EGG-SARE-9247
Revision 1
September 1991

**Idaho
National
Engineering
Laboratory**

*Managed
by the U.S.
Department
of Energy*

**RECOMMENDED METHODS FOR
STATISTICAL ANALYSIS OF DATA
CONTAINING LESS-THAN-DETECTABLE
MEASUREMENTS**

**Corwin L. Atwood
Larry G. Blackwood
Gerald A. Harris
Cynthia A. Loehr**

*Work performed under
DOE Contract
No. DE-AC07-76ID01570*

NOV 22 1991

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

**RECOMMENDED METHODS
FOR STATISTICAL ANALYSIS OF DATA
CONTAINING LESS-THAN-DETECTABLE MEASUREMENTS**

**Corwin L. Atwood
Larry G. Blackwood
Gerald A. Harris
Cynthia A. Loehr**

September 1991

**Idaho National Engineering Laboratory
EG&G Idaho, Inc.
Idaho Falls, ID 83415**

**Prepared for the
U.S. Department of Energy
Idaho Operations Office
Under DOE Contract No. DE-AC07-76ID01570.**

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

482

ABSTRACT

This report is a manual for statistical workers dealing with environmental measurements, when some of the measurements are not given exactly but are only reported as less than detectable. For some statistical settings with such data, many methods have been proposed in the literature, while for others few or none have been proposed. This report gives a recommended method in each of the settings considered. The body of the report gives a brief description of each recommended method. Appendix A gives example programs using the statistical package SAS, for those methods that involve nonstandard methods. Appendix B presents the methods that were compared and the reasons for selecting each recommended method, and explains any fine points that might be of interest.

This is an interim version. At present it is not clear whether funding will be received to complete the recommendations in a final version. Scattered throughout the report are promises of topics to be added or augmented in the final version, but these promises are all conditional on receipt of funding.

ACRONYMS

The numbers in parentheses are section numbers giving the primary introduction of the term. A term that is mentioned in the body of the report is also often discussed in the parallel-numbered portions of Appendix A and Appendix B.

ANOVA	Analysis of variance (6.1)
CV	Coefficient of variation (2.1.2)
DL	Detection limit (2.1.5, 3.2)
DL/2	Method involving replacement of LTD observation by half the detection limit (4.1.3, 5, 6)
EM	Expectation-maximization (B-5.2)
IQR	Interquartile range (2.1.2)
ILS	Iterative least squares (A-5, B-5)
LR	Lognormal regression (4.1.2)
LTD	Less than detectable (2.1.5, 3.1)
MDL	Method Detection Limit (3.2)
MLE	Maximum likelihood estimator or estimate (4, 5, 6)
MSE	Mean squared error (2.1.3)
MWW	Mann-Whitney-Wilcoxon test (B-5)

CONTENTS

ABSTRACT	iii
ACRONYMS	iv
1 INTRODUCTION	1
1.1 Purpose and Scope	1
1.2 Report Organization	2
2 BACKGROUND	4
2.1 Definitions and Notations	4
2.1.2 Populations and Samples	4
2.1.2 Moments and Percentiles	4
2.1.3 Point Estimators and Their Properties	8
2.1.4 Interval Estimators	10
2.1.5 Censored Data	10
2.2 The Lognormal Distribution	11
3 DETECTION LIMITS	12
3.1 LTD Measurements as They Arise in Practice	12
3.1.1 Examples of Analysis Methods That Can Yield LTD Measurements	12
3.1.2 Comments	17
3.3 Statistical Analysis Using Actual Values for LTD Measurements	18
4 ONE SAMPLE	22
4.1 Point Estimates and Confidence Intervals	22
4.1.1 The LR Estimate	22
4.1.2 Maximum Likelihood Estimate (MLE)	25
4.1.3 Simple Substitution Methods	26
4.1.4 Recommendations	27
4.2 Tolerance and Prediction Intervals	28

4.3 Goodness of Fit Tests	28
TWO SAMPLES	29
6 LINEAR MODELS	31
6.1 Two-Way Analysis of Variance	31
REFERENCES	32
APPENDIX A. EXAMPLE PROGRAMS IN SAS	A-1
APPENDIX B. TECHNICAL DETAILS AND SUPPORTING MATERIAL	B-1

FIGURES

Figure 2.1.1. Examples of lognormal densities with different skewnesses.	6
Figure 3.1. Gamma-ray spectrum of ^{131}Ba obtained with a 65-cm ³ Ge(Li) detector.	13
Figure 3.2. Traces from three components of PCB.	16
Figure 4.1.1. Graphical illustration of LR method, showing uncensored values as solid, and estimated censored values on fitted line. . .	24

RECOMMENDED METHODS FOR STATISTICAL ANALYSIS OF DATA CONTAINING LESS-THAN-DETECTABLE MEASUREMENTS

1 INTRODUCTION

1.1 PURPOSE AND SCOPE

Environmental monitoring data frequently contain measurements that are reported as "less than detectable" (LTD). Standard textbook statistical methods do not allow such data; instead they require all the measurements to have known numerical values. Less well known methods have been developed in some situations to perform valid statistical analyses when some of the measurements are LTD. Many of these methods are described in a review of the literature by Hertzler, Atwood, and Harris (1989). The number of possible approaches can be bewildering. Therefore, the present report is intended to give recommended methods to use in the most commonly encountered settings.

The body of this report considers a number of settings in which statistical analysis is necessary, and presents the recommended method in each case. This portion of the report is written for data analysts working with environmental data. Appendix A follows the same order of topics, for each topic giving, where possible, a worked example using the statistical package SAS. Appendix B is written for statisticians who are interested in the details of the possible methods. Following the same order of topics as in the body of the report, it gives the methods that were compared and the reasons for selecting each recommended method, and explains any fine points that might be of interest. The numbering of the major sections of the appendices agrees exactly with the numbering of the body of this report, for easy cross-referencing.

The present interim version of this report contains recommendations as of

September, 1991. Portions of the final report are not yet written. Future revisions will complete the recommendations.

1.2 REPORT ORGANIZATION

The next section of this report consists of background information—a review of definitions, and information on the most important probability distribution, the lognormal. Following this, the three major sections of the report consider three situations: the one-sample problem, the two-sample problem, and the linear model, described briefly here.

In the one-sample setting, there is one sample of observations taken from a single population. The goal of the statistical analysis is to characterize the population in one or more ways. The parameters of interest often are:

- moments, such as the mean and variance
- percentiles, such as the median and the 95th percentile

The analysis might require

- point estimates
- confidence intervals
- tolerance intervals
- tests of hypotheses

or some other type of inference. The choice of parameters and the required analysis might be dictated by some applicable regulation.

In the two-sample setting, there are two (independent) samples from two populations. One sample might be a control. The analysis normally involves comparing the two populations, typically by considering a difference or ratio of parameters, and computing a point estimate or interval estimate, or performing a hypothesis test.

In the linear-model setting there are more than two samples, taken under differing conditions. The analysis includes, as special cases, analysis of variance (ANOVA) and regression.

2 BACKGROUND

2.1 DEFINITIONS AND NOTATIONS

2.1.2 Populations and Samples

The following definitions are stated tersely, because many of them should be familiar to readers of this report. Some quantity is being measured, such as the concentration of a contaminant. The *population* is the set of all possible values that could be produced, together with the distribution that determines the probabilities of the different values. The distribution is included in our definition of the population, because it is not enough to simply say that all non-negative concentrations are possible; the probabilities of the various concentrations are part of the definition of the population. For example, a smokestack could be emitting sulphur dioxide at a constant rate, and the random wind pattern causes various concentrations at a monitoring station. The population consists of all possible concentrations at the station (i.e., all non-negative values), together with the probability distribution of these possible values (for example, some lognormal distribution.) As a second example, surface soil in a certain region could be contaminated with ^{137}Cs . The population would consist of all possible ^{137}Cs concentrations per 10-gram (or other sized) unit of soil, along with the probability distribution of these concentrations.

A *sample* is a set of values, called *observations*, randomly chosen according to the population distribution. The number of observations in the sample is called the *sample size*, and is denoted in this report by n . Note, some experimenters use the term "sample" for an individual observation, such as a soil sample; in this report, a sample is always a set of numerical observations.

2.1.2 Moments and Percentiles

Suppose that a random quantity X is generated from a probability distribution; in other words, X is a single observation to be taken from some population. We will use capital letters for random variables, and the

corresponding lower case letters for observed values of these random variables. The use of capital X in this case means that X is considered random; its value has not yet been observed. The *expected value* of X , denoted $E(X)$, is the long-term average value that would be obtained if X were sampled infinitely often. It is a fixed number depending on the distribution for the population, not a random variable.

A *population statistic* is a number that can be computed from the population distribution. A *sample statistic* is the analogous quantity computed from the sample observations, as if the sample were the entire population (with all the observations having equal probability). Sample statistics are based on observed values, and are used to estimate population statistics. Two important classes of statistics are *moments* and *percentiles*, defined below.

Moments are expected values of various expressions. If X denotes a single observation from a population with some distribution, the following population moments are commonly used.

$$\begin{aligned} \text{mean} &= \mu = E(X) \\ \text{variance} &= \sigma^2 = E[(X - \mu)^2] \\ \text{skewness} &= E[(X - \mu)^3] / \sigma^3 \\ \text{kurtosis} &= E[(X - \mu)^4] / \sigma^4 \\ \text{coefficient of variation} &= \sigma / \mu \end{aligned}$$

In casual use, the *standard deviation*, σ , is also often called a moment, even though technically it is not the expected value of anything, but is the square root of an expected value. The coefficient of variation (CV) can also be thought of as the relative standard deviation.

The skewness measures asymmetry of the distribution. A positive skewness indicates that values far from the mean are more likely on the right than on the left; a negative skewness indicates the reverse; zero skewness results from a symmetrical distribution such as the normal. Measurements of contaminants are typically positively skewed: they cannot be less than zero, and so cannot be too far to the left of the mean; however, values that are very large (relative to the mean) occasionally occur. This asymmetry yields a positive skewness. This is illustrated in Figure 2.1.1.

Figure 2.1.1.1. Examples of lognormal densities with different skewnesses. Skewness and coefficient of variation are labeled as SK and CV.

It is often the case that both the coefficient of variation and the skewness measure the asymmetry of the distribution. This is the case in Figure 2.1.1, where the most asymmetrical distribution has both the largest skewness and the largest coefficient of variation.

For a sample $(x_1, \dots, x_n)$, the corresponding sample moments are

$$\begin{aligned} \text{sample mean} &= \bar{X} = \Sigma x_i / n \\ \text{sample variance} &= s^2 = \Sigma (x_i - \bar{X})^2 / (n-1) \\ \text{sample skewness} &= \Sigma (x_i - \bar{X})^3 / ns^3 \\ \text{sample kurtosis} &= \Sigma (x_i - \bar{X})^4 / ns^4 \end{aligned}$$

Note that the sample variance is not the exact analogue of the population variance (with each of the n observations getting weight $1/n$), because the sample variance uses a divisor $n-1$ rather than n . This is done to make s^2 an unbiased estimator of σ^2 . The sample moments are simple estimators of the corresponding population moments, and if n is large, they are good estimators.

Percentiles differ from moments in that they are not expectations. Instead, they are values that satisfy certain probability statements. If X is a random variable (that is, an observable value obeying the distribution of some population), the k th percentile is the number a satisfying

$$P(X \leq a) \times 100\% = k\%.$$

The 50th percentile is called the *median*. On the average, half the observations from the population are below the median and half are above it. The 75th percentile minus the 25th percentile is sometimes called the *interquartile range* (IQR). For a lognormal distribution, the ratio

95th percentile / median

is sometimes called the *error factor*.

The *sample percentiles* are the analogues for a sample of the above population percentiles. They are often expressed in terms of the *ordered* observations, commonly written with subscripts enclosed in parentheses $x_{(1)} \leq x_{(2)}$

$\leq \dots \leq x_{(n)}$. For example, if there are five observations, their values, in the order originally observed, might be

$$x_1 = 5.2$$

$$x_2 = 2.7$$

$$x_3 = 10.8$$

$$x_4 = 7.4$$

$$x_5 = 3.1$$

The ordered observations are

$$x_{(1)} = 2.7$$

$$x_{(2)} = 3.1$$

$$x_{(3)} = 5.2$$

$$x_{(4)} = 7.4$$

$$x_{(5)} = 10.8$$

The *sample median* in this case is $x_{(3)}$, or 5.2. In general, if n is odd, the sample median is the middle ordered observation, $x_{(n+1)/2}$. If n is even, there is no universally accepted definition of the sample median, although the average of the two middle ordered observations is commonly used. The other sample percentiles are similarly defined, with similar difficulties of detail if n is not neatly divisible by the right number.

Some statistical techniques make use of *ranks*. The rank of an observation is an integer from 1 to n , with the smallest observation having rank 1, the next smallest having rank 2, and so forth. In the above example, the rank of x_1 is 3, the rank of x_2 is 1, etc.

2.1.3 Point Estimators and Their Properties

Consider now properties of estimators. An *estimator*, such as a sample moment or a sample percentile, is a quantity computed from a random sample and used to approximate some unknown quantity that depends on the population. The numerical value it takes based on a particular observed sample is called an

estimate. An estimator is a random variable; that is, it is based on a random sample, and if a different sample were taken the calculated estimate would be a different number. Because the estimator is a random variable, it has a mean and a variance. Let θ be an unknown parameter, such as μ or σ , and let U be an estimator of θ . For example, the sample mean $\bar{X}$ is a natural estimator of the population mean μ , the sample coefficient of variation $S/\bar{X}$ is a natural estimator of the coefficient of variation σ/μ , etc. The value of the estimator depends on the random sample that happens to be observed. The *bias* of U , as an estimator of θ , is defined as

$$\text{bias}(U) = E(U) - \theta .$$

The estimator U is *unbiased* if the bias is zero, so that the long term average value of U is the desired quantity θ . In environmental contexts, the term *accuracy* is often used to express this idea, with high accuracy corresponding to small bias. The bias of an estimator cannot be considered without also considering the variance of the estimator; an unbiased estimator is of little use if its variance is enormous. In environmental contexts, *precision* is often used to express this idea, with high precision corresponding to small variance. The variance and bias are sometimes combined to give an overall measure of the goodness of an estimator, the *mean squared error* (MSE):

$$\text{MSE of } U = E[(U-\theta)^2] = \text{variance}(U) + \text{bias}^2$$

Often, when there are several candidate estimators of a single quantity, the analyst will select the estimator with the smallest MSE. This method may not give the best estimate for the particular data set being analyzed (if we knew which estimate was best, we would not need to estimate anything), but on the average the method has the smallest squared error.

One more desirable property of an estimator is robustness. The properties of an estimator, such as its variance or bias, are usually derived based on some assumptions, for example the assumption that the population is lognormal. An estimator is *robust* if its properties do not change greatly when the underlying assumptions are only approximately true. Robustness studies are often empirical;

for example, an estimator derived under the lognormal assumption can be studied using simulated data sets from other skewed populations such as gamma or Weibull.

2.1.4 Interval Estimators

An unknown parameter such as μ or σ can be estimated by an interval as well as by a point estimate. The interval is intended to include the value of the parameter with high probability. Such an interval is called a *confidence interval*. It is computed from the sample; therefore different random samples would produce different intervals. For a $k\%$ confidence interval, the probability that the interval contains the unknown parameter is $k\%$. This confidence level, $k\%$, is a property of the method, not of the particular calculated interval. We never know whether we got an unlucky sample, so that the interval misses the unknown value of the parameter, or a more typical sample, so that the interval contains the parameter. We only know that if we took many samples and computed the many resulting confidence intervals, $k\%$ of the intervals would contain the true value of the parameter.

There are other kinds of intervals, tolerance intervals and prediction intervals. Their definitions will be added to a future revision, when the recommended methods are given.

2.1.5 Censored Data

Data are *censored* if some or all of the observation values are not known exactly. For a censored observation, it is only known that the value is bounded by some number, or is in some interval. In environmental contexts, some observations are reported as less than detectable (LTD); this means that the value is less than some known detection limit (denoted DL), but is not known more exactly.

In the literature on censored data, two types of censoring are distinguished. In *Type I* censoring, an observation is censored if it is beyond a predetermined value. This is the type of censoring that is relevant in the environmental setting, in which observations are reported as less than detectable

if they are below a fixed value DL. Note, the value of the detection limit does not have to be the same for all the observations, but it must exist in principle and be independent of the values of the observed data. By contrast, in *Type II* censoring the number of censored observations is some predetermined fraction of the sample size, and the cutoff value between the censored and uncensored observations is random. This kind of censoring occurs in survival tests, when it is decided in advance to watch a number of items until a certain fraction of them have failed. Type II censoring is not used in the context of this report.

2.2 THE LOGNORMAL DISTRIBUTION

Environmental measurements considered here are typically measurements of concentrations, measured in parts per million or picocuries per liter or similar units. The concentration being measured varies randomly, depending on the particular sample of air, soil, water, or other medium, and it cannot be negative.

The most important distribution used to model such a random variable is the lognormal distribution. The defining relation is that X is lognormal if its logarithm is normal. By custom, the natural logarithm $\ln(X)$ is used. The lognormal distribution is so widely used because of its simple relation to the normal distribution. If $x_1, \dots, x_n$ are a set of measurements, define $y_i = \ln(x_i)$. Assuming that the x 's come from a lognormal population means that the y 's come from a normal population. Therefore many classical methods can be used with the y 's, and the results translated back to give results about the population of the x 's. For an introductory discussion of the lognormal distribution, see the relevant section of Appendix B and the references mentioned there.

Many of the statistical methods presented here are based on lognormality. The assumption of lognormality is usually made for convenience, not for any theoretical reason. To be recommended, therefore, a method must be robust, that is, it must perform well even when the assumption of lognormality is not exactly correct.

3 DETECTION LIMITS

This section gives examples of settings in which LTD measurements arise in practice, and recommends using the actual numerical values for statistical analysis whenever possible. Formal definitions of detection limits, with mathematical formulas, are given in Section B.3 of Appendix B.

3.1 LTD MEASUREMENTS AS THEY ARISE IN PRACTICE

Several laboratory methods are described below, illustrating ways in which measurements could arise that are LTD. For statistical analysis, it is of interest to know if the actual measurement, rather than merely the notation LTD, is obtainable; therefore, this topic is discussed for each example. Finally, some general comments are made.

3.1.1 Examples of Analysis Methods That Can Yield LTD Measurements

Gamma Spectroscopy. Gamma spectrometers measure the activity of gamma radiation emitted by radionuclides at specific energies. The data produced are usually in the form of an energy spectrum, a graph with the activity measured on the vertical axis and energy on the horizontal axis. Generally, the unit of measure on the vertical axis is the log of the counts, although actual counts are sometimes used. The horizontal axis measures energy in either kilo-electron volts or in channel numbers (there being a direct correspondence between channel numbers and energy). An example graph is shown in Figure 3.1.

The presence of a specific nuclide in a sample is identified by a peak in the spectrum at an energy known to be produced by that particular nuclide. The instrument is calibrated by analyzing pure or mixed samples of known amounts of known nuclides and developing a library of characteristic peaks. Given a good calibration library, a sample containing many elements (and thus having many, perhaps overlapping, peaks) can be successfully analyzed.

Figure 3.1. Gamma-ray spectrum of ^{131}Ba obtained with a 65-cm^3 Ge(Li) detector. Peaks are labeled with the γ -ray energies in keV; those marked BG are background lines. The upper (lower-energy) curve is displaced upward by a factor 10. [From R. J. Gehrke et al., *Phys. Rev. C*14, 1896 (1976).]

Measuring the concentration of a particular nuclide in a sample involves defining a peak at the known location for the nuclide and estimating the peak height or area under the peak that is above the background. Background is usually determined in some manner by considering the counts in the channels adjacent to the peak. Various statistical methods for curve fitting, curve smoothing, and baseline construction are used to select the boundaries and approximate the shape of the peaks in the spectrum. The number of counts in the peak region is used to calculate the amount of the radionuclide present.

When it is not known what nuclides are in a sample (that is, there is no specific set of nuclides of interest), a peak search algorithm can be employed. In this case, empirical criteria are used to search for peaks, and then the library is consulted for nuclides with peaks in the same channels as identified in the search procedure.

In gamma spectroscopy, large uncertainties leading to LTD measurements can occur when the net count for a nuclide is small relative to background or when peaks occur close together. In the latter case, there is difficulty distinguishing whether they are clearly single full-energy peaks, Compton edges, backscatter peaks, or multiplets (composed of two or more overlapping peaks). It is also possible for the instrument's own electronics to contribute noise (usually only significant in the lower energy ranges), making it difficult to establish peaks. In all these cases various decision rules are used to determine if a detection has actually occurred.

In the simplest case, once a peak has been identified and a net count established, a measurement is LTD when the standard deviation (which, assuming a Poisson distribution, is the square root of the count) is too large relative to the count itself. Declaring a measurement LTD when the net count is less than either two or three times the its standard deviation is common. A measurement can also be declared LTD when there was difficulty in determining that the peak existed.

Most gamma spectroscopy analysis is automated to at least a certain extent by a computer. Whether or not numeric values can always be obtained for less

than detectable measurements is dependent on the algorithm used for analysis. If the channels for the peak of a constituent of interest are known from calibration, then a numeric value is obtainable (except in cases where software will not print out a value once it is determined to be LTD—in that case new software is called for). This value could be negative if the "peak" is inverted (the count in the sample is lower than the count for the adjacent channels).

Peak search algorithms may not always yield a numeric result because calculations do not occur unless a peak is detected. When a peak is found by a search algorithm, a numeric value is calculated, although the constituent may be declared not detected. In the case of an "inverted" peak, no peak would be found and thus no numerical value obtained.

Gas Chromatography (GC). Problems with obtaining numerical values with gas chromatography can be illustrated by considering the measurement of PCB contamination in oil. When testing oil for PCB contamination using GC, the presence of PCB produces a series of spikes on a graph. (The x-axis of the plot is travel time through a column, and the y-axis is detector response.) A schematic diagram is shown in Figure 3.2. The relationship of the spikes to each other is characteristic of the particular contaminant, such as PCB, and is recognized by the eye of a trained analyst or by smart software. The area under the curve is proportional to the concentration of contaminant in the oil. Unfortunately, the oil itself produces a trace (analogous to background). Therefore, if the height of most of the spikes is below this trace, the contaminant cannot be identified as PCB, much less quantified. This causes a minimum detection limit, below which no number can be obtained, and probably dominates the usual considerations of measurement uncertainty. This masking is called "matrix interference." If the oil is dirty, so that other contaminants are producing their own spikes, the presence of PCB may be masked. In this case the detection limit for PCB is larger than if the oil were clean.

