional manipulations to be easily made within a datastep. The macro capability of SAS is pivotal for performing the testing and comparison of the different methods of handling censored data. Macros make repetitive processing of procedures and datasteps possible and allow for inputting parameters. The flexibility and robustness of the SAS language makes it a good choice for programming these types of data problems.

METHOD:

To test the difference between the various techniques, distributions with known parameters are needed. Therefore, normally distributed datasets were generated with known means and standard deviations using the RANNOR function in SAS with numbers of observations of either 500 or 1000. The mean and standard deviation combinations were arbitrarily chosen to be 4 and 1.5, 10 and 5, and 50 and 5.

Cut-off points were calculated using the 5, 10, 25, 50 and 75 percentiles for the distribution determined using a PROC MEANS statement. The cut-off points will be used as a surrogate for limit of detection. These datasets were used to test each of the methods of dealing with the censored values.

To test the simple replacement techniques, a series of datasets were generated with the values below the cut-off points replaced using the standard replacement values of zero, cut-off value divided by two, or the cut-off value divided by the square root of two. The means for the new datasets were compared to those of the parent dataset using a student t-test. As was done in Newman et al. 1989, the significance level was set to 0.1. The percentage of times that the p-value was less than 0.1 was calculated for each cut-off level, the replacement technique, and the combination of cut-off level and replacement technique.

The method of extrapolation is a method of calculating the mean and standard deviation based on the regression line of the values and the inverse cumulative normal distribution for a function of the rank score of those values not censored. To perform this method, first the values must be rank ordered including those that are censored. Given there are n samples of which k are censored, for the n-k values that are not censored, calculate the Blom estimate of the normal score:

$$A_i = F^{-1}((i-3/8)/ (n+1/4)),$$

where

F^{-1} is the inverse cumulative normal distribution function
i is the rank score for the value X_i
n is the size of the sample.

In SAS this translates to

```
proc rank data=one out=two normal=blom;
     var X;
     ranks A;
run;
```

Then regress the X values against A using a simple linear regression model:

$$X_i = a + b A_i + e_i.$$

In SAS use the following code:

```
proc reg data=two outest=stat2;
     model X = A;
run;
```

The mean of the noncensored distribution is estimated by intercept and the standard deviation is estimated by the slope of the line.

The maximum likelihood estimator (MLE) of the mean and standard deviation are solved in an iterative process solving the following system of equations:

$$\mu = m - \sigma (k/n-k) f(\varepsilon)/F(\varepsilon)$$
$$\sigma^2 = [S^2 + (m-\mu^2)] / [1 + \varepsilon (k/n-k) f(\varepsilon)/F(\varepsilon)]$$

where:

$\varepsilon = (LOD-\mu)/ \sigma$
LOD= limit of detection
f(k) = the distribution function for the normal distribution
F(k) =the cumulative distribution function for the normal distribution,
m= the mean of all values above the LOD
S =The population standard deviation of all values above the LOD.

In SAS this is done by using LIFEREG

```
proc lifereg data=ttest1 outest=stat3;
      model (lower, Y) = / d=normal;
run;

where:
if X > cuts then lower = X;
     else lower = .;
Y = (X > cuts) * X + (X<=cuts)*cuts;

(Y is X is X is greater than the cuts,
otherwise Y is set to the cuts value)
```

The methods are compared using two different error measurements. The first is based on the percentage of times the derived mean is significantly different from the true mean using a student t-test at the 0.1 level. The second is the based on the relative difference in the derived and true mean. The mean relative difference is calculated over all the distribution parameters.

FINDINGS

If only a small percentage of the values have been censored, there is little bias introduced by any of the replacement techniques (Table 1). At the 5th percentile cut-off, the derived mean for these situations is never significantly different from the true mean for any of the replacement techniques. At the 10th percentile, some significant differences are observed. The largest error rate (8%) is for replacement with zero for the combination of the following parameters, N=1000, mean=4, and standard deviation=1.5.

Conversely, if the percentage of censored values is large, none of these replacement techniques are adequate. The percent of times that the true mean is significantly different from the derived mean ranges from 33.87 to 100% over all the methods when the 75th percentile is used as the cut-off. Even at the 50th percentile, the error rate is high for some combinations of sample size, mean and standard deviation. For N=1000, mean=50, and standard deviation =5 the error rate is 100%, indicating that all of the derived means are significantly different from the true mean.

A difference in the replacements techniques is also obvious at the 25th percentile cut-off level. Replacement with 0 has the highest error rate of 33.33%; replacement with the limit of detection divided by the square root of 2 has an error rate of 0. The overall error rate is 15%.

Table 1: Percentage of Student t-test p-values less than 0.10 for the replacement methods

N=500, mean=4, std=1.5

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
25	100.00	0.40	0.00	33.47
50	100.00	100.00	0.00	66.67
75	100.00	100.00	40.80	80.27
overall	60.00	40.08	8.16	36.08

N=1000, mean=4, std=1.5

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	7.60	0.00	0.00	2.53
25	100.00	69.20	0.00	56.40
50	100.00	100.00	0.00	66.67
75	100.00	100.00	90.00	96.67
overall	61.52	53.84	18.00	44.45

N=500, mean=10, std=5

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00
50	20.40	0.00	0.00	6.80
75	100.00	0.00	1.60	33.87
overall	24.08	0.00	0.32	8.13

N=1000, mean=10, std=5

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00
50	100.00	0.00	0.00	33.33
75	100.00	0.00	94.80	64.93
overall	40.00	0.00	18.96	19.65

N=500, mean=50, std=5

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00
50	0.00	100.00	99.60	66.53
75	100.00	100.00	100.00	100.00
overall	20.00	40.00	39.92	33.31

N=1000, mean=50, std=5

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00
50	100.00	100.00	100.00	100.00
75	100.00	100.00	100.00	100.00
overall	20.00	40.00	39.92	33.30

Overall Sample Types.

	Replacement Value			
Level	ZERO	LOD/2	LOD/√2	overall
05	0.00	0.00	0.00	0.00
10	1.27	0.00	0.00	0.42
25	33.33	11.60	0.00	14.98
50	70.07	66.67	33.27	56.67
75	100.00	66.67	71.20	79.29
overall	40.93	28.99	20.90	30.27

There is a difference in the overall error rate for the different methods of replacement. Replacement with the limit of detection divided by the square root of two has the smallest overall error rate (21%); while the replacement with zero has almost twice the overall error rate (41%).

The error rate is affected by the choice of mean and standard deviation. The mean of 10 and standard deviation of 5 produces the smallest overall error rate, 14%. Further investigation into the relationship between distribution parameters and the error rate is needed to assess the cause of this effect and how to interpret it.

The extrapolation method (Table 2) is a marked improvement over the simple replacement techniques. The largest error rate is 18% at the 75th percentile cut-off level. This method is not as sensitive to the mean and standard deviation of the original distribution. The overall error rate of 1.24% is markedly different from the rate of 20.90% from the best of the replacement techniques, LOD/√2.

Table 2: Percentage of Student t-test p-values less than 0.10 for the Extrapolation method

Distributions parameters	Cut-off Level 05	10	25	50	75	overall
N=500, mean=4, std=1.5	0.0	0.0	0.0	0.8	17.6	3.68
N=1000, mean=4, std=1.5	0.0	0.0	0.0	0.4	18.0	3.68
N=500, mean=10, std=5	0.0	0.0	0.0	0.0	0.0	0.0
N=1000, mean=10, std=5	0.0	0.0	0.0	0.0	0.0	0.0
N=500, mean=50, std=5	0.0	0.0	0.0	0.0	0.0	0.0
N=1000, mean=50, std=5	0.0	0.0	0.0	0.0	0.0	0.0
overall	0.0	0.0	0.0	0.2	6.0	1.24

The method of maximum likelihood estimation (Table 3) has the best overall error rate (0.7%). Only at the 75th percentile cut-off level does the error rate differ from zero. There is, however, an effect by the choice of mean and standard deviation on the error rate, as with the replacement techniques. The highest error rate observed is 11%.