When there is a high concentration of PCB or a different contaminant in the oil, the spikes may go off the scale of the instrument. In this case, the oil must be diluted and reanalyzed. This raises the detection limit. For example, if the normal detection limit is 2 ppm but the oil was diluted by a factor of 10,

FIG. 1 Column: 3 % OV-1, Carrier Gas: Nitrogen at 60 mL/min, Column Temperature: 170°C, Detector: Electron Capture¹³

FIG. 2 Column: 3 % OV-1, Carrier Gas: Nitrogen at 60 mL/min, Column Temperature: 170°C, Detector: Electron Capture¹³

FIG. 3 Column: 3 % OV-1, Carrier Gas: Nitrogen at 60 mL/min, Column Temperature: 170°C, Detector: Electron Capture¹³

NOTE 1—Improper chromatographic conditions can lead to poor PCB separation, evidenced by overlapping peaks with few intervals where the instrument response returns to base line. The results can be erroneous. Chromatograms closely comparable to those shown in the accompanying figures should be obtained for maximum accuracy.

Figure 3.2. Traces from three components of PCB. (From ASTM D 4059-91.)

then the detection limit becomes 20 ppm. This should not cause a problem if the only contaminant is PCB, because the dilution should only take place when there is enough PCB in the oil to be above the new detection limit. The combination of reported value of LTD together with a very large detection limit should only be a result of matrix interference by other contaminants, not a result of diluting oil that is clean except for PCB.

Liquid Chromatography. Grant et. al (1991) reported on calculating detection limits for the determination of 1,3-dinitrobenzene (DNB) by reversed-phase high performance liquid chromatography. Measurements are obtained by either manual or integrator evaluation of plots of peak height vs. concentration (chromatograms) obtained from strip chart recorders. Even though they were studying method detection limit and certified reporting limit calculations (so were measuring quite low values, down to 0.125 $\mu\text{g/L}$) they mentioned no problem with not ever getting a numerical value. In fact, blank samples were employed to determine baseline measurements. Since measurement values were always obtained for the blanks, obtaining reportable numeric values apparently was not a problem for this method.

Atomic Absorption. Grant et. al (1991) also studied detection limits in the determination of copper (Cu) concentration by graphite furnace atomic absorption (GFAA). Basically the same comments as given for liquid chromatography apply. Again, adjustments to the data based on measurement of furnace blanks indicates that a numeric value can always be obtained.

3.1.2 Comments

Errors in the analysis procedure can also result in a value that is known to be LTD with no numeric value available. An example of such an error is an improper dilution procedure where an initial measurement "peaks out" the instrument while the next dilution is too large and the instrument doesn't record any amount of the target analyte. Another situation in which gross procedural errors can lead to LTD measurements without reportable numeric values is when an instrument is set up or calibrated to only measure values in a certain interval,

but measurements come in below the lower bound. For example, certain pressure transducers need to be set for the anticipated range of pressures to be encountered. If the transducer is set for a middle to high range pressure and a low value is encountered, all that will be known is that the value was less than the lower limit the instrument was set to record.

It must be recognized that detection limits are not fixed. Grant et al. state that "detection capability is not a fundamental property of procedures, nor is it a constant ... it can be managed to a significant extent based on the experimental protocol (number of standards and blanks analyzed, distribution of standards, time span covered by standards measurements, number of replications of standards and unknowns, etc.)" In cases where LTD measurements are occurring frequently and when the current detection limit is close to or above concentration levels of interest, changes in methods should be employed that will lower the level.

Changes in procedures to lower detection limits should be made only if necessary since, given a fixed budget, the additional cost incurred in lowering the detection limits means reductions in other areas such as number of field samples obtained will have to be made. If a detection level is only 1% of the level of interest (e.g. some action level), little relative error would be introduced by even using the simplest statistical analysis methods--simply assuming everything LTD was actually zero or the detection limit. It is only when the detection limit becomes a sizeable value relative to levels of interest that their treatment in the statistical analysis becomes critical to the results.

3.3 STATISTICAL ANALYSIS USING ACTUAL VALUES FOR LTD MEASUREMENTS

One method of dealing with LTD measurements that has received less attention than others is avoidance—if there is a numeric value available for every data point, there is no censoring problem. While some LTD measurements result from an inability of an analysis method to yield any numeric value at all

for a measurement, in many cases a numerical value is obtained but reported as LTD because it did not meet some precision requirement. In this second case the LTD problem can be avoided by using the actual measurement obtained, in spite of (and perhaps accounting for) its imprecision. The availability of numeric values for LTD measurements for various types of environmental data is discussed in Section 3.1.

We are not arguing here that the LTD distinction is not useful. The use of the designation LTD for measurements came about from the desire to control false positive and negative conclusions regarding the presence of a constituent in individual samples. For individual data points the LTD distinction is appropriate (although we would argue that there is benefit from reporting the measurement value along with the disclaimer that it does not indicate detection with any certainty). However, when individual measurements are to be combined in some form via statistical analysis, the issue of false positive and negative values for individual measurements and thus the LTD distinction is irrelevant. So true numeric values should be used whenever possible. (False positive and negative errors for the combined data are still relevant and can be addressed in the final statistical results.)

The statistician may have difficulty obtaining these true numeric values. Lambert et al. (1991) put it well: "If the uncensored data are not requested before the data are collected, they may be difficult to retrieve. Moreover, even if chemists have the unconfirmed data, they may be reluctant to give it to statisticians. Chemists traditionally have made measurement-by-measurement decisions about the quality of their data and, 'staking their reputation on each number reported,' have withheld any measurement that failed their stringent quality control requirements. This stance may be reasonable when only one field sample is analyzed, but in large studies it fails to balance the chemist's responsibility for reporting only the highest quality measurements against the statistician's responsibility for ensuring that parameters on the pollutant are estimated as well as possible." Therefore, the statistician would be wise to establish an explicit agreement with the laboratory for the reporting of values that are LTD, before any samples are sent for analysis.

When numeric values are obtained for LTD measurements, the problem in statistical analysis with missing or censored observations is eliminated. This does not mean however, that statistical analysis is completely straightforward. In particular, there are two potential analyses that may need to be addressed when using such data:

- When a clearly identifiable peak exists but the count is low, the actual numeric value can be useful in further statistical analysis. In this case there may be some value from taking several LTD measurements, averaging them, and then coming up with a more precise (non-LTD) measurement of the actual level of a rare constituent. In GC, this averaging would be equivalent to increasing precision by extending the count time for a single measurement. When, on the other hand, the LTD designation arises from problems with peak definition, or from false peaks (due to electronic noise, improper classification of peaks due to multiplets, backscatter peaks, miscalibration, etc.), then the measurement obtained is not just a poor estimate of the constituent of interest, but a misidentified value altogether. All that is being measured is the bias and variability of the measurement system. In this case, the suggestion of averaging LTD values would result in averaging measurements that are totally unrelated to anything of interest, yet it might be declared non-LTD. This could be very misleading and, in many applications, there is no way of knowing if this situation exists. Sometimes, by using "professional judgement," an experienced analyst can identify spurious peaks by carefully studying the printed spectrum or chromatograph.
- Distributional assumptions and measurement models may need to be adjusted when these measurements are included in statistical analysis. For example, consider estimating the mean concentration of a constituent whose distribution in the environment is lognormal but is measured by a procedure whose error is normal. In the case where all values are high relative to the measurement error (that is, there are no LTD measurements), the lognormal character of the

true distribution of the constituent in the environment dominates over the normal measurement error and calculating the mean value assuming a lognormal distribution should yield good results. However, when many LTD measurements are present and the actual measurement values are used, the distribution no longer approximates a lognormal since the normally distributed measurement error would dominate. A different analysis might be called for.

4 ONE SAMPLE

4.1 POINT ESTIMATES AND CONFIDENCE INTERVALS

Three types of estimate are described: the LR estimate of Gilliom and Helsel (1986), the maximum likelihood estimate (MLE), and estimates based on substitution simple quantities for LTD values. For estimating the median and the interquartile range (IQR), the MLE is best, according to extensive simulations performed by Gilliom and Helsel. The LR method is nearly as good. It is natural to conjecture that the same qualitative findings hold when estimating percentiles such as the 90th percentile, but we do not know of studies on this.

For estimating moments of the distribution, such as the mean and variance, the properties of the estimators depend on the distribution's (unknown) coefficient of variation (CV). The MLE can be extraordinarily bad when the CV is large. Simple substitution methods have properties that are very dependent on the relationship between the substituted value and the CV. The LR method is never much worse than the method that would be best if the CV were known, and therefore is the method of choice.

We also recommend avoiding estimation of moments if the CV appears large. Unless regulations require it, do not estimate moments at all, but estimate percentiles instead. These points are explained in more detail in the sections below.

4.1.1 The LR Estimate

The term LR stands for "lognormal regression." The LR estimate is based on the assumption that the observations are lognormally distributed, but it seems to be robust, that is, it seems to work well even if the observations in fact follow a somewhat different distribution. The idea is simple: plot the uncensored values on lognormal probability paper (or equivalently, plot their logarithms on normal probability paper). The points should fall approximately

on a straight line. Fit a straight line to these points, and extrapolate the line back below the detection limit to estimate values for the censored observations. Use these calculated values, together with the observed uncensored values, to calculate the sample mean, sample variance, or sample percentiles.

To illustrate the method, suppose that the ordered observations are the following:

<u>number</u>	<u>observation</u>	<u>ln(observation)</u>
1	<4.1	-
2	<4.5	-
3	4.5	1.50
4	5.3	1.67
5	5.3	1.67
6	5.6	1.72
7	7.7	2.04
8	8.5	2.14
9	9.5	2.25
10	9.6	2.26
11	11.0	2.40
12	17.0	2.83
13	17.0	2.83

This data set is also used in Appendix A. There are 13 observations, of which two are less than detectable. On normal probability paper, plot the logs of the uncensored observations against the points 3/14, 4/14, ..., 13/14. (The number 14 is the number of observations plus 1, and the numerators go from 3 to 13 because the ordered observations 3 through 13 are uncensored.) This is shown graphically in Figure 4.1.1. Fit a straight line through the points, and extrapolate them backward to estimate the censored values, shown by open circles on the line in the figure. Convert these replacement log-values back to the original units and use the resulting data to calculate the sample mean and sample variance. Note, because fewer than half the measurements are LTD, the LR estimate of the median is just the ordinary sample median, 7.7.

Figure 4.1.1. Graphical illustration of LR method, showing uncensored values as solid, and estimated censored values on fitted line.

There is one detail that must be made more exact, in order to actually carry out the above procedure: how, precisely, should the fitted line be determined? The formulas for least-squares fitting, which should be used rather than the above eyeball fitting, are given in Appendix B. An example using SAS is also given in Appendix A.

It is possible that a calculated x , based on extrapolating back using the LR method, may be greater than the detection limit, even though the calculated value is supposed to be used in place of a measurement that was less than detectable. This is a warning indication, perhaps suggesting that the data do not fit the assumed model very well or that the distribution is highly skewed. In this case, consult a statistician.

To get confidence intervals, Helsel and Gilliom (1986) give a method, using tables based on extensive simulation results. The method allows confidence intervals for the mean, standard deviation, median, and interquartile range. It is valid if the data come (with equal probability) from one of the four distributions with one of the four CVs and one of the four censoring levels used in the simulations. These intervals seem to have approximately correct confidence coefficients: for example, 95% confidence intervals for the mean, calculated using Helsel and Gilliom's tables, cover the mean in various simulations in 91% to 98% of the cases. Because the method is rather involved, the formulas and tables are given in Appendix B.

4.1.2 Maximum Likelihood Estimate (MLE)

When the data are censored, the MLE must be computed using numerical iteration—there is no simple formula. Maximum likelihood estimation for censored data is therefore most easily implemented with survival analysis programs in commercial packages. These programs generally allow a number of distributional assumptions. For example, the LIFEREG procedure in SAS allows the Weibull, exponential, gamma, normal, lognormal, logistic, and loglogistic distributions. A distribution must be assumed before the MLE can be computed. Gilliom and Helsel (1986) assume the lognormal distribution, and show that the MLE estimates the median and IQR as well as other estimators do, even if the data

come from some distribution other than the assumed lognormal. The MLE also estimates moments well when the distribution is not too skewed, but is very biased for estimating the moments when the distribution is strongly skewed.

See Appendix A for an example calculation of the MLE of the lognormal parameters using LIFEREG. The MLEs of moments or percentiles are obtained by substituting the parameter estimates into the appropriate formulas.

Approximate confidence intervals can be obtained by using the asymptotically normal distribution for the MLE. The SAS procedure LIFEREG gives confidence intervals for the percentiles, and a SAS program given in Appendix A gives approximate confidence intervals for the mean and variance of the distribution. Better confidence intervals, however, are found using the tables of Helsel and Gilliom (1986), as explained in Appendix B.

The MLE for moments is very sensitive to extreme observations. Therefore, when the data come from a highly skewed distribution, the MLE can be very unstable, sometimes giving an estimate of the mean that are incorrect by a factor of 1000 or more. Because the MLE is no simpler to calculate than the LR estimate, and because it is never much better than the LR estimate but in some situation much worse, it is not recommended for estimation of moments.

4.1.3 Simple Substitution Methods

These methods replace the LTD measurements by a single value, such as 0, DL, or DL/2. Then the moments are estimated by the sample moments. The properties of the methods depend on the true shape of the distribution. To illustrate this, consider the DL/2 method, defined as the method in which all LTD values are replaced by half the detection limit. Consider the distributions in Figure 2.1.1, and suppose that the detection limit DL equals the median, so that on average 50% of the observations are censored. All the distributions in Figure 2.1.1 have median 1.0, so the replacement value DL/2 equals 0.5. For the highly skewed distributions, the value of 0.5 is a very plausible value for an observation, and substitution of 0.5 may lead to good estimation. For the least skewed distribution, however, the value 0.5 is in the extreme left tail of the

distribution; assigning this value to all the LTD measurements results in an estimate of the mean that is biased on the low side.

Because the substitution methods are never much better than the LR method, and in some situations substantially worse, substitution methods are not recommended in general. If the distribution appears to have moderate to large skewness, then the DL/2 point estimate may be used, because of its simplicity. For many simulated cases, confidence intervals based on simple substitution tended to be both biased and too short. Therefore they are not recommended.

4.1.4 Recommendations

To estimate moments, such as the mean and standard deviation, use the LR method. If the DL/2 method is tentatively tried first because of its simplicity, and if the resulting estimated coefficient of variation equals 1.0 or more, then the DL/2 point estimates may be used. If the estimated coefficient of variation is less than 1.0, however, the LR point estimates should be used. In all cases, the LR method of Helsel and Gilliom should be used for confidence intervals or for tests about the moments.

The moments of an extremely skewed distribution are not very practical descriptors of the distribution. When the underlying data distribution has heavy tails, the probability of an extreme occurrence is great enough to severely inflate the moments of the distribution. Yet a researcher might reasonably expect to do a lifetime of work without encountering one of these rare occurrences. Therefore, when dealing with highly skewed distributions, it is best to avoid using moments for characterizing the distribution. Even if the mean must be used, for example because an average exposure is of regulatory interest, try to avoid estimating the variance of the distribution; the variance is more heavily influenced by rare extreme values than is the mean. Based on informal study of *uncensored* data, when the true coefficient of variation is as large as 1.0, the variance should not be estimated.

To estimate percentiles, use the MLE or the LR method, as recommended by Gilliom and Helsel (1986). For confidence intervals for the median or IQR, or

equivalently for tests about these quantities, use the confidence intervals of Helsel and Gilliom based on the MLE. Sample percentiles, such as the sample median, are equivalent to the LR estimates when the censoring fraction is low; they are not quite as good as the MLEs.

Be alert for violations of the modeling assumptions. Possible signs of this are:

- The presence of many censored observations, plus some very large observations, without many intermediate observations
- Replacement values greater than the detection limit, as calculated by the LR method.

In such cases, consult a statistician. If possible, take more observations. Try to avoid using estimates of the moments (mean and variance), which are strongly influenced by extreme observations; instead estimate percentiles, which are less sensitive to the values of extreme observations.

4.2 TOLERANCE AND PREDICTION INTERVALS

This will be added in a future revision.

4.3 GOODNESS OF FIT TESTS

This will be added in a future revision. Verrill and Johnson (1988) give tables for tests of normality when the data are censored with Type II censoring. Our very preliminary and tentative recommendation, based in part on informal conversation with Verrill, is to use these tables even with Type I censored data, as long as the sample size is at least 20.

5 TWO SAMPLES

It is a common practice in environmental data analysis to compare two groups. For example, one may wish to compare site concentrations of a contaminant to background concentrations. In ground water monitoring, one often compares downgradient concentrations to upgradient concentrations to assess whether a contaminant has entered the ground water. As always, some of the data may be censored. In this situation, the Wilcoxon rank sum test (also called the Mann-Whitney test) is recommended. The technique is easy to use and widely available on commercial statistical software. The implementation using SAS is described in Appendix A.

When there is a single detection limit (DL), the test is done as follows. First replace all the LTD values by a value smaller than DL. It does not matter whether this value is $DL/2$ or 0 or some other small value; the resulting ranks will be the same. Now rank all the data from the two groups together, assigning the average rank to any tied data. Then for each of the two groups compute the sum of those ranks, and compare the two sums. When there is a large difference between the sums of the ranks, there is strong evidence of a difference between the groups. This use of ranks makes the test robust to the form of the distribution.

When there are multiple detection limits, the modifications are more complex, but are presented by Millard and Deveral (1988). They will be added to a future version of this report.

The recommendation is more tentative when the goal is estimation rather than testing. We have taken natural logarithms of the data, assumed that each resulting distribution is normally distributed, and considered estimating the two population means, the difference between the means, or the common variance. The maximum likelihood estimator (MLE) is difficult to program. An alternative simple procedure is to replace all the LTD measurements by half the detection limit (before taking logarithms), and then use linear regression software on the log-transformed data. Call this the $DL/2$ method. When the true normal variance

is 1.0 and ten observations are taken in each group, and when there is 20% or 60% censoring, the DL/2 estimators have smaller MSE than the MLE, but larger bias. The reason why we do not simply recommend using the DL/2 method is that it has not been tested for other values of the variance and other sample sizes.

6 LINEAR MODELS

The methods of this section assume that the original data are lognormal. Therefore taking natural logarithms of the observations transforms them to a sample from a normal population. Standard linear-model methods may be used on normal data, to perform inferences about the means under the assumption of a constant variance. Therefore, the results translate to inferences about the medians of the original lognormal data, under the assumption of a constant coefficient of variation.

6.1 TWO-WAY ANALYSIS OF VARIANCE

The following recommendation in this case is based on studies of a 2×2 table, with either 3 or 8 observations per cell, and up to 60% of the observations censored, as described in Section B-6.1. The recommended method is to replace each censored observation by half the detection limit (the DL/2 method), to take the natural logarithm of each observation, and to perform the usual analysis of variance on these transformed observations. The methods considered for comparison were the DL/2 method, uniform replacement of LTDs, the LR method, and the MLE, all explained in Appendix B. Of the methods considered for comparison, this was the only one that correctly maintained a Type I error of 0.05 when up to 60% of the data was censored, and that could be used when there was a cell containing only censored data.

When main effects are present, but no interaction is present, all the methods considered tend to falsely reject the no-interaction hypothesis with probability greater than 0.05. This is apparently an inherent feature of the two-way layout with censoring.

If the purpose of the analysis is to estimate cell means (of the log-transformed distributions) rather than to test for effects, the two best estimators are DL/2 and MLE. Neither is better than the other in all circumstances, but the MLE is much much harder to implement.

REFERENCES

Robert J. Gilliom and Dennis R. Helsel, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data: 1. Estimation Techniques," *Water Resources Research*, 22, 1986, pp. 135-146.

C. L. Grant, A. D. Hewitt, and T. F. Jenkins, "Experimental comparison of EPA and USATHAMA detection and quantitation capability estimators," *American Laboratory*, Feb. 1991, pp. 15-33.

Dennis R. Helsel and Robert J. Gilliom, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data: 2. Verification and Applications," *Water Resources Research*, 22, 1986, pp. 147-155.

C. L. Hertzler, C. L. Atwood, and G. A. Harris, *Current Methods of Handling Less-than-Detectable Measurements and Detection Limits in Statistical Analysis of Environmental Data*, EGG-SARE-8609, September 1989.

Diane Lambert, Bruce Peterson, and Irma Terpenning, "Nondetects, Detection Limits, and the Probability of Detection," *Journal of the American Statistical Association*, 86, 414, 1991, pp. 266-277.

S. P. Millard and S. J. Deveral, "Nonparametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Nondetect Limits," *Water Resources Research*, 24, 12, 1988, pp. 2087-2098.

Steve Verrill and Richard A. Johnson, "Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality," *Journal of the American Statistical Association*, 83, 404, 1988, pp. 1192-1197.

APPENDIX A

EXAMPLE PROGRAMS IN SAS

CONTENTS.

A-1	INTRODUCTION	A-5
A-2	THE LOGNORMAL DISTRIBUTION	A-5
A-3	DETECTION LIMITS	A-5
A-4	ONE SAMPLE	A-6
	A-4.1 Point Estimates and Confidence Intervals	A-6
	A-4.1.1 LR Estimates	A-6
	A-4.1.2 Maximum Likelihood Estimation	A-9
A-5	TWO SAMPLES	A-13
A-6	LINEAR MODEL	A-13
	A-6.1 Two-Way Analysis of Variance	A-13
	REFERENCES	A-14

APPENDIX A

EXAMPLE PROGRAMS IN SAS

A-1 INTRODUCTION

This appendix gives example SAS programs (SAS, 1988a,b,c) for performing some of the statistical techniques used in this report. To run one of the examples, copy the program printed here to a file in your SAS directory, with extension 'sas'. For example, suppose you name it 'myprog.sas'. Then you may run it by entering the DOS command 'sas myprog'. The results will appear in files named 'myprog.log' and 'myprog.lst'. To adapt the program to use your own data rather than example data, change the data lines in the example.

If no special programming is necessary in order to carry out the method in SAS, then no example is given here. Examples are given only when a nontrivial sequence of steps is needed. Therefore, nonparametric methods and methods for which censored values are replaced by an easily calculated number are not illustrated here.

A-2 THE LOGNORMAL DISTRIBUTION

This section is included to keep the numbering system parallel to that in the body of the report. The examples in this appendix either assume that the observations are lognormal, or that they have already been transformed by taking natural logarithms and are therefore normal.

A-3 DETECTION LIMITS

This section is also included to keep the numbering system parallel to that in the body of the report.