Table 3: Percentage of Student t-test p-values less than 0.10 for the Maximum Likelihood Estimate

Distributions parameters	Cut-off Level 05	10	25	50	75	overall
N=500, mean=4, std=1.5	0.0	0.0	0.0	0.0	10.8	2.2
N=1000, mean=4, std=1.5	0.0	0.0	0.0	0.0	10.8	2.2
N=500, mean=10, std=5	0.0	0.0	0.0	0.0	0.0	0.0
N=1000, mean=10, std=5	0.0	0.0	0.0	0.0	0.0	0.0
N=500, mean=50, std=5	0.0	0.0	0.0	0.0	0.0	0.0
N=1000, mean=50, std=5	0.0	0.0	0.0	0.0	0.0	0.0
overall	0.0	0.0	0.0	0.0	3.6	0.7

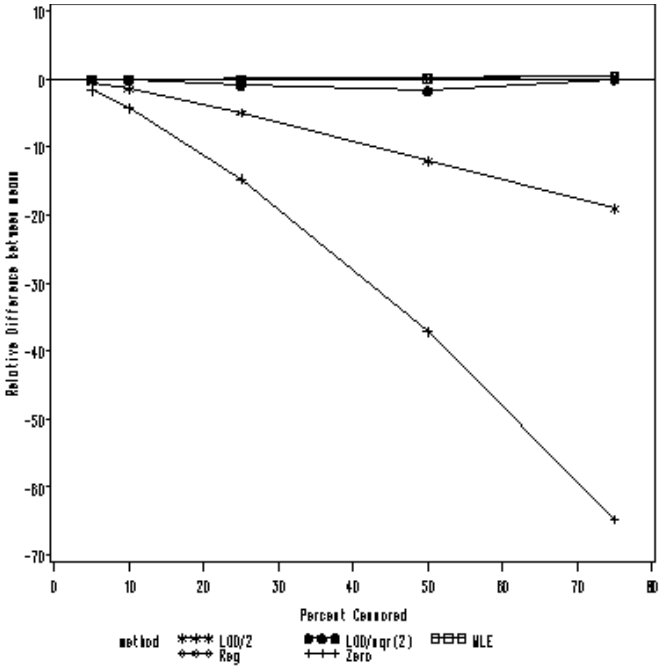
Bias can also be measured with the percent relative difference between the derived and true means. Using this measure, the effect of choice of parameters is removed. There is little

difference between the extrapolation and the MLE techniques (Table 4). The replacement techniques performed worse than the computational methods, with the replacement by zero having the largest relative difference (Figure 1). The LOD/√2 replacement has the smallest relative difference of all the replacement methods using these data.

Table 4: Percent Relative difference between derived and true means

Cut-off Level	MLE	Ext	Replacement Technique Zero	LOD/2	LOD/√2
5	-0.01	0.01	-1.67	-0.50	-0.01
10	-0.00	0.01	-4.32	-1.40	-0.18
25	-0.00	0.04	-14.71	-4.95	-0.90
50	0.06	0.13	-37.09	-12.08	-1.72
75	0.34	0.35	-64.73	-19.01	-0.07

Figure 1: Percent relative difference between the derived and true means



CONCLUSION

Distributions of normally distributed data were generated with three different mean and standard deviation combinations. These data were censored from 5 to 75 percent to test some common techniques for handling left-censored data. Simple substitution techniques are adequate if only a small percentage of the values have been censored. The error rate and relative difference in the means both increase once the percentage of censored values reaches 25%. Replacement with the LOD/√2 appears to be the best choice of replacement values. The reason for the large difference between the LOD/2 and LOD/√2 estimates is not readily apparent. The extrapolation and maximum likelihood techniques produce estimates with smaller bias and error rates than those from the replacement techniques. Furthermore, the error rate is less affected by the distribution parameters with the both the extrapolation and maximum likelihood technique. The maximum likelihood technique appears to have worked best, overall. Since the maximum likelihood technique is an iterative method, there are situations were there may not be a solution; however, the extrapolation technique has a closed form solution. In addition, the extrapolation technique is more intuitive and, thus, more transparent. Therefore, the extrapolation technique has some advantages over the maximum likelihood technique.