A-4 ONE SAMPLE

A-4.1 POINT ESTIMATES AND CONFIDENCE INTERVALS

A-4.1.1 LR Estimates

The following SAS program carries out the LR method for estimating the mean and variance of a population. The method is based on the assumption that the population is lognormal.

```
*INPUT DATA, CENS=1 IF MEASUREMENT CENSORED (I.E. Y IS A DETECTION LIMIT);
*CENS=0 OTHERWISE;
data one;
input y cens;
cards;
4.1 1
4.5 1
4.5 0
5.3 0
5.3 0
5.6 0
7.7 0
8.5 0
9.5 0
9.6 0
11.0 0
17.0 0
17.0 0
;

*SORT DATA SO THAT _N_ CONTAINS RANKS;
proc sort;
by y;

data one;
set one;
rank=_N_;

*CALCULATE LOGS OF UNCENSORED VALUES;
z=log(y);
if cens=1 then z='.';

*CALCULATE NORMAL SCORES FOR RANKS;
proc rank out=out normal=vw; var rank; ranks nscore;

*PERFORM REGRESSION, OUTPUT PREDICTED VALUES;
```

```

data one;
set out;
proc reg;
model z=nscore;
output out=out predicted=pred;

*CALCULATE MOMENTS USING PREDICTED VALUES FOR CENSORED DATA;
data one;
set out;
if cens=1 then y=exp(pred);
proc univariate;
var y;
run;

```

The results of running the above program are printed here, copied from the DOS file with extension 'lst'. The analysis-of-variance table is produced from the regression of the observations on the normal scores. The fact that R-square is large indicates that the fit is good. For a more careful check of this, residual plots should be examined. These are obtained by running PROC REG with option R; that is, the model line is

model z=nscore / r;

After the censored observations have been replaced based on extrapolating the regression line to the censored observations, the sample is described by PROC UNIVARIATE. The sample moments are given in the table of moments below. The sample mean is 8.267533 and the sample variance is 21.12309 (printed to unreasonably many digits.)

SAS 11:49 Monday, September 23, 1991 1

Model: MODEL1
Dependent Variable: Z

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	1.99825	1.99825	207.762	0.0001
Error	9	0.08656	0.00962		
C Total	10	2.08481			

Root MSE	0.09807	R-square	0.9585
Dep Mean	2.12008	Adj R-sq	0.9539
C.V.	4.62582		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	1.973496	0.03126956	63.112	0.0001
NSCORE	1	0.636632	0.04416777	14.414	0.0001

SAS 11:49 Monday, September 23, 1991 2

UNIVARIATE PROCEDURE

Variable=Y

Moments

N	13	Sum Wgts	13
Mean	8.267533	Sum	107.4779
Std Dev	4.595986	Variance	21.12309
Skewness	0.986277	Kurtosis	0.22215
USS	1142.054	CSS	253.4771
CV	55.59078	Std Mean	1.274697
T:Mean=0	6.48588	Prob> T	0.0001
Sgn Rank	45.5	Prob> S	0.0002
Num ^= 0	13		

Quantiles(Def=5)

100% Max	17	99%	17
75% Q3	9.6	95%	17
50% Med	7.7	90%	17
25% Q1	5.3	10%	3.646785
0% Min	2.831144	5%	2.831144
		1%	2.831144
Range	14.16886		
Q3-Q1	4.3		
Mode	5.3		

Extremes

Lowest	Obs	Highest	Obs
2.831144(	1)	9.5(	9)
3.646785(	2)	9.6(	10)
4.5(	3)	11(	11)
5.3(	5)	17(	12)
5.3(	4)	17(	13)

A-4.1.2 Maximum Likelihood Estimation

The following SAS program uses the SAS procedure LIFEREG to obtain maximum likelihood estimates (MLEs), assuming a lognormal population with mean α and standard deviation β . The corresponding normal population has mean and standard deviation μ and σ , respectively. MLEs are found for μ , σ , α , β^2 , and for selected percentiles of the lognormal population. The asymptotic variances are found for the estimators, so that approximate confidence intervals can be calculated. The procedure can easily be modified so that a distribution other than the lognormal is used; in this case the parameters called INTERCPT and SCALE by PROC LIFEREG no longer correspond to a normal μ and σ , but the other results are unchanged.

Even though many quantities are printed by this program, only the point estimates of the percentiles are recommended for use in the body of this report. For moments, the LR estimates are recommended. For a confidence interval for the median or IQR, the Helsel and Gilliom intervals explained in Section B-4.1.2 are recommended. For other quantiles, the MLE asymptotic confidence intervals may be used, with the understanding that they are probably too short.

```
data;
input y lower;
control=0;
if _n_=1 then control=1;
cards;
4.1 .
4.5 .
4.5 4.5
5.3 5.3
5.3 5.3
5.6 5.6
7.7 7.7
8.5 8.5
9.5 9.5
9.6 9.6
11.0 11.0
17.0 17.0
17.0 17.0
;
```

```

proc lifereg outest=out covout;
model (lower,y) = /d=lnormal covb;
output out=quant quantiles=.50 .90 std=std p=estquant control=control;
run;

```

```

*CALCULATE 95% CONFIDENCE INTERVAL FOR QUANTILE ESTIMATES;
data;
set quant;
estimate=log(estquant);
stde=std/estquant;
upper=exp(estimate + 1.96*stde);
lower=exp(estimate - 1.96*stde);
quantile=_prob_;

```

```

proc print;
var quantile estquant lower upper;
title 'Quantile estimates and 95% confidence limits';

```

```

*PUT PARAMETER ESTIMATES AND COV MATRIX INTO MACRO VARIABLES;
data one;
set out;
if type_='PARMS' then do;
call symput('M',intercep);
call symput('S',_scale_);
end;
if _name_='INTERCPT' then do;
call symput('A',intercep);
call symput('C',_scale_);
end;
if _name_='SCALE' then call symput('B',_scale_);

```

```

*RETRIEVE MACRO VARIABLES, CALCULATE LOGNORMAL MEAN AND VARIANCE AND ;
*COVARIANCE MATRIX;

```

```

data one;
*MU=NORMAL MEAN;
mu=symget('M');
*SIG=NORMAL STANDARD DEV.;
sig=symget('S');
*K11=VARIANCE OF MU;
k11=symget('A');
*K22=VARIANCE OF SIG;
k22=symget('B');
*K12=COVARIANCE OF MU AND SIG;
k12=symget('C');

```

```

*INVERT AND TAKE NEGATIVE OF K MATRIX TO GET INFO MATRIX J;
det=k11*k22-k12**2;
j11=-k22/det;
j22=-k11/det;
j12=k12/det;

```



```

*CALCULATE LOGNORMAL MEAN AND VARIANCE;
*A=LOGNORMAL MEAN;
a=exp(mu+.5*sig**2);
*B2=LOGNORMAL VARIANCE;
b2=exp(2*mu+sig**2)*(exp(sig**2)-1);

*CALCULATE INFORMATION MATRIX FOR LOGNORMAL MEAN AND VARIANCE;
*DUDA=DERIVATIVE OF MU WRT A, DSDB2=DERIVATIVE OF SIG WRT B2, ETC.;
duda=(1/a)*(1+1/(1+a**2/b2));
dudb2=-1/(2*(b2+a**2));
dsda=-1/(sig*a*(1+a**2/b2));
dsdb2=1/(2*sig*(b2+a**2));
k11=j11*duda**2+2*j12*dsda*duda+j22*dsda**2;
k22=j11*dudb2**2+2*j12*dsdb2*dudb2+j22*dsdb2**2;
k12=j11*duda*dudb2+j12*(dsda*dudb2+dsdb2*duda)+j22*dsda*dsdb2;

*INVERT AND TAKE NEGATIVE OF NEW K MATRIX TO GET VAR/COV MATRIX J;
det=k11*k22-k12**2;
va=-k22/det;
vb2=-k11/det;
cab2=k12/det;

file print;
title 'Estimates of lognormal mean and variance';
put;
put 'Lognormal mean, a = ' a;
put;
put 'Lognormal variance, b2 = ' b2;
put;
put 'Asymptotic variance of estimator of a = ' va;
put;
put 'Asymptotic variance of estimator of b2 = ' vb2;
put;
put 'Asymptotic covariance of estimators of a and b2 = ' cab2;
run;

```

The output from running the above example is printed here, copied from the DOS file with extension 'lst'. In the output of PROC LIFEREG, the variables called INTERCPT and SCALE are the normal parameters μ and σ , respectively.

SAS 10:51 Thursday, September 27, 1990

1

L I F E R E G P R O C E D U R E

Data Set =WORK.DATA1
Dependent Variable=Log(LOWER)
Dependent Variable=Log(Y)
Noncensored Values= 11 Right Censored Values= 0
Left Censored Values= 2 Interval Censored Values= 0

Log Likelihood for LNORMAL -10.86329568

SAS 10:51 Thursday, September 27, 1990

2

L I F E R E G P R O C E D U R E

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	1.97430789	0.151434	169.9744	0.0001	Intercept
SCALE	1	0.53518773	0.118413			Normal scale parameter

Estimated Covariance Matrix

	INTERCPT	SCALE
INTERCPT	0.022932	-0.001621
SCALE	-0.001621	0.014022

Quantile estimates and 95% confidence limits 3
10:51 Thursday, September 27, 1990

OBS	QUANTILE	ESTQUANT	LOWER	UPPER
1	0.5	7.2016	5.35214	9.6902
2	0.9	14.2988	9.57762	21.3474

Lognormal mean, $a = 8.31050888$

Lognormal variance, $b^2 = 22.905916033$

Asymptotic variance of estimator of $a = 1.7413221958$

Asymptotic variance of estimator of $b^2 = 241.86162368$

Asymptotic covariance of estimators of a and $b^2 = 14.411935268$

A-5 TWO SAMPLES

When there is only one detection limit, the Wilcoxon-Mann-Whitney test can be performed in SAS by using the WILCOXON option in PROC NPARIWAY. For this, assign the value 0 (or any other single number less than the detection limit) to all observations that are less than the detection limit; SAS then treats them as ties. When there are multiple detection limits, the method is more complex, and will be discussed in a future revision of this report.

The method of iterative least squares (ILS) is presented in Appendix B. It has not been programmed in SAS, because to do so would be quite difficult. Other languages, such as Fortran or Pascal, are more suited to tasks such as ILS.

A-6 LINEAR MODEL

A-6.1 TWO-WAY ANALYSIS OF VARIANCE

To implement the recommended method, replace all censored observations by half the detection limit. Take natural logarithms of the resulting set of values (uncensored observations plus replaced censored values), and perform a two-way analysis of variance using PROC ANOVA.

REFERENCES

SAS Institute Inc., *SAS Language Guide for Personal Computers, Release 6.03 Edition*, Cary, NC: SAS Institute Inc., 1988a.

SAS Institute Inc., *SAS Procedures Guide, Release 6.03 Edition*, Cary, NC: SAS Institute Inc., 1988b.

SAS Institute Inc., *SAS/STAT User's Guide, Release 6.03 Edition*, Cary, NC: SAS Institute Inc., 1988c.

APPENDIX B
TECHNICAL DETAILS AND SUPPORTING MATERIAL

CONTENTS

B-1	INTRODUCTION	B-7
B-2	THE LOGNORMAL DISTRIBUTION	B-8
B-2.1	Introduction	B-8
B-2.2	Justification for Using the Lognormal	B-8
B-2.3	Properties of Lognormal Variables and Their Implications for Statistical Analysis	B-10
B-2.3.1	Mathematical Properties	B-10
B-2.3.2	Measures of Central Tendency	B-16
B-2.3.3	Measures of Dispersion	B-19
B-2.3.4	Confidence Intervals and Hypothesis Tests	B-19
B-2.4	Discussion and Recommendations	B-21
B-2.4.1	Use of the Logarithmic Transformation	B-21
B-2.4.2	Alternatives to the Lognormal Distribution	B-21
B-2.4.3	Robustness of Lognormal Analysis Methods	B-22
B.3	FORMAL DEFINITIONS OF DETECTION LIMITS	B-24
B-4	ONE SAMPLE	B-28
B-4.1	Point Estimates and Confidence Intervals	B-28
B-4.1.1	The LR Method	B-29
B-4.1.2	Maximum Likelihood Methods	B-35
B-4.1.3	The DL/2 Method	B-37
B-5	TWO SAMPLES	B-38
B-5.1	Two-Sample Testing: Nonparametric Methods	B-38
B-5.1.1	Simulation Methods	B-39
B-5.1.2	Discussion	B-40
B-5.2	Two-Sample Testing: Iterative Least Squares	B-42
B-5.2.1	Specification of the Method	B-43
B-5.2.2	Simulations and Competing Methods	B-46

B-5.2.3 Results	B-48
B-5.3 Two-sample Testing: DL/2	B-52
B-5.3.1 Simulation Settings	B-52
B-5.3.2 Computing Methods	B-52
B-5.3.3 Results	B-53
B-6 LINEAR MODELS	B-56
B-6.1 Simulations on Two-Way Analysis of Variance with Replacements for Censored Values	B-57
B-6.1.1 Introduction	B-57
B-6.1.2 Simulation Technique	B-57
B-6.1.3 Simulation Results	B-66
B-6.1.4 Conclusions	B-104
B-6.1.5 SAS Programs for Data Generation and ANOVA Simulation	B-105
REFERENCES	B-113

FIGURES

Figure B-2.3.1. Lognormal densities with $\sigma = 1$ and various μ	B-13
Figure B-2.3.2. Lognormal densities with $\mu = 0$ and various σ	B-14
Figure B.3-1. Relation of L_c and L_D , from Currie (1968).	B-25
Figure B.3-2. Power curve illustrating L_c and L_D	B-25

TABLES

Table B-4.1.1. <i>Rmses</i> of LR and ML estimators, from simulations. . .	B-31
Table B-4.1.2. Biases of LR and ML estimators, from simulations. . .	B-32
Table B-4.1.3. Simulated Coverage Percentages of Confidence Intervals.	B-34
Table B-4.1.1. Nonparametric 2-sample test simulation results. . . .	B-41
Table B-5.2.1. Summary of testing results for the second set of simulations.	B-50
Table B-5.2.2. Summary of testing results for the third set of simulations.	B-51

Table B-5.3.1. Summary of two-sample comparison results.	B-54
Table B-6.1.1 Treatment combination means considered for normal distributions with standard deviation σ	B-58
Table B-6.1.2. Simulated power with 3 replicates per cell, DL/2 and uniform replacement.	B-67
Table B-6.1.3. Simulated power with 8 replicates per cell, DL/2 and LR methods.	B-68
Table B-6.1.4. Simulated power with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-69
Table B-6.1.5. Simulated power with 8 replicates per cell, DL/2 and adjusted MLE (EM).	B-70
Table B-6.1.6. Simulated relative bias of mean squares with 3 replicates per cell, DL/2 and uniform replacement.	B-73
Table B-6.1.7. Simulated relative bias of mean squares with 8 replicates per cell, DL/2 and LR methods.	B-74
Table B-6.1.8. Simulated relative bias of mean squares with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-75
Table B-6.1.9. Simulated relative bias of mean squares with 8 replicates per cell, DL/2 and adjusted MLE (EM).	B-76
Table B-6.1.10. Simulated variance of mean squares with 3 replicates per cell, DL/2 and uniform replacement.	B-78
Table B-6.1.11. Simulated variance of mean squares with 8 replicates per cell, DL/2 and LR methods.	B-79
Table B-6.1.12. Simulated variance of mean squares with 3 replicates per cell, DL/2 and adjusted MLE (EM) methods.	B-80
Table B-6.1.13. Simulated variance of mean squares with 8 replicates per cell, DL/2 and adjusted MLE (EM) methods.	B-81
Table B-6.1.14. Simulated relative MSE for the mean squares with 3 replicates per cell, DL/2 and uniform replacement.	B-82
Table B-6.1.15. Simulated relative MSE for the mean squares with 8 replicates per cell, DL/2 and LR method.	B-83
Table B-6.1.16. Simulated relative MSE for the mean squares with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-84
Table B-6.1.17. Simulated relative MSE for the mean squares with 8 replicates per cell, DL/2 and adjusted MLE (EM).	B-85

Table B-6.1.18. Simulated bias of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-87
Table B-6.1.19. Simulated bias of parameter estimators with 8 replicates per cell, DL/2 and adjusted MLE (EM).	B-89
Table B-6.1.20. Simulated mean squared error (MSE) of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-91
Table B-6.1.21. Simulated mean square error (MSE) of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-93
Table B-6.1.22. Simulated bias of estimators of cell means, with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-95
Table B-6.1.23. Simulated bias of estimators of cell means, with 8 replicates per cell, DL/2 and adjusted MLE (EM).	B-97
Table B-6.1.24. Simulated mean squared error (MSE) of estimators of cell means with 3 replicates per cell, DL/2 and adjusted MLE (EM).	B-99
Table B-6.1.25. Simulated mean squared error (MSE) of estimators of cell means with 8 replicates per cell, DL/2 and adjusted MLE (EM).	B-101

APPENDIX B

TECHNICAL DETAILS AND SUPPORTING MATERIAL

B-1 INTRODUCTION

This appendix contains fuller details regarding the methods recommended, as well as mention of methods that were considered but not ultimately recommended. The comparisons that form the basis of the recommendations are described. This appendix is written at a more demanding technical level than the body of the report. It is intended for statisticians who need to know the details.

B-2 THE LOGNORMAL DISTRIBUTION

B-2.1 INTRODUCTION

This section is adapted from Blackwood (1990). For a more complete discussion, refer to that report.

A random variable is said to be lognormally distributed if its natural logarithm follows a normal distribution. Lognormal distributions allow only positive, non-zero values and are positively skewed. Since radiological and other measurements in the environment are often intrinsically positive as well as highly skewed, the assumption of a lognormal distribution for these variables is common.

Much work already exists summarizing the properties and applications of the lognormal distribution. Most notable are the books by Aitchison and Brown (1957) and Crow and Shimizu (1988). However, these treatments concentrate more on the theoretical aspects of the distribution, or on non-environmental applications. The purpose of this section is to review the properties of the lognormal distribution in the context of the analysis of environmental data, emphasizing the issues related to data analysis and interpretation. The focus is on how the "usual" approach to estimation and hypothesis testing (i.e., the way we do it with normally distributed data) must be re-evaluated when considering lognormal data.

B-2.2 JUSTIFICATION FOR USING THE LOGNORMAL

Empirically the lognormal distribution seems to fit radiological data from environmental settings. This may be justification in itself for its use. Also, an adaptation of the model described as the theory of breakage by Aitchison and Brown (1957) can be readily adapted to the diffusion of radionuclides or pollutants to provide a theoretical justification for the lognormal distribution. This is summarized by Blackwood (1990).

Whatever theoretical basis exists for assuming a lognormal distribution for radionuclide or related data, pragmatic reasons for its use seem to dominate in the literature. The lognormal distribution has emerged as the most popular alternative to assuming normality most likely because it seems to fit most data sets adequately and because of the convenience it affords in data analysis. It is the next best thing to a normal distribution in the sense that, once the data is transformed, all the usual normal theory analysis methods apply. The following examples illustrate the pervasive empirical basis of the lognormal distribution for environmental data in general and radiological data in particular.

Eberhardt and Gilbert (1980), observing that frequency distributions of transuranic data are "often dramatically skewed," recommend using "a logarithmic transformation of the data before doing any statistical analysis involving significance tests based on the normal distribution." Georgopoulos and Seinfeld (1982) reported that "the two-parameter lognormal has been the most popular in representing air pollutant concentration data." Bernarie (1971) gives a justification for a lognormal distribution for ground level concentration of atmospheric dispersion of a single pollutant source in a single direction. He argues that all the generally accepted theoretical formulae for atmospheric dispersion indicate that ground level concentration is inversely proportional to the instantaneous wind speed. Furthermore, empirical observations have shown the distribution of wind speed to generally follow a lognormal distribution. So, since the inverse of a lognormal distribution is also lognormal, the ground level concentration should be lognormally distributed as well.

Perhaps the best description of the pragmatic use of the lognormal distribution comes from Dennis and Patil (1988) whose comments relative to general ecology studies also apply to radionuclides in the environment:

Possibly the main role of the lognormal in ecology is simply to serve as the handiest adjustable wrench in the toolbox of statistical distributions. Ecological abundance data are intrinsically positive, with a few enormously high data points

typically arising in every study. The lognormal distribution is an ideal descriptor of such data, with a positive range, right skewness, heavy right tail, and easily computed parameter estimates.

If data are skewed and do not show a constant variance across groups with different mean values, often a logarithmic transformation is used simply to stabilize the variance and make the data symmetrical so that applying normal theory analysis methods seems more appropriate. The implicit assumption in this situation is that, since normal theory methods are being applied to the log transformed data, the original data must be approximately lognormal. However, as discussed earlier, a logarithmic transformation will only be successful in stabilizing the variance if the error term in the original underlying model is multiplicative. Eberhardt and Gilbert (1973) argue for a multiplicative error structure for environmental data by noting that, while means range over many orders of magnitude across different data sets, coefficients of variation stay relatively constant, making normally distributed additive errors unlikely.

B-2.3 PROPERTIES OF LOGNORMAL VARIABLES AND THEIR IMPLICATIONS FOR STATISTICAL ANALYSIS

B-2.3.1 Mathematical Properties

The lognormal distribution implies a multiplicative rather than an additive model for the data. This can be most easily seen by beginning with a measurement model which has additive, normally distributed, error or distributional terms. To legitimately use normal theory methods on data, a measurement model such as

$$y_i = \mu + m_i,$$

is assumed, where y_i is the data value, μ is a mean value and m_i is a random, normally distributed term with mean zero and variance σ^2 (denoted $N(0, \sigma^2)$).

If such data are obtained through a logarithmic transformation of data, that is $y = \ln(x)$, then the original lognormal model must be of the form

$$\begin{aligned} x_i &= \exp(y_i) \\ &= \exp(\mu + m_i) \\ &= \exp(\mu) \exp(m_i) \end{aligned}$$

so the effects in the original lognormal model are multiplicative rather than additive. In other words, the variance of the X_i is a function of μ in the lognormal model but the variance of Y_i is not a function of μ in the normal model.

In order to discuss the statistical properties of the lognormal distribution, some more rigorous mathematical notation is helpful. If $Y = \ln(X)$ is normally distributed with mean μ and variance σ^2 , then X is said to be lognormally distributed with parameters μ and σ^2 . The normal distribution of Y is denoted $N(\mu, \sigma^2)$. Similarly $LN(\mu, \sigma^2)$ denotes the lognormal distribution of X . Some important properties of the lognormal distribution are

$$\text{mean} = e^{\left(\mu + \frac{1}{2}\sigma^2\right)}$$

$$\text{variance} = e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1)$$

$$\text{median} = e^{\mu}$$

$$\text{skewness} = (e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^{\frac{1}{2}}$$

$$\text{kurtosis} = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$$

$$\text{coefficient of variation} = (e^{\sigma^2} - 1)^{\frac{1}{2}}.$$

Parameters of a distribution can take many forms. Three basic types that relate to the various forms of the density function are location, scale, and shape parameters. In terms of a graph of the probability density function (e.g. Figures B-2.3.1 and B-2.3.2), a location parameter establishes the position of a distribution along the x axis. Changing the location parameter is equivalent to adding a constant to every x value, thus shifting a distribution left or right along the x axis. Scale parameters are multiplicative and provide the measurement scale for the x axis. Changing a scale parameter amounts to multiplying each x value by a constant, shrinking or expanding the graph but not otherwise changing its shape. Shape parameters affect the general configuration of a distribution function in other ways than just changing the scale or location. Examples are parameters determining skewness and kurtosis.

For a normal distribution, $N(\mu, \sigma^2)$, μ is the location parameter and σ is the scale parameter. Normal distributions do not have shape parameters, thus the skewness and kurtosis are fixed for all normal distributions.

Unlike the normal distribution, the lognormal distribution has no location parameter but does have a shape parameter. It has no location parameter because it is always tied to zero, the minimum value. That it has a shape parameter means varying values of skewness and kurtosis can be obtained in lognormal distributions.

The scale parameter for a lognormal is e^μ , the median of the distribution. Thus increasing e^μ in a lognormal distribution stretches the distribution in an analogous manner to increasing σ in a normal distribution. This is illustrated in Figure B-2.3.1 which shows the probability density function for three lognormal distributions with varying μ , but the same σ^2 value.

Increasing μ also increases the mean, mode, and median of the distribution, so is related to measures of central tendency. Skewness and kurtosis are constant for all three distributions in the figure, since they are functions of σ^2 alone. That is, any two lognormal distributions with the

Figure B-2.3.1. Lognormal densities with $\sigma = 1$ and various μ .

Figure B-2.3.2. Lognormal densities with $\mu = 0$ and various σ .

same σ^2 value but different μ values will still have the same shape (in the same sense that two normals with differing values of σ^2 have the same shape).

The shape parameter for the lognormal distribution is σ . The range of shapes obtainable with the lognormal distribution and the relationship of the parameters to the shape of the distribution can be observed in Figure B-2.3.2, which show the graphs of several lognormal probability density functions for varying values of σ but a constant μ value. The graphs illustrate that, the greater the value of σ , the more skewed the distribution and the heavier the tail (kurtosis). The coefficient of variation also increases with larger values of σ .

Because the lognormal distribution is always positively skewed, the mode of a lognormal distribution is always less than the median which is in turn less than the mean. Note that for the smallest value of σ in Figure B-2.3.2, the shape and symmetry of the lognormal might, at least at a casual glance, be mistaken for that of a normal distribution. The mode, median, and mean for this distribution are .94, 1.0, and 1.03 respectively, indicating the near symmetry of the distribution. The skewness is .78 and kurtosis is 4.1 compared to the corresponding values of 0 and 3 for any normal distribution. For this distribution and a sample size of 25, the probability of detecting a departure from normality is less than 20% (for a significance test with a 5% Type I error rate). This compares to a probability of essentially 100% when $\sigma = 1.27$ (the most skewed distribution in the graph).

For any lognormal distribution the power for detecting departures from normality is a decreasing function of σ^2 . This is true because, as σ^2 goes to zero, the standardized lognormal distribution tends to a standardized normal distribution (Shimizu and Crow, 1988). That the difference between a normal and lognormal distribution is hard to detect in small samples with small σ^2 values may explain why the normal distribution is sometimes found to provide an adequate fit to radiological data more typically expected to be lognormal.

B-2.3.2 Measures of Central Tendency

Measures of central tendency are indicators of typical or average values used to summarize a distribution. The most common measures of central tendency are the (arithmetic) mean, median, and mode. For symmetric, unimodal distributions such as the normal, all three measures take on the same value, so there is no problem with choosing the most appropriate measure. For the lognormal distribution however, a decision must be made as to the best summary measure of central tendency. The choice will vary depending on what the purpose of measurement is.

The Mean. Summary statistics are often used to compute total dosage, lifetime exposure, the total amount of a constituent in a specific area, or other similar measurements. In this case an estimate of the mean of the untransformed data is the proper choice of summary value. Denote the population mean by α . As noted above, α is a function of both μ and σ^2 . Thus it is not purely a measure of central tendency in the same sense as the mean for a normal distribution is.

There are various ways of estimating α , some better suited for particular purposes than others. Given the relationship of α to μ and σ^2 , that is

$$\alpha = e^{\left(\mu + \frac{1}{2}\sigma^2\right)},$$

one method of estimation is to replace μ and σ^2 in this equation with the sample estimates from the transformed data,

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 :$$

This gives the maximum likelihood (ML) estimator for α . (Note that $\hat{\sigma}^2$ is the ML estimator of σ^2 rather than the usual unbiased estimator which has $n - 1$ rather than n as the divisor.) This estimator has several desirable properties. Besides being easy to calculate, it approaches unbiasedness and minimum variance for large samples. However, for small samples, it can be extremely biased and can have infinite variance (Shimizu, 1988, p. 38). For further discussion, see Blackwood (1990).

The simplest reasonable estimator of α for the lognormal is the sample mean, the (arithmetic) mean of the untransformed sample data:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

The sample mean is unbiased, but has a larger variance than the MLE. However, for smaller values of σ^2 and small sample sizes, the sample mean is nearly as efficient as the MLE. Koch and Link (1980) suggest using the sample mean for lognormal distributions whenever the coefficient of variation is estimated to be less than 1.2. For larger values of σ^2 , the sample mean estimator still compares favorably when only the types of data sets actually likely to be analyzed are considered (Blackwood, 1991).

The sample mean is unbiased regardless of the true underlying distribution of the data. Thus even if an error is made in specifying the distribution, or if the distribution is only approximately lognormal, the sample mean still retains its optimal property. This is not true of the other estimators.

One caution in using the sample mean as an estimator of α needs to be mentioned. Even though the sample mean is unbiased for the lognormal distribution, it will underestimate α most of the time when small sample sizes are combined with large σ^2 values. This result seems surprising at first, but it is an obvious result of the skewness of the distribution. For any distribution, the most likely value will be the mode, which will be considerably smaller than the mean in a highly skewed distribution. While an occasional large observation (much larger than the mode or the mean) will occur which brings the average mean calculation up sufficiently for the estimator to be unbiased, these large values will occur only rarely in small samples. So most of the time the sample mean will underestimate α .

The Median. When evaluation of total exposure is not the purpose of calculating a summary measure, the median, $e^{\bar{y}}$, may be a more reasonable measure of central tendency for the lognormal distribution. In addition to being a more "typical" value when the data are highly skewed, the median for the lognormal distribution is a function of μ alone so does not have the interpretation problem associated with α . That is, the median can change without there being a corresponding change in the shape of the distribution (skewness, kurtosis, etc.).

The usual sample estimate of the median is the geometric mean,

$$\bar{x}_g = e^{\bar{y}} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} .$$

This is the maximum likelihood estimator for the median and thus has desirable asymptotic (large sample) properties. However, the estimator is positively biased and for small samples could be misleading. (The estimate $\bar{y}$ is unbiased for μ but the exponential transformation, since it is not linear, produces a biased estimate of the median of the x population.)

The simplest estimator of the population median is just the sample median of the x 's. This nonparametric estimator will not be fully efficient for truly lognormal data. However, it is easy to calculate, considerably less sensitive to outliers than the geometric mean, and it does not depend on the form of the population distribution.

B-2.3.3 Measures of Dispersion

It is tempting to relate the variance of the lognormal distribution, denoted here by β^2 , to dispersion in the same manner as the variance of a normal distribution is used. However, as with the use of α described above, β^2 is related to both the scale and shape parameters of the lognormal so the analogy with the normal distribution is not strictly valid. Furthermore, the skewness of the distribution affects the interpretation. For a normal distribution, the variance relates directly to how far from the mean value an observation is likely to lie, in either direction. However in the lognormal case the probability of an observation occurring a fixed distance from the mean is different in one direction than it is in the other.

Because of the difficulties of interpretation, the value of β^2 as a measure of dispersion for a lognormal distribution is unclear (with the exception of lognormals that are close to being normal). Alternative measures of dispersion include stating the value of σ^2 , or calculating percentiles, interquartile ranges, etc. The geometric standard deviation, defined as e^{σ} , and the error factor (the ratio of the 95th percentile to the 50th percentile) are used sometimes as well.

B-2.3.4 Confidence Intervals and Hypothesis Tests

Confidence intervals for functions of μ alone, such as the median e^{μ} , can be easily calculated by making use of the normally distributed transformed data. If (a, b) defines the usual normal theory confidence interval for μ based on the log transformed data (i.e. $\bar{y} \pm t s_y / n$, where t is the appropriate Student's t value and s_y is the standard deviation of the observed y values), then an appropriate confidence interval for any monotone function $f(\mu)$ is just

$(f(a), f(b))$. The interval for $f(\mu)$ will be optimal in the same sense that (a, b) is optimal for μ (e.g. uniformly most accurate).

For the median the confidence interval is (e^a, e^b) . Note that the interval is not symmetric about the median estimate. That is

$$e^b - e^{\bar{y}} \neq e^{\bar{y}} - e^a .$$

However, the ratios are equal,

$$\frac{e^b}{e^{\bar{y}}} = \frac{e^{\bar{y}}}{e^a} .$$

Functions of σ^2 alone are also calculated easily by referring to the chi-square distribution of $(n - 1)s_y^2/\sigma^2$ and then transforming. Functions of σ^2 of interest might be the geometric standard deviation (e^σ), the coefficient of skewness, the coefficient of kurtosis, and the coefficient of variation.

Since the usual normal theory does not consider confidence intervals for joint functions of μ and σ^2 , interval estimates for measures such as α and β^2 , and for other functions of both μ and σ^2 , are more difficult. Some methods are discussed by Blackwood (1990).

The usual advice for hypothesis tests involving lognormal variables is to transform the data and use standard normal theory methods. If there is no need to express results in the original scale, then no further changes in the standard methods are necessary. This means that a hypothesis that involves only μ or only σ , such as a hypothesis about the median e^μ , is most easily tested. For further discussion, see Blackwood (1990).

B-2.4 DISCUSSION AND RECOMMENDATIONS

B-2.4.1 Use of the Logarithmic Transformation

In terms of making the most use of standard normal theory methods, the most straightforward method of dealing with lognormal data is to do a logarithmic transformation of the data and state all estimates, hypotheses, and test results in terms of the transformed data. Then everything proceeds as in the usual normal case. Eberhardt and Gilbert (1980) recommend that, if the results of statistical tests are of primary interest, then they should be given in terms of the transformed data.

If it is not practical to discuss everything in terms of the transformed data, in particular if at least some statement of results in the original scale are required, then stating required estimates and tests in terms of the median of the distribution affords the most direct analysis. Normal theory methods can be used for estimates and tests of the median with only minor modification (e.g. stating hypotheses in terms of ratios, and getting confidence intervals not symmetric about the estimate).

If the mean on the lognormal scale (or other compound functions of μ and σ^2) is of interest, methods other than those based on normal theory are required. Two possibilities are maximum likelihood estimators (and likelihood ratio tests), or nonparametric estimators such as the sample mean.

B-2.4.2 Alternatives to the Lognormal Distribution

Application of the lognormal distribution to certain classes of data have been so common that researchers sometimes fail to consider other distributions that might work better. The lognormal assumption is often just one of convenience for positive highly skewed data. It allows some justification for using normal theory methods but the required assumptions of multiplicative errors etc., may not be consistent with the true underlying distribution. Even if multiplicative errors apply, the transformed data may still be far

from normal so that even appealing to the robustness of certain normal theory methods may not make sense. In that case, other distributions or methods may provide a more reasonable alternative.

Variations of the two-parameter lognormal are sometimes appropriate. The delta lognormal, which allows for zero values, is useful in certain applications, such as when there is no naturally occurring radiation so there can be true zero values in samples. The three parameter lognormal applies in cases where there is an unknown minimum value other than zero. While between-sample variability in radiological measurements may be lognormally distributed, individual radiological measurements are basically count data usually thought to follow a Poisson distribution. So a lognormal-Poisson mixture distribution may be useful in some applications as well. In this case, substitution of a gamma distribution for the lognormal is analytically convenient, because the gamma-Poisson mixture distribution is a negative binomial distribution.

Other parametric distributions can also be used successfully with positive, skewed data. Eberhardt and Gilbert (1973) studied the gamma distribution in relation to the distribution of radionuclide concentrations in biota and substrates. They reported some empirical evidence that the gamma distribution provides a better fit than the lognormal. Speer and Waite (1975) fit Weibull distributions to radiological environmental surveillance with some success. Johnson and Symons (1980) reported a better fit of the Weibull compared to the lognormal when applied to a set of ambient air quality data.

Nonparametric methods should also be considered as an alternative to assuming lognormality. In the analysis of weather modification data, one area where the lognormal distribution has been traditionally assumed, nonparametric methods are now preferred due to their wider range of validity (Crow, 1988).

B-2.4.3 Robustness of Lognormal Analysis Methods

Given that lognormal methods are often used on pragmatic grounds, it is of interest to consider what happens when they are used on data that are in

fact from other distributions. In a study of censored data analysis methods, Gilliom and Helsel (1986) found that estimation methods based on assuming lognormality worked well for data from gamma distributions, delta lognormal distributions, and mixtures of two lognormal distributions. Blackwood (1991) found that lognormal methods produced acceptable results when used on normally distributed data. Eberhardt and Gilbert (1973), using simulations to compare the performance of F-tests, concluded that assuming either the gamma or lognormal distribution provide about the same levels of power and robustness regardless of which of the two distributions the data actually originated from. They concluded that assuming lognormality for convenience has little penalty in terms of accuracy of results.

B.3 FORMAL DEFINITIONS OF DETECTION LIMITS

In a widely cited article, Currie (1968) defines three levels. Let X be a measurement, with mean μ and standard deviation σ . Currie's article assumes that X is normally distributed. Note, X is not a raw measurement, but has been transformed to an unbiased estimator of μ . For example, to measure radiation in water, X is not a raw count; the background has been subtracted, and the units have been converted, for example, to pCi/L.

The *critical level*, L_c , is defined by

$$P[X > L_c \mid \mu = 0] = \alpha,$$

where α is a specified Type I error, such as 0.05. Calculation of L_c must be based on repeated measurements of blanks; blanks are used to make $\mu = 0$, and the repeated measurements provide an estimate of σ . Currie also calls L_c a "decision limit," at which one may decide whether or not the result of an analysis indicates detection.

The *detection limit*, L_D , is defined by

$$P[X < L_c \mid \mu = L_D] = \beta,$$

where β is a specified Type II error, such as 0.05. If σ is nearly constant as μ ranges from 0 to L_D , the same value of σ that was used to calculate L_c may be used here. Currie says that L_D is the level at which a given analytical procedure may be relied upon to lead to detection. Figure B-3.1 illustrates these two definitions. Figure B-3.2 also illustrates the two definitions in terms of a power curve. In this context, imagine a test of the hypothesis $H_0: \mu = 0$. The hypothesis is rejected whenever $X > L_c$. The probability of rejecting H_0 depends on the actual value of μ . It is α when $\mu = L_c$ and $1 - \beta$ when $\mu = L_D$.

The *determination limit*, L_Q , is defined by $\sigma_Q/L_Q =$ a required relative standard deviation, such as 10%. Here σ_Q denotes the standard deviation of X when $\mu = L_Q$. Currie says that for μ above L_Q , the procedure will be sufficiently precise to yield a satisfactory quantitative estimate. L_Q is also called a "quantitation limit."

Figure B.3-1. Relation of L_c and L_D , from Currie (1968).

Figure B.3-2. Power curve illustrating L_c and L_D .

The smallest of the three numbers, L_c , can be used for any particular analysis, to yield a decision of whether the contaminant of interest is present or not. Note that L_c is compared to the measurement X , not to the true mean μ , in its definition. Currie calls this analysis of a particular measurement the *a posteriori* problem, the problem of how to use a particular measurement. The other two numbers, L_D and L_Q , describe the average behavior of the measurement process, and do not refer to the decision to be made in any particular analysis. Note that L_D and L_Q are equated to the mean μ , not to a measurement X , in their definitions. Currie calls this analysis of the process the *a priori* problem. Although the true value of μ is not known, in principle the measurements can be used in the following ways, depending on the true value of μ :

$\mu < L_D$	unreliable detection
$L_D < \mu < L_Q$	detection: qualitative analysis
$L_Q < \mu$	determination: quantitative analysis.

Currie recommends using every measurement, and reporting a one-sided confidence interval if $X < L_c$ and a two-sided confidence interval if $X > L_c$. Current practice in environmental work does not follow this recommendation.

The *Method Detection Limit (MDL)* is defined by the EPA (Glaser et al., 1981) as L_c with $\alpha = 0.01$. Clayton et al. (1987) deal with the calibration problem, based on a finite number of measurements. They recommend $\beta = 0.5$ for defining L_D ; this makes $L_D = L_c$. They point out that if $\alpha = 0.01$, then the detection limit L_D equals the MDL. They also note that if $\mu = \text{MDL}$, this method yields LTD observations 50% of the time.

The frequently-used formula $L_D = 3s$ results from applying the above recommendations based on 8 measurements of blanks. This is because Student's t distribution has $t_{0.99}(7) = 3.00$. Therefore, if X is a new observation with $\mu = 0$, and s is the estimated standard deviation based on the 8 observations of blanks, we have

$$P[X > 3.00s] = P[X/s > t_{0.99}(7)] = 0.01.$$

Glaser et al. try to give careful guidelines for analytical laboratories when measurements are taken to determine the MDL; their formulas rely on assumptions of normality with constant relative standard deviation. Although the MDL is often used as "the" detection limit, the practice is not universal. For example, the 1988 Love Canal study, described by Lambert et al. (1991), instead used Currie's L_D , with $\alpha = 0.05$ and $\beta = 0.05$. Lambert et al. also cite Currie (1988) as reporting great inconsistency in definitions of detection limit as used in recent peer-reviewed literature.

B-4 ONE SAMPLE

B-4.1 POINT ESTIMATES AND CONFIDENCE INTERVALS

Three estimation methods are presented in the subsections below, LR estimation, maximum likelihood (ML) estimation, and DL/2 replacement. The first two methods give the best estimators, according to the extensive simulations of Gilliom and Helsel (1986). Three methods not considered in those simulations are the bias-adjusted MLE with censored data (Cohn, 1988), the DL/2 substitution method, and linear combinations of order statistics (Gupta, 1952, and Sarhan and Greenberg, 1962). The bias-adjusted MLE is mentioned below in the discussion of the MLE. The DL/2 method is discussed in Section B-4.1.3 below. Linear combinations of order statistics are discussed here:

Linear combinations of order statistics are estimators of the form

$$\mu^* = \sum \beta_i X_{(i)}$$

$$\sigma^* = \sum \gamma_i X_{(i)}$$

The order statistic notation is defined in Section 2.1.2 of the body of this report, and the sums here are over the uncensored order statistics. Values of the coefficients β_i and γ_i are tabulated by Gupta (1952) and Sarhan and Greenberg (1962), for various levels of censoring. The coefficients are optimal in the following sense: among all estimators that are linear combinations of order statistics, the coefficients presented give the minimum-variance unbiased estimates, under Type II censoring. Although the estimators can be used with Type I censoring, they are biased estimators in this case, and have larger MSEs than the MLEs have (Blackwood, 1989). Therefore they are not recommended for use with Type I censored data, and are not considered further here.

B-4.1.1 The LR Method

The LR method is recommended by Gilliom and Helsel (1986). Although the method can be thought of informally as a graphical method, a more precise calculational procedure is needed for computer simulations. This more precise method is given here.

Denote the observations by $x_1, \dots, x_n$, of which c are censored, that is, c are below the detection limit. Order the observations as $x_{(1)}, \dots, x_{(n)}$, with the uncensored observations being $x_{(c+1)} \leq \dots \leq x_{(n)}$. Let Φ denote the standard normal cumulative distribution function, tabulated in any statistics book. $\Phi(x)$ takes values from 0 to 1 as x goes from $-\infty$ to ∞ . Let Φ^{-1} be the inverse function; for example $\Phi^{-1}(0.95) = 1.645$, and $\Phi^{-1}(0.025) = -1.96$. For each uncensored observation $x_{(i)}$, compute the normal score $z_i = \Phi^{-1}[i/(n+1)]$. We assume that the observations come from a lognormal population, that is, that the values of $\ln(x)$ come from a normal population. For the uncensored observations, perform a least squares fit of $\ln(x_{(i)})$ on z_i ; this can be done with a commercial statistics package as a simple linear regression of $\ln(x)$ on z . Graphically, the values of z can be plotted on the horizontal axis and the values of $\ln(x)$ on the vertical axis. The fitted line passes approximately through the points. Now use the fitted line to extrapolate back and find the values of x that correspond to z_1 through z_c . The actual values are unknown, but less than the detection limit; therefore use these calculated values, based on extrapolating the fitted line, instead. (In the rare case that a fitted value is greater than the detection limit, it seems reasonable to use the detection limit rather than the fitted value, although Gilliom and Helsel always used the fitted values in their simulations.) Take the entire set of x values, c calculated and the rest observed, and calculate the sample mean and the sample variance. These are the LR estimates of the population mean and variance. An example using SAS is presented in Appendix A.

For confidence intervals, Helsel and Gilliom (1986) present tables, based on simulations. The simulations used four distributional forms (lognormal, gamma, mixture of two lognormals, and delta-lognormal), four CVs (0.25, 0.5, 1.0, and 2.0), four censoring levels (20%, 40%, 60%, and 80%), and three

sample sizes (10, 25, and 50). The confidence intervals below are based on the assumption that the data set comes from a population that was drawn at random from one of the distributional forms, CVs, and censoring levels of the simulation.

The tables are reproduced here as Tables B-4.1.1 and B-4.1.2. To use them, the *relative interquartile range (rqr)* of the data must be calculated. This is defined as the interquartile range of the uncensored data, divided by the detection limit. The tables show the bias and root mean squared error (*rmse*) of the estimators for certain sample sizes (*n*), true censoring fractions (to be estimated by the observed censored fractions), and *rqr* values.

Denote the unknown parameter of interest by θ ; this may be either the mean, the standard deviation, the median, or the IQR. Denote the point estimate by θ^* . This is the LR estimate of the mean or standard deviation, or the MLE of the median or the IQR. To obtain a confidence interval for θ , calculate the *rqr* from the data set, and read the *rmse* from Table B-4.1.1 and the bias *b* from Table B-4.1.2. Define

$$\begin{aligned}\sigma_e^2 &= rmse^2 - b^2 \\ \sigma_y^2 &= \ln[1 + \sigma_e^2/(b+1)^2] \\ \mu_y &= \ln(b+1) - \sigma_y^2/2\end{aligned}$$

Then a $(1-\alpha)$ confidence interval for θ is given by

$$\theta^* \exp[-\mu_y - z_{1-\alpha/2} \sigma_y] \leq \theta \leq \theta^* \exp[-\mu_y + z_{1-\alpha/2} \sigma_y]$$

where $z_{1-\alpha/2}$ is the 100(1 - $\alpha/2$) percentile of the standard normal distribution.