REFERENCES

Cohen, AC., Jr. Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics 1:217-237, 1959.

Finkelstein MM, Verma DK. Exposure estimation in the presence of nondetectable values: another look. AIHAJ 62:195-198, 2001.

Gleit A. Estimation for small normal data sets with detection limits. Environ Sci Technol 19:1201-1206, 1985.

Helsel DR, Gilliom RJ. Estimation of distributional parameters for censored trace level water quality data: 2. Verification and applications. Water Resources J 22(2):147-155, 1986.

Helsel DR. Less than obvious: statistical treatment of data below the detection limit. Environ Sci Technol 24(12)1766-1774, 1990.

Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. Appl Occup Environ Hyg. 5(1):46-51, 1990.

Lambert D, Peterson B, Terpenning I. Nondetects, detection limits, and the probability of detection. J Am Stat Assoc 86(414):266-276, 1991.

Lyles RH, Fan D, Chuachoowong R. Correlation coefficient estimation involving a left censored laboratory assay variable. Statist Med 20:2921-2933, 2001.

Newman MC, Dixon PM, Looney BB, Pinder JE. Estimating Mean and Variance for Enviromental Samples with Below Detection Limit Observations. Water Resources Bulletin 25(4): 905-915, 1989.

Özkaynak H, Xue Jianping, Butler DA, Haroun LA, MacDonell, MM, Fingleton DJ. Addressing data heterogeneity: lessons learned from a multimedia risk assessment. Presented at 84th Annual Meeting of the Air and Waste Management Association, Vancouver, BC, 1991.

Taylor JK. Quality Assurance of Chemical Measurements. Chelsea, MI: Lewis Publishers, 1987.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name:	Carry W. Croghan
Company:	US-EPA
Address:	HEASD (MD E205-1)
City state ZIP:	RTP, NC 27709
Work Phone:	(919) 541-3184
Fax:	(919) 541-0239
Email:	croghan.carry@epa.gov

This paper has been reviewed in accordance with the United States Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ATTACHED CODE

```
***************************;
proc means data=ttest1 noprint;
        by group;
        var below;
        output out=stat1 n=n sum=k;
run;

***************************;
proc rank data=two out=temp2n normal=blom;
        var x;
        ranks prob;
run;
****************************************;

proc reg data=temp3 outest=stat2 noprint;
        by group;
        model x = prob;
run;

****************************************;
*** Since there is only a mean and std and not
data, the t-test is calculated within a
datastep.
***;

data testit;
        merge stat2
              stat1 (keep=group n k) end=eof;

        replace= substr(group, 1, 1);
        level  = substr(group, 2);

** Retaining the values for the first record.
This is the parent distribution to which all
others are compared. **;

        retain mean1 s1 n1 ss1;

        if _n_=1 then do;
                mean1= intercept;
                s1 = prob **2;
                n1 = n;
                SS1 = s1*(n1-1);
        end;

        else do;
                mean2= intercept;
                s2   = prob **2;
                n2   = n;
                SS2  = s2*(n2-1);

                sp   = (ss1 + ss2) / (n1 + n2 -
2);

*** Student t ***;
                t = (mean1 - mean2) /
                   sqrt(sp**2/n1 + sp**2/n2);

                p = probt(t,n1+n2-2);



** Relative difference in the means **;
                bias = (mean2-mean1) / mean1;

        end;

run;
```