The following example of a confidence interval for the mean is taken from Helsel and Gilliom. Let the data set be

LTD, LTD, LTD, LTD, LTD, 6, 6, 6, 8, 10, 10, 10, 11, 20,

Table B-4.1.1.1. Rmses of LR and ML estimators, from simulations. (From Helsel and Gilliom)

Method	Censored at 20th Percentile						Censored at 40th Percentile						Censored at 60th Percentile						Censored at 80th Percentile						
	N	x	s	m	LM	iqr	N	$\bar{x}$	s	m	LM	iqr	N	$\bar{x}$	s	m	LM	iqr	N	$\bar{x}$	s	m	LM	iqr	
n = 10	127	•11/24	•48/59	•10/16	•37/43	•13/26	187	•13/26	•52/60	•9/20	39/43	176	•13/30	•56/67	•14/41	43/45	•42/52	39/43	297	•19/35	50/50	•22/130	39/80		
n = 25	104	•6/13	37/42	•5/7	26/27	•7/16	160	•7/16	39/44	•6/9	27/30	176	•9/15	43/47	•8/26	26	•30/38	32		12	43	22			
n = 50		4	33	4	18	5	35	5	35	4	23	23	11	37	5	26							32		
		$Rqr < 0.47$						$Rqr < 0.35$						$Rqr < 0.25$						$Rqr < 0.16$					
n = 10	355	20/23	•43/48	17/18	38/40	21/24	322	21/24	45/49	18/21	39/41	279	•21/29	47/48	27/64	43/45	43/45	44/43	268	24/29	42/41	82/100	44/43		
n = 25	377	•11/18	•32/39	10/12	25/28	•12/19	366	•12/19	•34/40	11/14	26/31	298	16/20	38/40	15/20	35/35	35/35	44/43		18	31	39	44		
n = 50		7	22	7	18	8	25	8	25	8	20	20	10	29	10	34									
		$Rqr = 0.47-1.2$						$Rqr = 0.35-0.84$						$Rqr = 0.25-0.60$						$Rqr = 0.16-0.41$					
n = 10	301	32/35	57/52	26/30	•45/64	30/37	236	30/37	55/52	28/35	•44/61	213	31/36	53/47	63/53	46/59	46/59	43/98		22	40	94	38		
n = 25	304	22/27	47/44	19/18	31/42	•21/27	243	•21/27	44/46	20/17	•29/43	246	23/25	46/45	27/27	•33/46	•33/46	43/98		22	40	94	38		
n = 50		16	37	13	23	15	36	15	36	14	21	21	15	36	17	23									
		$Rqr = 1.2-3.8$						$Rqr = 0.84-2.1$						$Rqr = 0.60-1.4$						$Rqr = 0.41-0.92$					
n = 10	134	55/75	65/61	•37/65	•77/130	62/67	172	62/67	73/63	46/62	•81/130	144	51/37	65/60	•130/48	•70/120	•70/120	•37/60		44/40	56/57	•240/68	•37/60		
n = 25	132	34/44	53/43	•25/42	•40/94	35/42	148	35/42	•57/44	27/32	•41/96	143	31/38	54/50	•77/38	38/58	38/58	•37/60		32	48	200	28		
n = 50		25	49	18	26	26	51	26	51	19	28	28	25	50	46	28									
		$Rqr > 3.8$						$Rqr > 2.1$						$Rqr = 1.4-3.7$						$Rqr > 0.92$					
n = 10																									
n = 25																									
n = 50																									
		$Rqr > 3.7$						$Rqr > 3.7$						$Rqr > 3.7$						$Rqr > 3.7$					
n = 10																									
n = 25																									
n = 50																									

The number before each slash is the percentage rmse from the simulation results, and the number after each slash is the percentage rmse from the verification results. There are no verification results for sample size $n = 50$. The value of N is the number of verification data sets falling in each class and from which the rmse was calculated. In the simulation, data were censored at percentiles of the parent distribution, and in the verification, data sets were censored at the sample percentiles.

*Significantly different at $\alpha = 0.05$.

Table B-4.1.2. Biases of LR and ML estimators, from simulations. (From Helsel and Gilliom)

Method	Censored at 20th Percentile				Censored at 40th Percentile				Censored at 60th Percentile				Censored at 80th Percentile			
	LR		LM		LR		LM		LR		LM		LR		LM	
	$\bar{x}$	s	m	iqr	$\bar{x}$	s	m	iqr	$\bar{x}$	s	m	iqr	$\bar{x}$	s	m	iqr
	<i>Rqr < 0.47</i>				<i>Rqr < 0.35</i>				<i>Rqr < 0.25</i>				<i>Rqr < 0.16</i>			
$n = 10$	-3	-34	-3	-13	-2	-39	-2	-13	4	-46	5	-15				
$n = 25$	0	-21	0	-10	-1	-22	-1	-4	-1	-28	-1	-5	9	-39	8	-14
$n = 50$	0	-17	0	-1	0	-17	-1	4	0	-23	-1	1	3	-32	2	-9
	<i>Rqr = 0.47-1.2</i>				<i>Rqr = 0.35-0.84</i>				<i>Rqr = 0.25-0.60</i>				<i>Rqr = 0.16-0.41</i>			
$n = 10$	-5	-23	-5	-6	-5	-23	-2	-5	1	-23	10	0				
$n = 25$	-1	-11	-2	-7	-1	-10	-1	-3	-3	-10	-1	1	3	-12	18	5
$n = 50$	0	-7	-2	-1	-1	-6	-1	1	-3	-3	-2	9	-3	-5	3	8
	<i>Rqr = 1.2-3.8</i>				<i>Rqr = 0.84-2.1</i>				<i>Rqr = 0.60-1.4</i>				<i>Rqr = 0.41-0.92</i>			
$n = 10$	-8	-29	-7	-4	-4	-29	0	-1	4	-25	24	7				
$n = 25$	-4	-18	-4	-9	-4	-20	-2	-9	-3	-17	2	-4	1	-9	43	8
$n = 50$	-2	-12	-3	-3	-2	-13	-2	-4	-2	-12	1	-2	-3	-7	19	5
	<i>Rqr > 3.8</i>				<i>Rqr > 2.1</i>				<i>Rqr = 1.4-3.7</i>				<i>Rqr > 0.92</i>			
$n = 10$	15	-16	2	35	23	-6	10	37	17	-13	56	30				
$n = 25$	7	-16	0	10	9	-12	1	10	4	-12	19	2	15	8	90	7
$n = 50$	5	-10	-3	6	7	-9	-1	7	4	-9	7	2	7	0	78	-2
									<i>Rqr > 3.7</i>							
$n = 10$									58	31	130	51				
$n = 25$									25	7	130	4				
$n = 50$									12	2	130	7				

with a detection limit of 5. There are $n = 14$ observations, of which 6 (36%) are censored and 9 are uncensored. The *rqr* is based on the interquartile range of the uncensored values. These uncensored values correspond to the quantiles 1/10, 2/10, ..., 9/10, with the denominator 10 equal to the number of uncensored values plus one. The 20th and 30th sample percentiles are both 6, while the 70th and 80th are 10 and 11. Therefore the IQR of the uncensored values is $(10 + 11)/2 - 6 = 4.5$. The *rqr* is this IQR divided by the detection limit, so $rqr = 4.5/5 = 0.90$. The LR estimate of the mean can be calculated by the SAS program listed in Appendix A, and is 7.56. (This differs slightly from Helsel and Gilliom's value of 7.42. Helsel and Gilliom do not describe their software.) From Tables B-4.1.1 and B-4.1.2, conservatively using the portions for $n = 10$ and 40% censoring, we obtain $rmse = 0.30$ and $b = -0.04$. Therefore, from the above equations, we have $\sigma_e^2 = 0.088$, $\sigma_y^2 = 0.092$, and $\mu_y = -0.087$. The 95% confidence interval is

$$\begin{aligned} 7.56\exp[0.087 - 1.96 \times 0.303] &\leq \text{mean} \leq 7.56\exp[0.087 + 1.96 \times 0.303] \\ 4.55 &\leq \text{mean} \leq 14.94 \end{aligned}$$

If interpolated values instead of conservative values had been taken from the tables, the interval would have been somewhat shorter.

To verify that their confidence intervals have approximately the correct coverage, Helsel and Gilliom simulated data sets from 16 populations (4 distributional forms with 4 Cvs), 4 censoring levels, and 3 sample sizes. From these censored simulated data sets, 95% confidence intervals were constructed, and the fractions of intervals that covered the estimated true parameters are shown in Table B-4.1.3. The coverage fractions range from 81% to 98%. Helsel and Gilliom comment that only for the standard deviation and the three lowest *rqr* classes is there a consistent tendency to underestimate the width of the confidence intervals.

The LR estimate of the standard deviation is 4.54. A naive LR confidence interval would use these estimates in the customary way to obtain $7.42 \pm (1.96 \times 4.54)/\sqrt{14}$, or (5.04, 9.94). The normal percentile 1.96 is used rather

Table B-4.1.3. Simulated Coverage Percentages of Confidence Intervals. (From Helsel and Gilliom)

Method	Censored at 20th Percentile				Censored at 40th Percentile				Censored at 60th Percentile				Censored at 80th Percentile			
	LR		LM		LR		LM		LR		LM		LR		LM	
	$\bar{x}$	s	m	iqr	$\bar{x}$	s	m	iqr	$\bar{x}$	s	m	iqr	$\bar{x}$	s	m	iqr
	<i>Rqr < 0.47</i>				<i>Rqr < 0.35</i>				<i>Rqr < 0.25</i>				<i>Rqr < 0.16</i>			
$n = 10$	95	89	95	91	95	87	92	92	94	81	98	91				
$n = 25$	96	87	81	93	95	89	95	95	93	89	94	94	95	85	95	92
$n = 50$	95	82	98	94	94	85	95	96	95	85	94	95	94	87	99	94
	<i>Rqr = 0.47-1.2</i>				<i>Rqr = 0.35-0.84</i>				<i>Rqr = 0.25-0.60</i>				<i>Rqr = 0.16-0.41</i>			
$n = 10$	93	88	93	92	93	86	94	92	94	86	99	93				
$n = 25$	93	94	93	94	94	92	94	95	95	92	94	95	95	91	99	95
$n = 50$	95	94	94	95	95	94	95	95	94	94	95	95	94	93	99	96
	<i>Rqr = 1.2-3.8</i>				<i>Rqr = 0.84-2.1</i>				<i>Rqr = 0.60-1.4</i>				<i>Rqr = 0.41-0.92</i>			
$n = 10$	93	88	92	93	93	88	94	94	95	89	98	95				
$n = 25$	91	93	94	93	94	90	95	93	95	89	94	96	94	91	96	95
$n = 50$	94	93	93	95	94	93	95	94	93	91	94	95	93	92	97	95
	<i>Rqr > 3.8</i>				<i>Rqr > 2.1</i>				<i>Rqr = 1.4-3.7</i>				<i>Rqr > 0.92</i>			
$n = 10$	98	96	93	97	98	98	95	98	97	98	98	96				
$n = 25$	96	95	94	96	96	97	93	95	95	97	96	95	96	98	98	94
$n = 50$	95	96	95	95	95	96	94	95	95	97	96	94	95	97	98	94
									<i>Rqr > 3.7</i>							
$n = 10$									98	97	99	99				
$n = 25$									97	97	98	92				
$n = 50$									96	97	97	88				

$\alpha = 0.05$; therefore percentages should equal 95.

than a Student's t percentile, because the data are not believed to be normal. The interval obtained in this way is substantially shorter than the Helsel-Gilliom interval.

Simulations were performed for this report to see if a simpler confidence interval of this naive form could be used. The simulations modeled all of Gilliom and Helsel's CVs, censoring levels, and sample sizes, and two of their four distributions (lognormal and gamma). In the cases that had at most 40% censoring and $CV \leq 1$, the naive one-sided 95% confidence upper limit

$$(\text{estimated mean}) + (1.645 * \text{est'd s.d.}) / \sqrt{n}$$

fell below the true mean from 6.3% to 11.6% of the time, more than the nominal 5%. This shows that the naive interval is too short at the upper end. If all four of Gilliom and Helsel's distributional forms had been simulated, the naive interval might have appeared worse. For comparison, note in Table B-4.1.3, when there was at most 40% censoring, that Helsel and Gilliom 95% confidence intervals covered the true mean for 93% to 98% of the cases. Therefore, the Helsel-Gilliom intervals seem preferable, at least for the population mean.

B-4.1.2 Maximum Likelihood Methods

First consider estimation of moments. Maximum Likelihood Estimation (MLE) methods work reasonably well for larger sample sizes and smaller variances. They work badly for small sample sizes and large variances when the goal is to estimate moments. It is likely that in this case there will sometimes be estimators of σ^2 that have infinite variance (this is known to be true in the case of zero censoring — see Shimizu, 1988, p. 38). Maximum likelihood estimation for censored data is easily implemented with existing survival analysis programs. These programs generally allow a number of distributional assumptions. For example, the LIFEREG procedure in SAS allows the Weibull, exponential, gamma, normal, lognormal, logistic, and loglogistic distributions.

For lognormal data, adjustments are available for MLE estimates for censored data that make them nearly unbiased minimum variance estimates. (Cohn, 1988) They are relatively difficult to implement, although FORTRAN programs for their calculation are available. Moreover, the bias of the MLE is only substantial when estimating moments of a strongly skewed distribution, in which case we do not recommend estimating moments at all. For these two reasons, the bias-adjusted MLE is not considered further here.

Now consider estimation of percentiles. Gilliom and Helsel showed that the MLE is the best estimator for the median and IQR. This may also be the case for other percentiles.

For estimating the median or IQR, Tables B-4.1.1 and B-4.1.2 allow construction of confidence intervals based on the MLE. For example, consider estimation of the median, based on the SAS calculations in Section A-3.1.2. The data, copied from that section, are

<4.1, <4.5, 4.5, 5.3, 5.3, 5.6, 7.7, 8.5, 9.5, 9.6, 11.0, 17.0, 17.0

There are 11 uncensored values; they are the 1/12, 2/12, ..., 11/12 sample quantiles of the uncensored data. The IQR of the uncensored data is therefore the 9/12 sample quantile minus the 3/12 sample quantile, i.e. $11.0 - 5.3 = 5.7$. The (average) detection limit is 4.3, so the *rqr* is $5.7/4.3 = 1.33$. The MLE of the median, shown in the output file in Section A-3.1.2, is 7.2016. From Tables B-4.1.1 and B-4.1.2, conservatively using the portions for $n = 10$ and 20% censoring, we obtain $rmse = 0.26$ and $b = -0.07$. Therefore, following the same steps as for the LR estimate of the mean, we have $\sigma_e^2 = 0.0627$, $\sigma_y^2 = 0.0700$, and $\mu_y = -0.1076$. The 95% confidence interval is

$$7.20\exp[0.1076 - 1.96 \times 0.2646] \leq \text{median} \leq 7.20\exp[0.1076 + 1.96 \times 0.2646]$$

$$4.78 \leq \text{median} \leq 13.47$$

This is substantially wider than the interval (5.35214, 9.6902) given in the SAS output in Section A-3.1.2, based on the asymptotic distribution of the MLE. According to Table B-4.1.3, the coverage probability of the wide

interval is probably not too far from 95%; therefore, the wide interval should be used, not the asymptotic interval given in the SAS output.

B-4.1.3 The DL/2 Method

In this method, the LTD values are replaced by half the detection limit. Then the mean and standard deviation are estimated by the sample mean and sample standard deviation. Suppose first that the true CV is small. As mentioned in the body of this report, the DL/2 method is severely biased in this case. The resulting confidence intervals for the mean, of the form

$$\bar{X} \pm z_{1-\alpha/2} s / \sqrt{n} ,$$

are shifted to the left, and therefore cover the true mean too infrequently. As n increases, the coverage frequency decreases.

In the body of this report, it was suggested that the DL/2 point estimates could be used when the estimated CV is 1.0 or larger. Consider now only data sets satisfying this restriction. The DL/2 confidence intervals are wider (and therefore better) than the naive LR confidence intervals, but apparently not as good as the Helsel-Gilliom intervals. Based on simulations from the gamma and lognormal distributions, the DL/2 one-sided 95% confidence upper limit tends to be too large when the true CV equals 1.0, and tends to be substantially too small when the true CV equals 2.0. Therefore, the DL/2 method is not recommended for confidence intervals or tests of hypotheses.

B-5 TWO SAMPLES

It frequently occurs in environmental sampling that two samples are independently collected and compared to see if both samples are from the same population. The most common example is comparison to background. There are a number of ways of making the comparison in the case when data are uncensored. The most commonly used methods involve comparing a measure of location from each sample. In the parametric case, the location measure is typically the mean. In the nonparametric case, it is the median.

Five types of method were evaluated for their ability to test for equality of means (medians for nonparametric tests) in the presence of censored data. Three of the techniques, the detection limit test (DL test), the Wilcoxon rank sum test (also known as the Mann-Whitney-Wilcoxon test, abbreviated MWW), and the Prentice test, are nonparametric techniques. The use of the latter two tests in the presence of censored data is described in Millard and Deveral (1988). The fourth technique involves replacing the LTD values with half the detection limit and computing a simple t -test. The fifth type of technique we call Iterative Least Squares (ILS); it encompasses at least three different techniques, each using an iterative algorithm based on least squares regression.

The MWW test and the DL/2 test both perform well and have approximately the same power to detect differences. These tests are also capable of handling multiple detection limits and comparisons of more than two groups (we only consider two groups here because it is our experience that it is the most common situation in environmental practice). The ILS methods are difficult to implement and do not maintain the Type I error rates as well as the other methods. The MWW test is widely available in most statistical packages and is easily modified to account for censoring when there is a single detection. It requires a little more work when there are multiple detection limits, but this should still be possible in most statistical packages. Hence, for its simplicity, availability, and performance, the MWW test is recommended for general use.

For estimation of parameters when σ equals 1, the DL/2 estimates have larger bias but smaller MSE than the MLEs obtained by the ILS technique.

B-5.1 TWO-SAMPLE TESTING: NONPARAMETRIC METHODS

Simulations were run on three different nonparametric 2-sample comparison tests. The three tests are 1) the Wilcoxon test, 2) the Prentice test, and 3) the DL test. A fourth test, the Runs test, was discontinued due to a poor showing in the early rounds (and to make more efficient use of the

computer).

The Wilcoxon test (also called the Mann-Whitney rank sum test) ranks all the data and then sums the ranks within each sample. The greater the difference between the sums of the ranks, the more likely the two samples have different means or medians. This test will be denoted the MWW test for the remainder of this report. Details of the test can be found in Hollander and Wolfe (1973). As applied to censored data, the test is discussed in Millard and Deveral (1988). For this simulation, a normal approximation was used to compute probabilities (see Steel and Torrie, 1980).

The Prentice test is based on computation of a linear combination of the ranks of the data. The choice of coefficients used in the linear equation is selected based on the distribution of the data in order to maximize the power of the test. In this sense, the technique is not truly nonparametric. The test is related to the Wilcoxon test. Details can be found in Millard and Deveral (1988). For this simulation, normal scores (see Table 5 of Millard and Deveral, 1988), and the permutation variance (see equation 10 in Millard and Deveral) were used, as was a normal approximation to compute probabilities.

The DL test (detection limit test) is essentially the same as the median test, using the detection limit as the break point rather than the median. For more detail on the median test, see Steel and Torrie (1980). For this simulation, the exact hypergeometric probability was computed.

B-5.1.1 Simulation Methods

A sample of size 20 was selected from a pseudo-standard normal distribution (with mean 0 and variance 1). The first ten pseudo-random numbers were assigned to the first group. The latter ten had their values shifted by addition of the mean, which took values 0, 0.5, 1.0, and 1.5, and were assigned to the second group. 500 samples were so constructed.

Each sample was censored at the 20, 40, 60, and 80% point of the joint density function of the two groups. For each censoring level, a single censoring point was calculated. Any observation falling below that point was censored. If all the observations fell below the censoring level, then that sample was not analyzed. This occurred about 1% of the time for the 80% censoring level, and occurred once at the 60% censoring level.

At each censoring level, the three test statistics and probabilities of a larger value were calculated. The null hypothesis was of equal medians, the alternative hypothesis was that the second group median was larger than the first. Hence, all tests were against a one-sided alternative. This one-sided alternative was used to mimic the common practice in environmental monitoring of comparing contaminated areas to background. In these circumstances, it is generally not of interest to test if the

concentration in contaminated areas is less than background.

All sample generation and calculations were performed using S/PC (Technology Partners, Inc., 1988). The computing was done on an IBM PS/2 Model 70.

B-5.1.2 Discussion

Simulation results are shown in Table B-5.1.1. As can be seen, the DL test performed poorly compared to the other two tests in terms of both power and maintained α -level (i.e. 5% rejected at $\mu_2 = 0$). The Prentice test performed slightly better than the MWW test in terms of power, but not quite as well as the MWW test in terms of maintained α -level. Generally, for the MWW test and the Prentice test, as the censoring level increased, the power decreased. For the DL test, power was consistently highest at the 60% censoring level. As would be expected, the power of the tests increased as the difference between the means of the two groups increased.

Simulation results are only reported for samples of size 10 for each group. Some simulation was done with samples of size 20. The results from these latter simulations showed a significant increase in the power of the tests, particularly the DL test and for all 3 tests when the differences between means was small (0.5).

The poor performance of the DL test with respect to the other two tests is not surprising since it uses less information contained in the sample. However, it was hoped that at high censoring levels, the DL test would compare favorably to the other two tests. This was not the case. Specifically, when the mean of the second group was 0.5 or 1.0, the DL test was approximately half as powerful as the other two tests at the 80% censoring level.

Because of symmetry in the hypergeometric statistic on which the DL test is based, it was expected that the power of the DL test would be symmetric about 50% censoring. However, the results show that the power is slightly higher for censoring levels above 50% than for the corresponding censoring level below 50%. The DL test was also poor at maintaining the significance level, as can be seen from the results when the two group means are equal. This makes this test a conservative test. Based on these results, there is no reason to recommend the DL test over the remaining two tests.

The Prentice test was consistently more powerful than the MWW test. However, since the samples were from normal distributions, the simulation was slanted in favor of the Prentice test based on normal scores. From this simulation, it is unclear whether the Prentice test would have still been more powerful than the MWW test if the data had come from a non-normal distribution. Further, these results are based on only 500 simulations. When many more simulations were run, shown in Section B-5.3, the Prentice test had a Type I error less than 5%, resulting in lower power than MWW

Table B-5.1.1. Nonparametric 2-Sample Test Simulation Results

n	μ_2	% Censored	% Rejected ^a		
			Wilcoxon	Prentice	DL test
10	0	20	4.6	5.4	2.0
		40	4.6	6.4	1.6
		60	6.4	7.2	2.6
		80	5.9	6.5	1.2
10	0.5	20	30.4	33.2	6.8
		40	28.4	33.8	12.8
		60	26.1	30.0	13.6
		80	17.2	21.2	8.5
10	1.0	20	65.6	69.2	18.2
		40	62.8	67.2	35.2
		60	60.6	65.2	37.4
		80	46.7	51.1	24.0
10	1.5	20	93.8	94.7	40.0
		40	92.7	93.8	69.1
		60	86.5	90.4	70.2
		80	64.1	66.9	42.7

a. $\alpha = 0.05$, one-sided alternative.

when β_1 was non-zero, especially at the 60% censoring level. Moreover, MWW does not depend on the normality of the distribution, while the Prentice test does. Therefore, although the reasons are not all apparent from Table B-5.1.1, we recommend the MWW over the Prentice test, and both over the DL test. Another reason to recommend either the MWW or Prentice test over the DL test is that they have been modified to allow for multiple detection limits, as discussed in Millard and Deveral (1988). It is common in chemical analytical data to have multiple detection limits due, for example, to dilution of the original sample. However, few statistical techniques allow for multiple detection limits. Millard and Deveral (1988) give some simulation results of these methods in the presence of multiple detection limits and get mixed results, dependent on the parent distribution.

B-5.2 TWO SAMPLE TESTING: ITERATIVE LEAST SQUARES

In this section, a method is given for comparing two samples by using a linear model and least squares analysis in the presence of censored data. The method involves iterations and has been called iterative least squares (ILS) by Schmee and Hahn (1979). It resembles the widely used EM algorithm (Dempster, Laird, and Rubin, 1979) used for obtaining the MLE. The EM algorithm, however, replaces the sufficient statistics (including $\sum y_i^2$) by their expectations, whereas the ILS method replaces individual missing observations by their expectations. The Schmee and Hahn ILS algorithm therefore is not the EM algorithm, and in fact it does not yield the MLE. A related algorithm by Wolynetz (1979b) is claimed to be an implementation of the EM algorithm. Yet it also is not, because it does not consider replacement of the sufficient statistics. Instead, it replaces individual missing observations, and then modifies the normal equations to obtain the updated estimates of the parameters. We use the Wolynetz algorithm, and have accepted the claim that it indeed yields the MLE, but we have not tested this claim. In this discussion, we use the term ILS to refer to all these iterative methods.

The model that is used to describe the data is

$$y_{ij} = \mu + \beta_i + \epsilon_{ij} \quad (\text{B5.2-1})$$

where $i = 1, 2$ is a group indicator, $j = 1(1)n_i$, and n_i is the number of observations in the i 'th group. In this parameterization, the β_i is the parameter of most interest as it represents the group effect (often called the treatment effect in the standard terminology of experimental design). If there is a significant difference between the β_i , then there is evidence that the two samples are not from the same population.

B-5.2.1 Specification Of The Method

Consider the model

$$y = \mu + \beta_1 x_1 + \epsilon \quad (\text{B5.2-2})$$

where y is the response variable, x_1 is the grouping variable, ϵ is the random error term (residual), and μ and β_1 are the parameters to be estimated. It is assumed that the expected value of ϵ is 0 and the variance is σ^2 . We first review the regression calculations for the uncensored case.

If $n = n_1 + n_2$ independent observations of the response are taken for n realizations of the set of xs , then for each observation the model is

$$y_i = \mu + \beta_1 x_{1i} + \epsilon_i, \quad i = 1(1)n. \quad (\text{B5.2-3})$$

In this model, the y_i and x_{1i} are known, while μ , β_1 and the ϵ_i s are unknown and must be estimated. The model (B5.2-3) with n observations can be rewritten in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (\text{B5.2-4})$$

where $\mathbf{y}$, the vector of response values, and $\boldsymbol{\epsilon}$, the vector of residuals, are $n \times 1$ vectors, $\mathbf{X}$ is an $n \times 2$ matrix with the first column being a column of 1s and the second column being an indicator (grouping) vector of 0s and 1s depending on whether the corresponding observation is taken from group 1 or 2, respectively. $\boldsymbol{\beta} = (\mu, \beta_1)^T$ is the vector of coefficients. The expected value of $\boldsymbol{\epsilon}$ is $\mathbf{0}$ with dispersion matrix $\sigma^2 \mathbf{I}$.

The vector $\mathbf{y}$ and matrix $\mathbf{X}$ in Eqn. (B5.2-4) are known from the n observations. The estimate of $\boldsymbol{\beta}$ is found as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (\text{B5.2-5})$$

And the residuals are given by

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (\text{B5.2-6})$$

The variance of the residuals, σ^2 , is estimated as

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{N-2} . \quad (\text{B5.2-7})$$

Finally, an important property of the estimate $\hat{\beta}$ is the variation in the estimation. This is contained in the estimated dispersion matrix, $\text{var}(\hat{\beta})$, which is calculated as

$$\text{var}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} . \quad (\text{B5.2-8})$$

The structure of $\text{var}(\hat{\beta})$ is

$$\text{var}(\hat{\beta}) = \begin{bmatrix} \text{var}(\hat{\mu}) & \text{cov}(\hat{\mu}, \hat{\beta}_1) \\ \text{cov}(\hat{\mu}, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) \end{bmatrix} .$$

A simple way of assessing whether there is a difference between groups is to test whether β_1 is zero. This can be done with a t -test by dividing the estimate of β_1 by its associated standard error. That is,

$$t = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} , \quad (\text{B5.2-9})$$

where $\hat{\beta}_1$ and $\text{var}(\hat{\beta}_1)$ are calculated as shown in Eqns. (B5.2-5) and (B5.2-8), respectively. The t -value has degrees of freedom $n - p - 1 = n - 2$, since p is the number of independent (grouping) variables. This calculated t -value is compared to the tabled t -value at the appropriate significance level.

Iterative Least Squares. The ILS algorithm essentially requires that one put in replacement values for the censored data points and compute $\hat{\beta}$ and $\hat{\sigma}^2$. With these estimates, one then finds new replacement values for the censored data. These new estimates of the censored points are then used to recompute $\hat{\beta}$ and $\hat{\sigma}^2$. The iterations proceed to convergence.

The initial replacement values for the censored data are typically taken as the censoring points (detection limits), DL_i . $\hat{\beta}$ is then computed by Equation (B5.2-5). The residual variance is then computed using a method analogous to Equation (B5.2-7), but that accounts for the censoring. We identified three different estimates for the residual variance in the presence of censoring in the literature. The first, due to Schmee and Hahn (1979) is to use Equation (B5.2-7) with the censored observations represented by their replacement value. The second method, identified by Aitkin (1981), is to calculate the residual variance as

$$\hat{\sigma}^2 = \frac{\sum_{i \in u} (y_i - \hat{\mu}_i)^2}{\left(n - c - \sum_{i \in c} \hat{z}_i S(\hat{z}_i) \right)} \quad (\text{B5.2-10})$$

where

$$\begin{aligned} i \in u &= \text{all uncensored observations,} \\ i \in c &= \text{all censored observations,} \\ n - c &= \text{the number of uncensored observations,} \\ \hat{\mu}_i &= \hat{\mu} + \hat{\beta}_1 x_{1i}, \\ \hat{z}_i &= (\hat{y}_i - \hat{\mu}_i)/\hat{\sigma}, \\ \hat{y}_i &= \hat{\mu}_i - \hat{\sigma} S((y_i - \hat{\mu}_i)/\hat{\sigma}), \\ S(v) &= \phi(v)/[1 - \Phi(v)] . \end{aligned}$$

Here Φ is the standard normal cumulative distribution function and ϕ is the corresponding density.

The third method for estimating the residual variance is due to Wolynetz (1979b). In this case,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \hat{\mu}_i)^2}{n - c + \sum_{i \in c} T(-\hat{H}_i)} \quad (\text{B5.2-11})$$

where

$$\begin{aligned} \hat{y}_i &= \begin{cases} y_i & \text{if } i \in u, \\ \hat{\mu}_i - \hat{\sigma} S(-\hat{H}_i) & \text{if } i \in c, \end{cases} \\ \hat{H}_i &= (DL_i - \hat{\mu}_i)/\hat{\sigma}, \\ T(v) &= S(v)\{S(v) - v\}, \end{aligned}$$

and all other symbols are as above. We also considered the denominator in Equation (B5.2-11) with an adjustment for the number of parameters in the model. Note that the definition of $\hat{y}_i$ in Equation (B5.2-11) differs from that defined in Equation (B5.2-10). The form of the calculation of the replacement values $\hat{y}_i$ given in (B5.2-11) was used regardless of the method for estimating $\hat{\sigma}^2$, since it was not clear in Aitkin's paper what the value of y_i for censored data is in the equation for $\hat{y}_i$.

Finally, two different estimates of $\text{Var}(\hat{\beta})$ were used. The first estimate is given in Equation (B5.2-8), with $\hat{\sigma}^2$ calculated using one of the techniques given above. The degrees of freedom associated with the t -test given in Equation (B5.2-9) was considered to be the denominator of the estimate of $\hat{\sigma}^2$ rounded to the nearest integer.

The second method uses the asymptotic estimate of $\text{Var}(\hat{\beta})$, due to Wolynetz (1979b). The asymptotic information matrix, $G(\beta)$, is first estimated. Then $\text{Var}(\hat{\beta})$ is estimated as the inverse of

the estimated information matrix. The elements of the information matrix, g_{jk} , are given by

$$g_{jk} = \hat{\sigma}^2 \left\{ \sum_{i \in u} x_{ij} x_{ik} + \sum_{i \in c} x_{ij} x_{ik} T(-\hat{H}_i) \right\}, \quad (\text{B5.2-12})$$

where x_{ij} represents the element in the matrix $\mathbf{X}$ corresponding to the i th row and j th column. When using the asymptotic estimate of $\text{Var}(\hat{\beta})$ for testing $\beta_1 = 0$, the degrees of freedom were assumed to be ∞ , and hence the standard normal distribution was used rather than the t -distribution.

B-5.2.2 Simulations and Computing Methods

Because of concerns expressed in Schmee and Hahn (1979) and Aitkin (1981) with respect to bias in the estimate of the residual variance, simulations were implemented to see if it was possible to come up with an unbiased estimate of the variance. There was also a need to look at Wolynetz's (1979a, 1979b) estimators for the variance, as these had not been considered in previous simulations evaluating performance.

The simulations were done in three parts. The first part was used to determine if any methods could be "thrown out" immediately as they did not perform well versus the other methods. The methods considered involved different estimates of the variance and were due to Aitkin (1981), Schmee and Hahn (1979), and Wolynetz (1979b). Under the latter estimates we adjusted the degrees of freedom for the model parameters being estimated. The first set of simulations involved up to 500 runs.

The second set of simulations looked only at Wolynetz's (1979b) method for estimating the variance with a variety of degrees of freedom adjustments. The general form of the variance is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \hat{\mu}_i)^2}{n - c + \sum_{i \in c} T(-\hat{H}_i) - k}, \quad (\text{B5.2-13})$$

where $k = 0(1)3$ is the adjustment factor studied. Equation (B5.2-13) is identical to Equation (B5.2-11) when $k = 0$. The hypothesis test of $H_0: \beta_1 = 0$ was tested using the non-asymptotic t -test shown in Equation (B5.2-9). This set of simulations used 1,000 runs.

The third set of simulations extended the second set and included two testing methods: the non-asymptotic t -test of the hypothesis $\beta_1 = 0$ and the asymptotic version of the same test using the standard error of β_1 as given by Wolynetz (1979b). The variance shown in Equation (B5.2-13) was used with two different values of k , 0 and 2. Thus, the third set of simulations actually encompassed four methods: non-asymptotic test with the variance estimator's denominator adjusted by $k = 0$,

non-asymptotic test with $k = 2$, asymptotic test with $k = 0$, and asymptotic test with $k = 2$. The exact form of the non-asymptotic test is given in Equation (B5.2-9), where the resulting t -value is compared to a t -distribution with degrees of freedom corresponding the nearest integer value of the denominator of the residual variance. The exact form of the asymptotic test is

$$z = \frac{\hat{\beta}_1}{\sqrt{G(\hat{\beta})_{22}^{-1}}} , \quad (\text{B5.2-14})$$

where $G(\hat{\beta})_{22}^{-1}$ is the second diagonal element of the inverse of the information matrix, $G(\hat{\beta})$. Calculation of the elements of $G(\hat{\beta})$ is shown in Equation (B5.2-12). The third set of simulations used 5,000 runs.

The first set of simulations will not be discussed in detail here. They were run using an S/PC program, which proved to be very time consuming. It immediately picked up problems with Aitkin's method, as the denominator of the variance estimate frequently went below zero for 60 and particularly 80 percent censoring. Schmee and Hahn's method showed significant bias when estimating the residual standard deviation.

Simulation Settings. Sampling was from two normal populations, corresponding to two groups. The first population had mean 0 and variance 1. The second population had its mean varied at 0, 0.5, and 1.0 and its variance constant at 1. The censoring was done at 20, 40, 60, and 80 percent levels of the joint population density function. That is, the censor points (i.e. detection limits in environmental data) were determined such that when considering both populations as one, 20, 40, 60, and 80 percent of the data would be censored. Thus only one detection limit was assumed to be in effect. This may not be realistic as most environmental data sets have multiple censoring levels (note, though, that these techniques can all handle multiple censoring levels). Also two sample sizes were considered. In the first, 10 observations were taken from each group. 20 observations were taken from each group in another set of simulations. A simulation was dropped when there were less than 2 uncensored observations.

Computing Methods. Apart from the first set of simulations, which used S/PC, the simulations were run using a Pascal program compiled in TurboPascal. The regression algorithm is due to Givens (Kennedy and Gentle, 1980, p. 273). The random number generator has two parts. The first part used a linear congruential generator (LCG) that produced uniform random numbers between zero and one. The LCG coefficients had previously been tested and found to be adequate. The uniform pseudo-random numbers were then transformed to normal random variates using the Box-Muller method.

For each setting of β_1 (i.e. 0, 0.5, and 1.0), up to 5000 data sets were produced. Each data set was then censored at the 20, 40, 60, and 80 percent levels. Depending on the set of simulations, there were up to 5 estimates of the residual variance used on each data set and censoring level. In the third set of simulations, there were also two tests of $\beta_1 = 0$.

There were two convergence criteria, both of which had to be met. First, the square of the difference between two successive estimates of the residual variance had to be less than 0.00001. Second the sum of the square of the difference between two successive estimates of μ and the square of the difference between two successive estimates of β_1 had to be less than 0.00001. If the program did not converge in 100 iterations, the iterations stopped and the most current estimates were used. This was only a problem at the 80% censoring level. Since a large fraction of the 80% results are based on simulations that were not allowed to converge, they may be suspect. However, it is also typically true that by the end of 100 iterations the estimates are stable (i.e. there are no gross differences between the estimates from one iteration to the next).

B-5.2.3 Results

The methods were evaluated for testing for a difference between two groups. They were primarily judged based on their ability to maintain the α -level when $\beta_1 = 0$ and their power when $\beta_1 > 0$. The α -level was set at 5%. It controls the Type I error rate; i.e. it establishes the allowable frequency of rejecting the null hypothesis that $\beta_1 = 0$, when in fact the null hypothesis is true. When the null hypothesis is false ($\beta_1 > 0$), the Type II error rate measures the frequency of incorrectly accepting the null hypothesis. The methods were also compared for properties of the parameter estimators, as discussed in the text below but not tabulated.

The first set of simulations (which used up to 500 runs) involved comparing the performance of ILS when using Schmee and Hahn's variance estimate, Aitkin's variance estimate, and Wolynetz's estimate. These simulations showed Wolynetz's method to be the most stable, that is, the estimator had the smallest variance. Aitkin's estimate produced negative variances at high censoring levels. Schmee and Hahn's method was not stable at high censoring levels, but performed well at low censoring levels. The Wolynetz estimate was found consistently to underestimate the residual variance. Since Wolynetz makes no adjustment for the number of parameters estimated, it was decided to look at adjusting the denominator of the residual variance in the Wolynetz estimate. It was hoped that this might remove some of the bias in the estimate.

In the second set of simulations (which used 1,000 simulations) Wolynetz's residual variance estimator was adjusted by subtracting 0, 1, 2, and 3 from the denominator. It is important to note

that this adjustment was made at each step of the iterative algorithm, not merely at the end. This affected the estimates of all the parameters. Thus, only the 0-adjustment estimates are claimed to be MLEs. The results of the simulations in terms of % Rejected are shown in Table B-5.2.1. None of the adjustments were able to maintain a 5% alpha level at the 60 and 80% censoring levels. However, when the denominator was adjusted by 3, the Type I error rate remained below 8%; when adjusted by 2, the Type I error rate was less than 9%. On the other hand, and not surprisingly, the method using no adjustment was consistently the most powerful. Since these methods were using a t -test of $H_0: \beta_1 = 0$, it was not truly clear what the degrees of freedom should be. We used the denominator of the residual variance rounded to the nearest integer.

The third set of simulations (which used 5,000 simulations) used only Wolynetz estimate with either no adjustment or an adjustment of 2. The purpose of the third set of simulations was to determine whether the asymptotic test of $H_0: \beta_1 = 0$ performed better than the ad-hoc t -test. The results are presented in Table B-5.2.2. The non-asymptotic results resemble the corresponding results in Table B-5.2.1, since they are the same methods, but use different random number generator seeds. The asymptotic test was more conservative in that it was less likely to reject the null hypothesis. Its power is sharply reduced at the 60% censoring level compared to the non-asymptotic test. On the other hand, the asymptotic test did not tend to inflate the α -level as does the non-asymptotic test; indeed, the α -level falls off at 60 and 80% censoring in the asymptotic test.

Although not presented in detail here, the method that provided the most unbiased estimates and smallest standard errors for μ and β_1 was the Wolynetz estimate with no adjustment. This method also provided the best estimate of σ^2 at the higher censoring levels. At low censoring levels, the adjustment of 2 to the denominator appeared to work best at estimating σ^2 .

However, at low censoring levels, virtually all the Wolynetz estimates and both the asymptotic and non-asymptotic testing methods work well. It is only at high censoring levels that any differences in performance become apparent. Thus selecting the "best" method will depend upon the objectives of the researcher. Certainly if one is only interested in estimating the model parameters μ and β_1 it is best to use no adjustment on the residual variance. For testing the hypothesis of no group effect, the recommendations are less clear. If one was most interested in controlling the Type I error rate, it is probably best to use the asymptotic test with no adjustment to the residual variance. The non-asymptotic test has a smaller Type II error, at the price of a larger Type I error.

Table B-5.2.1. Summary of testing results for the second set of simulations. The percent of runs in which the hypothesis $H_0: \beta_1 = 0$ was rejected is shown in the body of the table. The last four columns correspond to the residual variance estimate used, as shown in Equation (B5.2.13). The results are based on 1,000 runs.

n	β_1	% Censored	% Rejected ^a			
			$k = 0$	$k = 1$	$k = 2$	$k = 3$
10	0	0 ^b	5.6	5.6	5.6	5.6
		20	7.3	6.9	5.6	5.1
		40	8.5	6.2	6.0	5.6
		60	10.5	8.1	7.7	7.8
		80	11.6	11.0	8.8	7.1
10	0.5	0	28.2	28.2	28.2	28.2
		20	30.4	28.8	27.1	25.1
		40	30.2	28.6	27.0	25.1
		60	31.0	29.1	26.2	24.1
		80	33.5	30.7	28.0	26.3
10	1.0	0	68.2	68.2	68.2	68.2
		20	71.4	69.6	66.8	66.5
		40	69.9	65.9	63.4	61.8
		60	70.1	65.2	62.5	61.6
		80	64.8	65.7	61.2	57.5

^a. $\alpha = 0.05$, one-sided alternative.
^b. At 0% censored, the residual variance used $n - p - 1$ degrees of freedom.

Table B-5.2.2. Summary of testing results for the third set of simulations. The percent of runs in which the hypothesis $H_0: \beta_1 = 0$ was rejected is shown in the body of the table. The k value refers to the residual variance adjustment as shown in Equation (B5.2-13). Non-asymptotic results are based on a t -test. Asymptotic results are based on a z -test. The results are based on 5,000 runs.

n	β_1	% Rejected ^a				
		% Censored	Non-asymptotic		Asymptotic	
			$k=0$	$k=2$	$k=0$	$k=2$
10	0	0 ^b	5.1	5.1	5.1	5.1
		20	6.4	5.2	6.4	5.5
		40	6.9	5.8	6.2	4.9
		60	8.6	6.9	5.3	3.7
		80	11.5	8.3	1.6	0.8
10	0.5	0	28.1	28.1	28.1	28.1
		20	31.5	27.8	32.7	29.0
		60	33.9	29.4	22.3	17.6
10	1.0	0	70.3	70.3	70.3	70.3
		20	73.3	69.5	74.6	71.0
		60	70.4	66.0	52.6	45.8
20	0	0	4.9	4.9	4.9	4.9
		20	5.6	5.1	5.6	5.1
		40	6.2	5.6	5.4	5.0
		60	8.1	7.5	5.2	4.3
		80	12.2	11.2	3.4	2.3
20	0.5	0	46.9	46.9	46.9	46.9
		20	48.5	46.6	48.5	46.6
		60	49.5	47.5	39.1	36.3
20	1.0	0	92.9	92.9	92.9	92.9
		20	93.2	92.4	93.2	92.4
		60	90.2	89.3	84.5	82.9

a. $\alpha = 0.05$, one-sided alternative.
b. At 0% censored, the residual variance used $n - p - 1$ degrees of freedom.

B-5.3 TWO-SAMPLE TESTING: DL/2

A final set of simulations was performed to assess the effectiveness of the replacement of LTD values with half the detection limit. The DL/2 had performed reasonably well in early ANOVA simulations, and is possibly the most commonly used method for handling LTD data by environmental scientists. Its performance in the two-sample situation was compared to that of the unadjusted and adjusted Wolynetz EM method, and the nonparametric methods of Mann-Whitney-Wilcoxon and Prentice.

It was assumed that the data come from a lognormal distribution. The data analyzed have been log-transformed to the normal distribution. The model is that given by Equation (B5.2-2) and discussed in Section B-5.2.1. The replacement of censored values in the lognormal scale by half the detection limit corresponds to replacement of the censored values with $\ln(DL) - \ln(2)$ on the normal scale.

The test of differences between the group means on the normal scale corresponds to a test for differences in the group medians in the lognormal scale. The assumption of a common normal variance corresponds to the assumption of a common CV for the two lognormal distributions. The testing procedure is the same as that discussed in Section B-5.2.1; a simple t -test of the null hypothesis $\beta = 0$ is implemented using Equation (B-5.2-9), with the censored values replaced by $\ln(DL) - \ln(2)$. All tests were performed with a 0.05 Type I error rate.

B-5.3.1 Simulation Settings

Sampling was from two normal populations, corresponding to two groups. The first population had mean 0 and variance 1. The second population had its mean varied at 0, 0.5, and 1.0, with a constant variance of 1. These settings correspond to model values of $\mu = \beta_0 = 0$ and $\beta_1 = 0, 0.5, \text{ or } 1.0$. The censoring was done at the 20% and 60% levels for the overall population density. Only a sample size of 10 per group was considered. For each censoring level and population mean difference, 5000 data sets were generated. To more precisely estimate the Type I error rate under the null distribution with $\beta_1 = 0$, the number of simulations was extended to 47,500. We also extended the simulations of the MWW and Prentice tests to 47,500 to get a more precise estimate of the Type I error rate under the null distribution.

B-5.3.2 Computing Methods

All simulations were done in S-Plus (Statistical Sciences, Inc., 1988). The samples were generated using the `nrnd()` function. The `Two.t.test()` function was used to perform the calculations

for the DL/2 method. The MWW and Prentice methods were programmed using the S-Plus language. The analyses were run on a DECstation 3100.

B-5.3.3 Results

The results are shown in Table B-5.3.1. This table summarizes the DL/2 simulations as well as some of the earlier simulations, for easy comparison of the results. The standard error of the power (or Type I error when $\beta_1 = 0$) is shown in parentheses. The quantity labeled SE is the standard error of the tabulated bias. The two EM methods yield the same estimates, the MLEs, although the properties of the estimates are only shown once. The tests of $\beta_1 = 0$ differ in that EM-N uses the simple formula (B5.2-8) for the estimate of $\hat{\beta}_1$, and a tabulated t -value, while EM-A uses the asymptotic variance of $\hat{\beta}_1$ based on Expression (B5.2-12), and a tabulated z -value. The denominator of Expression (B5.2-11) is used for the degrees of freedom in the t -test.

The DL/2 method maintains the Type I error rate under the null distribution at 20% censoring, and very nearly maintains it at 60% censoring. The bias in the parameter estimates increases in magnitude with an increase in the censoring. At 60% censoring, the DL/2 estimate of β_0 tends to be biased high, the estimate of β_1 is biased low when the true value is non-zero, and the estimate of σ^2 is biased low. At 20% censoring, the DL/2 method introduces little bias in the parameter estimates. Most of the MSE in the parameter estimates is due to variance rather than to bias, the exception being estimation of β_0 at 60% censoring.

A perusal of Table B-5.3.1 shows that MWW is best at maintaining Type I error rate under the null distribution, followed closely by DL/2. This comparison has only been made when σ equals 1.0. Because MWW is nonparametric, it should be independent of the value of σ . The DL/2 estimators are dependent on σ , but it was found in Section B-6 that the DL/2 test maintains the Type I error of 0.05 quite well in an ANOVA setting, for σ equal to 1.0 or 0.25. The apparent reason is that under the null distributions the DL/2 biases all the cell means equally, and therefore maintains the correct Type I error for testing for differences of the cell means. Therefore, the conclusions seen here may be valid for different values of σ as well.

For parameter estimation, the EM estimators with no adjustment (i.e., the MLEs) tend to have lower biases than the DL/2 estimators, but larger MSEs. These comparisons of estimators may be dependent on the particular value of σ used.

Table B-5.3.1. Summary of the two sample comparison results. DL/2 = replacement with one half the detection limit; EM-N = MLE using non-asymptotic results; EM-A = MLE using asymptotic results; MW = Mann-Whitney/Wilcoxon test; Prentice = Prentice test.

β_1	LTD	Method	N	Power(SE)	β_0			β_1			σ^2		
					Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.0	20	DL/2	5000	0.0492 ^a (0.0010)	-0.0233	0.0046	0.1055	-0.0087	0.0065	0.2093	0.0352	0.0043	0.0919
		EM-N	5000	0.0640(0.0035)	-0.0063	0.0046	0.1073	0.0028	0.0065	0.2095	-0.0806	0.0052	0.1422
		EM-A	5000	0.0642(0.0035)									
		MWW	47500	0.0509(0.0010)									
		Prentice	47500	0.0433(0.0009)									
0.0	60	DL/2	4981	0.0534 ^a (0.0010)	0.1233	0.0035	0.0749	-0.0020	0.0049	0.1113	-0.3986	0.0031	0.2053
		EM-N	4981	0.0861(0.0040)	-0.0130	0.0064	0.2060	-0.0026	0.0084	0.3492	-0.0604	0.0087	0.3776
		EM-A	4981	0.0532(0.0032)									
		MWW	47500	0.0505(0.0010)									
		Prentice	47500	0.0363(0.0009)									
0.5	20	DL/2	5000	0.2806(0.0064)	-0.0210	0.0045	0.1007	-0.0067	0.0065	0.2105	0.0164	0.0041	0.0863
		EM-N	5000	0.3148(0.0066)	-0.0124	0.0047	0.1120	0.0051	0.0066	0.2149	-0.0732	0.0054	0.1512
		EM-A	5000	0.3270(0.0066)									
		MWW	1000	0.2770(0.0142)									
		Prentice	1000	0.2630(0.0139)									
0.5	60	DL/2	4985	0.2518(0.0061)	0.2238	0.0031	0.0966	-0.1659	0.0050	0.1522	-0.4077	0.0030	0.2113
		EM	4990	0.3387(0.0067)	-0.0496	0.0085	0.3653	0.0369	0.0093	0.4360	-0.0540	0.0089	0.3963
		EM-A	4990	0.2320(0.0059)									
		MWW	1000	0.2100(0.0129)									
		Prentice	1000	0.1740(0.0120)									

Table B-5.3.1 (contd). Summary of the two sample comparison results. DL/2 = replacement with one half the detection limit; EM-N = MLE using non-asymptotic results; EM-A = MLE using asymptotic results; MW = Mann-Whitney/Wilcoxon test; Prentice = Prentice test.

β_1	%	LTD	Method	N	Power(SE)	β_0			β_1			σ^2		
						Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
1.0	20		DL/2	5000	0.6844(0.0066)	-0.0142	0.0043	0.0940	-0.0061	0.0064	0.2021	-0.0031	0.0043	0.0944
			EM-N	5000	0.7332(0.0063)	-0.0054	0.0049	0.1196	0.0109	0.0066	0.2184	-0.0847	0.0052	0.1433
			EM-A	5000	0.7458(0.0062)									
			MWW	1000	0.6780(0.0148)									
			Prentice	1000	0.6550(0.0150)									
1.0	60		DL/2	4990	0.5713(0.0070)	0.3653	0.0025	0.1652	-0.3528	0.0049	0.2438	-0.4162	0.0030	0.2183
			EM-N	4991	0.7035(0.0065)	-0.0951	0.0105	0.5541	0.0978	0.0111	0.6205	-0.0760	0.0082	0.3432
			EM-A	4991	0.5257(0.0071)									
			MWW	1000	0.5600(0.0157)									
			Prentice	1000	0.4920(0.0158)									

a. % Reject based on 47,500 observations.

B-6 LINEAR MODELS

In a standard linear model, we assume that the observations are independent and normally distributed with a constant variance, and that the means of the observations depend in some linear way on unknown parameters. That is, the observations Y_i are normal(μ_i, σ^2), where the μ_i 's depend on a small number of parameters. In the context of environmental measurements, suppose that the measurements X_i have a lognormal(μ_i, σ^2) distribution. That is, they all have the same coefficient of variation, and they have different medians $\exp(\mu_i)$. If the linear-model assumptions relating the μ_i 's are correct, then linear-model methods may be applied to analyze a data set of x_i 's by letting $y_i = \ln(x_i)$ and analyzing the y_i 's.

When trying to apply intuition from the one-sample setting to linear models, remember that much of the one-sample work dealt with moments, in particular with means and standard deviations of the lognormal measurements. In the linear model setting, on the other hand, the inference in terms of the original lognormal measurements concerns a percentile, the median, under the assumption of a common CV.

The detection limit DL for the original x 's corresponds to a limit $\ln(DL)$ for the y 's. Replacement of a censored x by $DL/2$ means that the corresponding y is replaced by $\ln(DL/2) = \ln(DL) - \ln 2$. Other replacement techniques for the x 's correspond similarly to replacement techniques for the y 's. The behavior of the $DL/2$ method will depend to some extent on where $\ln(DL) - \ln 2$ falls in the normal(μ, σ^2) distribution. This depends in part on the relative sizes of $\ln 2$ and σ .

The methods discussed in this section assume that the data are normally distributed, that is, that the log transformation has already been applied to the original data. The notation y is used for an observation of the normally distributed random variable.

B-6.1 SIMULATIONS ON TWO-WAY ANALYSIS OF VARIANCE WITH REPLACEMENTS FOR CENSORED VALUES

B-6.1.1 Introduction

This section discusses the methods and results of two-way analysis of variance (ANOVA) simulations performed on censored data sets, using several replacement techniques for the censored values. Different censoring levels and normal parameters are simulated. The approach is patterned after Gilliom and Helsel (1986), described in Section B-4.1.1, but is much more limited. In particular, only lognormal data (i.e. normal Y 's) are simulated, and only 20% and 60% censoring are considered. The sample sizes are distributed over four cells, and differ from Gilliom and Helsel's sizes of single samples.

The LR method of Gilliom and Helsel can only be applied on a cell-by-cell basis. This limits the method's applicability to cases in which no cell is fully censored. The reason why the method cannot be used with the entire pooled data set is simple: The method uses the uncensored measurements to assign values (all different) to the LTD measurements. In an ANOVA setting with LTDs in more than one cell, the LR method does not give a way to decide which values should be assigned to which cells. The calculated ANOVA statistics will depend on which way the assigned values are placed in the cells. Therefore, there is no straightforward way to use the LR method with the pooled data. The LR method can be applied separately to the data within each cell, but only if each cell has at least two observations. The extension in which the plots of normal scores are assumed to have a common slope but possibly different intercepts in the different cells has not been investigated.

B-6.1.2 Simulation Technique

Distributions. A two-factor fixed-effects model is used in the simulations. Each factor has two levels. Data are randomly generated to reflect different distributions per cell, or treatment combination. Each cell

distribution is normal and has standard deviation 1.0 or 0.25. In the one-sample work of Section B-4, the DL/2 method performed fairly well for estimating the lognormal mean when the lognormal CV was 1.0 or larger, corresponding to a normal $\sigma \geq .83$, and the DL/2 method was badly biased when the lognormal CV was 0.25, corresponding to a normal $\sigma = 0.25$. Therefore, in this study of the ANOVA setting, the normal distributions are assumed to have either $\sigma = 1.0$ or $\sigma = 0.25$. To assess the Type I error and the power, we consider both the null distribution and one distribution designed to produce significant main effects. Table B-6.1.1 shows cell means and coefficients of variation for the simulations performed.

Table B-6.1.1 Treatment combination means considered for normal distributions with standard deviation σ .

Distribution	σ	Mean			
		$a_1b_1^a$	a_1b_2	a_2b_1	a_2b_2
0	1.0	0	0	0	0
1	1.0	0	$1.5\sigma=1.5$	$1.5\sigma=1.5$	$3\sigma=3.0$
2	0.25	0	0	0	0
3	0.25	0	$1.5\sigma=0.375$	$1.5\sigma=0.375$	$3\sigma=0.75$

a. a_ib_j indicates that level i of factor A and level j of factor B are applied. The corresponding cell is Cell ij .

Constant variance of the normal distribution was used to meet the assumption of ANOVA. This corresponds to a constant CV for the original lognormal population. For Distributions 0 and 2, the null distributions, the factor effects are all zero. Distributions 1 and 3 have non-zero main effects but no interaction effect.

Censoring and Replacements.

Censoring—Type I censoring was performed by determining a cutoff (analogous to a detection limit in environmental data sets) from the overall distribution for each simulation. The cutoff z_c was defined such that

$$P(Y < z_c) = c$$

where

Y = random variable for the overall distribution

c = censoring level.

The censoring was based on the overall distribution because it is more realistic than simply censoring each cell equally. The probabilistic approach was used in preference to percentile censoring based on the data because the latter would represent Type II censoring, not Type I censoring. The censoring levels used corresponded to 0%, 20%, and 60% censoring.

Methods—For the simulations, four methods for analyzing two-way tables were considered:

- the adjusted MLE, described below
- the LR method, carried out separately for the data in each cell.
- the DL/2 method, in which each censored normal value is replaced by $z_c - \ln 2$
- uniform replacement between zero and the detection limit in the lognormal scale.

These methods are explained more fully in the following paragraphs.

The MLE was found by the Wolynetz (1979b) version of the EM algorithm, presented in Section B-5.4.2. The extension to the two-way analysis of variance involved only modifying the design matrix. The two-way model can be written as

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad (B6.1-1)$$

where $i = 1$ or 2 , $j = 1$ or 2 , and k takes values from 1 to the number of replicates, 3 or 8 . The random error terms ϵ_{ijk} are independently and identically normally distributed. The constraints used are $\tau_1 = \beta_1 = (\tau\beta)_{11} = (\tau\beta)_{12} = (\tau\beta)_{21} = 0$. The EM calculations were performed with a program written in the C language. After the MLE was found, the resulting estimate of σ^2 was multiplied by $n/(n-4)$ in an attempt to reduce to bias. (Here, n is the total number of observations, 12 or 32 .) If there were no censoring, multiplication by $n/(n-4)$ would change the MLE of σ^2 into the usual unbiased estimator. This bias adjustment is the reason for the term "adjusted MLE." The sums of squares used in the ANOVA tests were not calculated directly by summing observations, but instead were calculated algebraically from the MLEs of the model parameters μ , τ_2 , β_2 , and $(\tau\beta)_{22}$. When the program was run on a sample data set with no censoring, the results agreed with a SAS analysis of the same data set. With censored data, the degrees of freedom used in the denominator for the F-test was not $n - 4$, but instead was the denominator in the Wolynetz estimator of σ^2 , involving an adjustment for the number of censored observations.

The LR method is explained in Sections 4.1.1 and B-4.1.1, and was applied separately to the data for each cell, for data sets having at least 2 uncensored observations in each cell.

The DL/2 method is described in Sections 4.1.3, B-4.1.3, and above in the opening paragraphs of Section B-6. It results in replacing censored normal values by $z_c - \ln 2$, where z_c is the censoring cutoff for the normal distribution.

The final method is performed by letting

$$y_{ij} = z_c + \ln[i/(n_{cj} + 1)], \quad i = 1, 2, \dots, n_{cj}$$

where

y_{ij} is the i th replacement value for the j th cell of the data set
 z_c is the censoring cutoff for data generated from normal distributions,
 n_{cj} is the number of censored values in the j th cell.

This is motivated by the fact that $\exp(y)$ is lognormally distributed with censoring cutoff $\exp(z_c)$, and that

$$\exp(y_{ij}) = \exp(z_c) \times [i/(n_{cj} + 1)],$$

are uniformly spaced between 0 and the cutoff $\exp(z_c)$. In the following discussion, the method is called the *uniform replacement* method. It models more variability in the replacement values than does the DL/2 replacement, and therefore may reflect the true data values better. If a cell has only one censored observation, the uniform replacement is identical to the DL/2 method.

In tabulations of the results, the abbreviations EM, LR, DL/2, and UN are used to denote the four methods. The term EM is used instead of MLE to prevent casual readers from thinking that the MLE of σ^2 is used, when in fact a bias adjustment has been applied to the MLE.

Sample Size. Simulations were run with 3 and 8 replicates per cell. For simulations comparing uniform replacement to DL/2 replacement, and those comparing the LR method to the DL/2 method, the same seed was used to generate data sets using the SAS function RANNOR (SAS Institute Inc., 1988a). For the 3-replicate case, 5000 data sets were generated and for 8 replicates, 2000 data sets were generated. Programs used to create data sets are found in Section B-6.1.5. The simulations required that at least two uncensored observations remained in each data set following censoring. In addition, at least two uncensored observations were required in each cell for simulations using the LR method, because a separate regression calculation was performed

for each cell. Because of these requirements, the actual number of data sets used in simulations varies from the original number generated.

For simulations comparing the adjusted MLE to the DL/2 method, the same seed was used within any one choice of distribution and censoring level. The random numbers were generated based on a Box-Muller transformation of uniform pseudo-random variates generated by a linear congruential generator. The programming was written in the C language, and is available on request. For each distribution and censoring level, 5000 data sets were generated. These simulations required that at least two uncensored observations remained in each data set following censoring.

Simulation Statistics. There is great potential for confusion in the terminology. ANOVA calculations involve sums of squares and their sample means. These means of sums of squares (commonly called mean squares) are random variables, which will be denoted here by *ms*. If there were no censoring, their expected values would be given by algebraic formulas; these expected values are denoted here by *ems*. One can think of *ms* as a random variable that tries to estimate *ems*. In addition, there are other parameters such as treatment effects and cell means, which can be estimated by random quantities. The mean squared error (MSE) is the expected value of the squared difference between the fixed quantity, *ems* or the parameter, and its estimator. The MSE of a parameter estimator is familiar, but the MSE of *ms* can sound confusing at first.

For each simulation, power of mean squares for main and interaction effects and the relative bias of mean squares for main effects, interaction, and within-cell error are computed. Variability in mean squares is expressed in terms of variance and relative mean squared error (MSE). For power and relative bias a standard error is also reported, to show that the number of simulations was large enough to give accurate estimates.

Relative bias of *ms* is calculated in the following manner:

N

$$\{[(1/N)\sum_{i=1} ms_i] - ems\}/ems$$

where

N is the number of data sets with a given censoring and type of replacement,

ms_i is the mean square of the i th data set with a given censoring and type of replacement, and

ems is the mean square that is expected with no censoring.

Formulas for determining ems for two-way ANOVA with fixed effects may be found in Neter and Wasserman (1974, p. 577). For the null distributions (Distributions 0 and 2 in Table B-6.1.1), ems equals σ^2 . For Distributions 1 and 3, and 3 replicates per cell, $ems = 7.75\sigma^2$ for main effects and $ems = \sigma^2$ for interaction and error. For these distributions with 8 replicates, the main effects ems is $18\sigma^2$ and the interaction and error ems are both σ^2 . Examination of the biases of the sums of squares can reveal information about the behavior of the ANOVA tests.

The simulation standard error in relative bias is computed by

$$s_{ms}/[ems \times (N^{1/2})]$$

where s_{ms} is the sample standard deviation of the mean squares for data sets with a given censoring and type of replacement, and N and ems are as described above.

The relative mean square error is

$$\frac{N}{[(1/N)\sum_{i=1} (ms_i - ems)^2]/ems^2}$$

This is equal to

$$[s_{ms}^2 + B_{ms}^2]/ems^2$$

where s_{ms} and ems are defined above, and B_{ms} is the sample bias of the mean squares for data sets with a given censoring and type of replacement.

Power was computed from the count of the number of data sets that rejected the hypothesis of no main or interaction effects, for a given censoring level and type of replacement, and a nominal Type I error of 0.05. For the simulations comparing the adjusted MLE to the DL/2 method, model parameter estimates are also computed. The methods' abilities to estimate the parameters are assessed by comparing the bias and mean squared error of the estimates.

Simulations Performed. Three groups of simulations were performed:

- 1) simulations with 3 replicates per cell and the DL/2 and uniform replacement methods
- 2) simulations with 8 replicates per cell and the DL/2 and LR methods
- 3) simulations with 3 and 8 replicates per cell, and the DL/2 and adjusted MLE methods.

The first two sets of simulations used Distributions 0 and 1 shown in Table B-6.1.1, while the third set used all four distributions. All sets used censoring levels of 20% and 60%. All replacements were done by cell. (With the DL/2 method, replacement by cell within each data set is equivalent to replacement by data set.)

The motivation for the first set of simulations, using three replicates per cell, is twofold: 1) environmental data are typically expensive to collect, especially in regulatory settings and 2) the volume of material available for "replication" is limited. (Often true replicates are not

possible, only collocated or split samples.) Therefore, a small sample size is used to more realistically simulate practical situations. The requirement of at least two uncensored values in each data set is applied, since as a matter of practicality, data sets with fewer uncensored values would probably not be subjected to analysis. As a result of this requirement, a larger number of data sets contribute to the statistics at lower censoring levels compared to higher censoring levels. The number of data sets is also dependent on the joint distribution of the data generated.

The second set of simulations, using eight replicates per cell, was performed primarily to provide a basis for evaluating the regression method in the two-way ANOVA application. A larger sample size is used in these simulations for the following reasons:

- 1) reasonable use of the regression method requires replacement by cell and
- 2) the regression method requires two uncensored values for estimation.

Therefore, the requirement of two or more uncensored values in each cell of a data set is applied to the simulations. To meet this requirement and in order to have a large number of acceptable data sets generated in a reasonable amount of computing time, a larger sample size is required.

The third set of simulations was performed 1) to assess the performance of the adjusted MLE in the two-way ANOVA model, and 2) to assess the DL/2 method when the variance was reduced.

The simulations comparing the DL/2, LR, and uniform replacement methods were run using SAS procedures MEANS, UNIVARIATE, FORMAT, and TABULATE (SAS Institute Inc., 1988b) and ANOVA (SAS Institute Inc., 1988c). The simulation programs are found in Section B-6.1.5. The simulations comparing the DL/2 and adjusted MLE methods were run with a program written in C. The least squares calculations were based on the decomposition of the design matrix by Givens'

method. The design matrix was generated to be non-singular based on the parameter constraints discussed below Equation (B6.1-1).

B-6.1.3 Simulation Results

Power. Power is shown in Tables B-6.1.2 through B-6.1.5. For these tables, the distribution is defined in Table B-6.1.1, CENSOR is the average percent censored, TYPE denotes one of the methods described earlier, N is the number of data sets used, SOURCE is a source of variation for the two-way setting with factors A and B, POWER is the power or Type I error, as appropriate, and STDERROR or SE is the standard error for power based on binomial distribution.

Table B-6.1.2 shows the power results for the first set of simulations, performed using 3 replicates per cell. The techniques examined are DL/2 and uniform replacement. The table shows the following conclusions:

- The power generally decreases with an increase in censoring from 20% to 60%, for both techniques and both distributions.
- At 60% censoring, the uniform replacement method does not have the desired Type I error: At 60% censoring under the null distribution, confidence intervals for the power using uniform replacement do not include the nominal value (0.05) or overlap with confidence intervals for simulations with no censoring. That is, from the A column of the table, $0.0316 + 2 \times 0.0025$ is well below 0.05, and even below $0.0486 - 2 \times 0.0030$; similar results are seen for the B and A*B columns.
- The interaction term for Distribution 2 with 60% censoring has anomalously high power, which in this instance should ideally be 0.05. Reasons for this behavior are discussed below for the third set of simulations, when cell mean estimates are considered.

Table B-6.1.2. Simulated power with 3 replicates per cell, DL/2 and uniform replacement.

SAS

				SOURCE							
				A		B		A*B		ERROR	
DISTN	CENSOR	TYPE	N	POWER	STDERROR	POWER	STDERROR	POWER	STDERROR	POWER	STDERROR
0	0	none	5000	0.0486	0.0030	0.0482	0.0030	0.0496	0.0031	.	0.0000
	20	DL/2	5000	0.0506	0.0031	0.0520	0.0031	0.0510	0.0031	.	0.0000
		UN	5000	0.0498	0.0031	0.0492	0.0031	0.0484	0.0030	.	0.0000
	60	DL/2	4901	0.0473	0.0030	0.0453	0.0030	0.0463	0.0030	.	0.0000
		UN	4901	0.0316	0.0025	0.0286	0.0024	0.0304	0.0025	.	0.0000
	0	none	5000	0.6224	0.0069	0.6142	0.0069	0.0496	0.0031	.	0.0000
1	20	DL/2	5000	0.6144	0.0069	0.6058	0.0069	0.0502	0.0031	.	0.0000
		UN	5000	0.6104	0.0069	0.6018	0.0069	0.0442	0.0029	.	0.0000
	60	DL/2	4976	0.4445	0.0070	0.4582	0.0071	0.2034	0.0057	.	0.0000
		UN	4976	0.3754	0.0069	0.3903	0.0069	0.1202	0.0046	.	0.0000
	0	none	5000	0.6224	0.0069	0.6142	0.0069	0.0496	0.0031	.	0.0000
	20	DL/2	5000	0.6144	0.0069	0.6058	0.0069	0.0502	0.0031	.	0.0000

Table B-6.1.3. Simulated power with 8 replicates per cell, DL/2 and LR methods.

SAS

				SOURCE							
				A		B		A*B		ERROR	
DISTN	CENSOR	TYPE	N	POWER	STDERROR	POWER	STDERROR	POWER	STDERROR	POWER	STDERROR
0	0	none	1998	0.0405	0.0044	0.0470	0.0047	0.0455	0.0047	.	0.0000
1	0	none	1782	0.9820	0.0031	0.9804	0.0033	0.0404	0.0047	.	0.0000
	20	DL/2	1782	0.9792	0.0034	0.9764	0.0036	0.0483	0.0051	.	0.0000
		LR	1782	0.9736	0.0038	0.9635	0.0044	0.1021	0.0072	.	0.0000

Table B-6.1.4. Simulated power with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Source					
				A		B		A*B	
				Power	SE	Power	SE	Power	SE
0	0	None	5,000	0.0496	0.0031	0.0488	0.0030	0.0454	0.0029
		DL/2	5,000	0.0530	0.0032	0.0510	0.0031	0.0508	0.0031
		EM	5,000	0.0800	0.0038	0.0780	0.0038	0.0752	0.0037
	60	None	5,000	0.0480	0.0030	0.0548	0.0032	0.0514	0.0031
		DL/2	4,577	0.0444	0.0030	0.0531	0.0033	0.0496	0.0032
		EM	4,577	0.2580	0.0065	0.2550	0.0064	0.2482	0.0064
	1	None	5,000	0.6256	0.0068	0.6266	0.0068	0.0498	0.0031
		DL/2	5,000	0.6096	0.0069	0.6200	0.0069	0.0510	0.0031
		EM	5,000	0.6814	0.0066	0.6880	0.0066	0.1470	0.0050
1	0	None	5,000	0.6288	0.0068	0.6214	0.0069	0.0492	0.0031
		DL/2	4,829	0.4649	0.0072	0.4645	0.0072	0.2063	0.0058
		EM	4,829	0.7287	0.0064	0.7267	0.0064	0.1655	0.0053
	60	None	5,000	0.0456	0.0030	0.0546	0.0032	0.0550	0.0032
		DL/2	5,000	0.0418	0.0028	0.0512	0.0031	0.0468	0.0030
		EM	5,000	0.0728	0.0037	0.0834	0.0039	0.0826	0.0039
	2	None	5,000	0.0484	0.0030	0.0464	0.0030	0.0510	0.0031
		DL/2	4,585	0.0495	0.0032	0.0471	0.0031	0.0545	0.0034
		EM	4,585	0.3967	0.0072	0.3950	0.0072	0.4057	0.0073
2	0	None	5,000	0.7902	0.0058	0.8024	0.0056	0.0024	0.0007
		DL/2	5,000	0.6814	0.0066	0.7000	0.0065	0.0188	0.0019
		EM	5,000	0.6426	0.0068	0.6660	0.0067	0.0074	0.0012
	60	None	5,000	0.6152	0.0069	0.6354	0.0068	0.0476	0.0030
		DL/2	4,847	0.4574	0.0072	0.4557	0.0072	0.1632	0.0053
		EM	4,847	0.7353	0.0063	0.7520	0.0062	0.2860	0.0065
	3	None	5,000	0.0456	0.0030	0.0546	0.0032	0.0550	0.0032
		DL/2	5,000	0.0418	0.0028	0.0512	0.0031	0.0468	0.0030
		EM	5,000	0.0728	0.0037	0.0834	0.0039	0.0826	0.0039

Table B-6.1.5. Simulated power with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Source					
				A		B		A*B	
				Power	SE	Power	SE	Power	SE
0	0	None	5,000	0.0474	0.0030	0.0516	0.0031	0.0518	0.0031
		DL/2	5,000	0.0468	0.0030	0.0536	0.0032	0.0540	0.0032
		EM	5,000	0.0516	0.0031	0.0568	0.0033	0.0578	0.0033
	60	None	5,000	0.0424	0.0028	0.0460	0.0030	0.0512	0.0031
		DL/2	5,000	0.0442	0.0029	0.0502	0.0031	0.0518	0.0031
		EM	5,000	0.1254	0.0047	0.1294	0.0047	0.1344	0.0048
	1	None	5,000	0.9842	0.0018	0.9838	0.0018	0.0466	0.0030
		DL/2	5,000	0.9776	0.0021	0.9802	0.0020	0.0516	0.0031
		EM	5,000	0.9818	0.0019	0.9816	0.0019	0.0684	0.0036
1	0	None	5,000	0.9812	0.0019	0.9848	0.0017	0.0486	0.0030
		DL/2	5,000	0.8996	0.0043	0.8996	0.0043	0.3754	0.0068
		EM	5,000	0.9510	0.0031	0.9510	0.0031	0.3390	0.0067
	60	None	5,000	0.9812	0.0019	0.9848	0.0017	0.0486	0.0030
		DL/2	5,000	0.8996	0.0043	0.8996	0.0043	0.3754	0.0068
		EM	5,000	0.9510	0.0031	0.9510	0.0031	0.3390	0.0067
	2	None	5,000	0.0522	0.0031	0.0536	0.0032	0.0468	0.0030
		DL/2	5,000	0.0518	0.0031	0.0482	0.0030	0.0484	0.0030
		EM	5,000	0.0548	0.0032	0.0564	0.0033	0.0524	0.0032
2	0	None	5,000	0.0514	0.0031	0.0528	0.0032	0.0492	0.0031
		DL/2	4,999	0.0502	0.0031	0.0546	0.0032	0.0498	0.0031
		EM	4,999	0.1404	0.0049	0.1458	0.0050	0.1358	0.0048
	60	None	5,000	0.0514	0.0031	0.0528	0.0032	0.0492	0.0031
		DL/2	4,999	0.0502	0.0031	0.0546	0.0032	0.0498	0.0031
		EM	4,999	0.1404	0.0049	0.1458	0.0050	0.1358	0.0048
	3	None	5,000	0.9842	0.0018	0.9830	0.0018	0.0502	0.0031
		DL/2	5,000	0.9588	0.0028	0.9588	0.0028	0.1250	0.0047
		EM	5,000	0.9822	0.0019	0.9812	0.0019	0.0736	0.0037
3	0	None	5,000	0.9820	0.0018	0.9852	0.0017	0.0500	0.0031
		DL/2	5,000	0.8882	0.0045	0.8892	0.0044	0.2522	0.0061
		EM	5,000	0.9512	0.0030	0.9544	0.0030	0.4250	0.0070

- The DL/2 method is clearly preferable to uniform replacement in terms of Type I error. It performs acceptably in these simulations except for its tendency to falsely reject the no-interaction hypothesis at 60% censoring when main effects are present.

For the second set of simulations, the DL/2 and LR methods are compared at 20% censoring under Distribution 1. The results, shown in Table B-6.1.3, indicate that the two methods have virtually the same power for detecting the main effects. However, the DL/2 method is clearly preferable to LR in testing for interaction when only main effects are present, because it has a lower probability of falsely declaring the presence of an interaction. Through an oversight, simulations were not performed under Distribution 0; this will be corrected in a future revision.

The results of the third set of simulations are displayed in Tables B-6.1.4 and B-6.1.5. These tables show the following conclusions.

- The DL/2 method is superior to the adjusted MLE at maintaining a Type I error rate at 0.05. This is true at both levels of censoring and both values of σ . At 60% censoring, the adjusted MLE has a Type I error rate of about 0.33 for the 3-replicate simulations, and about 0.14 for the 8-replicate simulations. (It is not known why this Type I error depends on σ .) If the estimate of σ^2 had not been adjusted (increased) to reduce the bias, the Type I errors would presumably have been even larger.
- Both methods falsely reject the no-interaction hypothesis under Distributions 1 and 3, at rates higher than the desired 0.05.
- For the DL/2 method, the power decreases with increased censoring and reduced replication.
- For the adjusted MLE, the power decreases with reduced replication. As censoring increases, the major change in power is that the Type I error) increases; in addition, the power under Distributions 1

and 3 increases when there are three replicates, and decreases slightly when there are 8 replicates.

Relative Bias. Relatives biases of mean squares are shown in Tables B-6.1.6 through B-6.1.9. For these tables, the distributions are defined in Table B-6.1.1, CENSOR is the average percent censored, TYPE denotes one of the methods described earlier, N is the number of data sets used, SOURCE is a source of variation for the two-way setting with factors A and B, RELBIAS or RB is the relative bias of mean squares, RBERROR or RBE is the standard error of relative bias.

Table B-6.1.6 shows relative bias for the simulations with 3 replicates per cell and using the DL/2 and uniform replacement techniques. The results show that the absolute value of the bias at the 60% censoring level is larger than at the 20% censoring level. The trend is more pronounced with Distribution 1 except for the relative bias of the interaction mean square with 60% censoring.

There are no differences in relative bias between all combinations of censoring levels and replacement types. For 20% censoring of the null distribution, DL/2 replacement is preferable, as the relative bias is less for DL/2 than for uniform replacement. However, for 60% censoring, uniform has lower (in absolute value) relative bias. The preferred replacement technique for Distribution 1 based on relative bias is uniform for both censoring levels.

The relative bias of mean squares is much larger at 60% censoring than at lower censoring. This indicates that, for the most part, use of DL/2 or uniform replacement results in severely biased mean squares in situations where 60% of the measurements are censored, irrespective of the underlying distribution. As is the case for power results, however, the interaction term for the Distribution 1 at 60% censoring is an anomaly. The relative bias is not as extreme as for other sources of variation. Even so, a test for significant interactions would probably be inaccurate because the error mean square is severely biased. The power results bear this out.

Table B-6.1.6. Simulated relative bias of mean squares with 3 replicates per cell, DL/2 and uniform replacement.

SAS

				SOURCE							
				A		B		A*B		ERROR	
				RELBIAS	RBERROR	RELBIAS	RBERROR	RELBIAS	RBERROR	RELBIAS	RBERROR
DISTN	CENSOR	TYPE	N								
0	0	none	5000	0.0052	0.0004	0.0154	0.0004	-0.0094	0.0004	0.0041	0.0000
	20	DL/2	5000	0.0281	0.0004	0.0431	0.0004	0.0238	0.0004	0.0327	0.0000
		UN	5000	0.0557	0.0004	0.0718	0.0005	0.0500	0.0004	0.0536	0.0000
	60	DL/2	4901	-0.3976	0.0001	-0.3978	0.0001	-0.3883	0.0002	-0.3945	0.0000
		UN	4901	-0.3148	0.0002	-0.3136	0.0002	-0.3036	0.0002	-0.2541	0.0000
	0	none	5000	-0.0024	0.0001	-0.0009	0.0058	-0.0094	0.0004	0.0041	0.0000
1	20	DL/2	5000	-0.0830	0.0001	-0.0776	0.0049	-0.0727	0.0003	-0.0640	0.0000
		UN	5000	-0.0582	0.0001	-0.0526	0.0052	-0.0527	0.0004	-0.0245	0.0000
	60	DL/2	4976	-0.6407	0.0000	-0.6383	0.0011	0.0624	0.0003	-0.4650	0.0000
		UN	4976	-0.6039	0.0000	-0.6013	0.0013	0.1018	0.0003	-0.3074	0.0000
	0	none	5000	-0.0024	0.0001	-0.0009	0.0058	-0.0094	0.0004	0.0041	0.0000
	20	DL/2	5000	-0.0830	0.0001	-0.0776	0.0049	-0.0727	0.0003	-0.0640	0.0000

Table B-6.1.7. Simulated relative bias of mean squares with 8 replicates per cell, DL/2 and LR methods.

SAS

				SOURCE							
				A		B		A*B		ERROR	
				RELBias	RBERROR	RELBias	RBERROR	RELBias	RBERROR	RELBias	RBERROR
DISTN	CENSOR	TYPE	N								
0	0	none	1998	-0.0400	0.0007	0.0060	0.0007	-0.0584	0.0007	0.0057	0.0001
1	0	none	1782	0.0110	0.0003	0.0118	0.0046	-0.1027	0.0007	0.0164	0.0001
	20	DL/2	1782	-0.0691	0.0002	-0.0649	0.0043	-0.0877	0.0007	-0.0516	0.0001
		LR	1782	0.0252	0.0004	0.0305	0.0071	0.6769	0.0018	0.0155	0.0003

Table B-6.1.8. Simulated relative bias of mean squares with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Source							
				A		B		A*B		Error	
				RB	RBE	RB	RBE	RB	RBE	RB	RBE
0	0	None	5,000	0.0297	0.0208	-0.0006	0.0199	-0.0101	0.0194	-0.0014	0.0070
		DL/2	5,000	0.0645	0.0214	0.0371	0.0205	0.0335	0.0202	0.0401	0.0067
		EM	5,000	0.3497	0.0390	0.3111	0.0373	0.2809	0.0341	0.0546	0.0087
	60	None	5,000	-0.0141	0.0194	-0.0093	0.0196	0.0011	0.0203	-0.0105	0.0071
		DL/2	4,577	-0.3895	0.0124	-0.3809	0.0127	-0.3829	0.0128	-0.3774	0.0046
		EM	4,577	2.5055	0.1065	2.3355	0.0890	2.4864	0.0996	-0.0787	0.0118
	1	None	5,000	-0.0092	0.0098	0.0027	0.0099	-0.0063	0.0197	-0.0082	0.0070
		DL/2	5,000	-0.0786	0.0092	-0.0684	0.0091	-0.0383	0.0190	-0.0664	0.0066
		EM	5,000	0.2844	0.0174	0.2950	0.0174	0.9710	0.0481	0.0138	0.0084
1	0	None	5,000	0.0142	0.0099	-0.0007	0.0099	-0.0309	0.0195	0.0009	0.0070
		DL/2	4,829	-0.6274	0.0044	-0.6252	0.0044	0.0895	0.0181	-0.4604	0.0047
		EM	4,829	0.5586	0.0277	0.5521	0.0270	1.6602	0.0688	-0.0048	0.0122
	60	None	5,000	-0.0276	0.0196	0.0360	0.0208	0.0211	0.0205	-0.0088	0.0069
		DL/2	5,000	1.9548	0.0595	2.0423	0.0609	2.0182	0.0608	1.9757	0.0198
		EM	5,000	0.3038	0.0358	0.3805	0.0383	0.3670	0.0396	0.0605	0.0092
	1	None	5,000	-0.0136	0.0197	-0.0212	0.0200	0.0240	0.0208	0.0170	0.0072
		DL/2	4,585	2.1204	0.0617	2.0569	0.0608	2.1705	0.0657	2.1943	0.0155
		EM	4,585	4.5874	0.1117	4.5168	0.1108	4.5731	0.1118	0.0106	0.0126
2	0	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	60	None	5,000	-0.0080	0.0098	0.0108	0.0100	-0.0192	0.0199	0.0021	0.0072
		DL/2	4,847	0.4659	0.0167	0.4894	0.0168	2.3051	0.0586	1.1298	0.0156
		EM	4,847	0.9629	0.0254	1.0192	0.0261	2.5983	0.0731	-0.0033	0.0118
	1	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
3	0	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	60	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	1	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
4	0	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	60	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	1	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
5	0	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	60	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059
	1	None	5,000	-0.0913	0.0059	-0.0723	0.0059	-0.6314	0.0075	0.0029	0.0038
		DL/2	5,000	0.7190	0.0121	0.7570	0.0119	0.0953	0.0201	1.1929	0.0091
		EM	5,000	-0.0460	0.0076	-0.0281	0.0075	-0.5070	0.0134	0.3297	0.0059

Table B-6.1.9. Simulated relative bias of mean squares with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Source							
				A		B		A*B		Error	
				RB	RBE	RB	RBE	RB	RBE	RB	RBE
0	0	None	5,000	-0.0359	0.0192	0.0287	0.0210	0.0437	0.0204	0.0008	0.0038
		DL/2	5,000	-0.0083	0.0196	0.0661	0.0217	0.0796	0.0212	0.0335	0.0035
		EM	5,000	0.0210	0.0205	0.0995	0.0227	0.1133	0.0222	0.0246	0.0045
	60	None	5,000	-0.0187	0.0201	-0.0062	0.0193	0.0181	0.0204	-0.0049	0.0037
		DL/2	5,000	-0.4100	0.0119	-0.4064	0.0119	-0.3973	0.0122	-0.4008	0.0025
		EM	5,000	1.3037	0.0800	1.1930	0.0639	1.3066	0.0773	-0.0003	0.0071
	1	None	5,000	0.0590	0.0068	0.0520	0.0066	-0.0132	0.0194	0.0007	0.0038
		DL/2	5,000	-0.0212	0.0063	-0.0290	0.0061	-0.0444	0.0190	-0.0626	0.0036
		EM	5,000	0.1038	0.0087	0.0943	0.0084	0.3267	0.0398	0.0073	0.0045
1	0	None	5,000	0.0532	0.0067	0.0635	0.0067	-0.0201	0.0196	-0.0027	0.0038
		DL/2	5,000	-0.6290	0.0029	-0.6242	0.0029	1.0108	0.0264	-0.4638	0.0025
		EM	5,000	0.7299	0.0170	0.7645	0.0177	3.1543	0.0765	0.0027	0.0067
	60	None	5,000	0.0219	0.0201	-0.0117	0.0201	-0.0152	0.0198	0.0023	0.0038
		DL/2	5,000	2.0382	0.0592	1.9110	0.0586	1.9256	0.0583	1.9923	0.0106
		EM	5,000	0.1042	0.0220	0.0682	0.0221	0.0661	0.0218	0.0535	0.0048
	1	None	5,000	0.0112	0.0202	0.0213	0.0207	0.0020	0.0203	-0.0009	0.0038
		DL/2	4,999	2.0607	0.0616	2.1312	0.0628	2.0655	0.0615	2.0401	0.0067
		EM	4,999	1.7622	0.0881	1.7828	0.0890	1.7105	0.0884	0.0723	0.0075
2	0	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
	60	None	5,000	0.0021	0.0065	-0.0009	0.0064	0.0238	0.0199	-0.0068	0.0037
		DL/2	5,000	0.4056	0.0110	0.4081	0.0109	4.3017	0.0794	1.1045	0.0079
		EM	5,000	0.7847	0.0150	0.7685	0.0145	3.5250	0.0675	0.0230	0.0066
	1	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
3	0	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
	60	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
	1	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
3	0	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
	60	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043
	1	None	5,000	-0.0056	0.0064	-0.0025	0.0064	-0.0115	0.0200	-0.0017	0.0037
		DL/2	5,000	0.8797	0.0128	0.8801	0.0128	2.0216	0.0546	1.1682	0.0089
		EM	5,000	0.0390	0.0081	0.0406	0.0080	0.3082	0.0346	0.0070	0.0043

Table B-6.1.7 gives the results for the second set of simulations. The relative bias of the interaction mean squares is much better when the DL/2 method is used rather than the LR method. In fact, the error mean square is not different from the case of no censoring. It is interesting to note that for the treatment effects the relative biases of mean squares using the LR method are similar to those for no censoring. All this is at the 20% censoring level. At 60% censoring Distribution 1 rarely yields a data set with enough uncensored values in Cell 11 to allow use of the LR method.

The results of the third set of simulations are displayed in Tables B-6.1.8 and B-6.1.9. Both the DL/2 and adjusted MLE methods have biased estimators of the treatment mean squares, and this bias increases in magnitude as censoring increases. The DL/2 method tends to be biased low then the within-cell standard deviation is 1.0, and biased high when the standard deviation is 0.25. The adjusted MLE tends to overestimate the treatment mean squares; however it provides a nearly unbiased estimator of the error mean square.

The DL/2 method tends to have smaller bias (in absolute value) than the adjusted MLE when σ equals 1, and larger bias when $\sigma = 0.25$, although there are exceptions in each case. Under the non-null distributions, the bias in estimating the main effect mean squares is typically much lower than the bias in estimating the interaction mean square. Further, the biases of the adjusted MLE tend to decrease in magnitude as the number of replicates is increased, while the biases of the DL/2 mean squares show no clear pattern as the number of replicates is increased.

Relative MSE of Mean Squares. Tables B-6.1.10 through B-6.1.17 show the variances and relative MSEs of the mean squares. For these tables, the distributions are defined in Table B-6.1.1, CENSOR is the average percent censored, TYPE denotes one of the methods described earlier, N is the number of data sets used, SOURCE is a source of variation for the two-way setting with factors A and B, VARMS is the variance of the mean square, and RELMSE the relative MSE of mean square. Standard errors are not shown.

Table B-6.1.10. Simulated variance of mean squares with 3 replicates per cell, DL/2 and uniform replacement.

SAS

				SOURCE			
				A	B	A*B	ERROR
				VARMS	VARMS	VARMS	VARMS
DISTN	CENSOR	TYPE	N				
0	0	none	5000	2.02	2.04	1.96	0.25
		DL/2	5000	2.00	2.15	2.07	0.22
	20	UN	5000	2.13	2.30	2.18	0.22
		DL/2	4901	0.69	0.70	0.74	0.10
	60	UN	4901	0.87	0.89	0.93	0.09
		DL/2	4901	0.87	0.89	0.93	0.09
1	0	none	5000	29.07	28.93	1.96	0.25
		DL/2	5000	24.41	24.39	1.72	0.22
	20	UN	5000	25.77	25.78	1.80	0.22
		DL/2	4976	5.30	5.25	1.47	0.11
	60	UN	4976	6.33	6.27	1.63	0.11
		DL/2	4976	6.33	6.27	1.63	0.11

Table B-6.1.11. Simulated variance of mean squares with 8 replicates per cell, DL/2 and LR methods.

SAS

				SOURCE			
				A	B	A*B	ERROR
				VARMS	VARMS	VARMS	VARMS
DISTN	CENSOR	TYPE	N				
0	0	none	1998	1.70	2.07	1.84	0.07
1	0	none	1782	65.72	66.78	1.62	0.07
	20	DL/2	1782	56.89	58.61	1.68	0.06
		LR	1782	152.90	161.85	10.80	0.32

Table B-6.1.12. Simulated variance of mean squares with 3 replicates per cell, DL/2 and adjusted MLE (EM) methods.

Distn	Censor	Type	N	Source			
				A	B	A*B	Error
				VarMS	VarMS	VarMS	VarMS
0	0	None	5,000	2.1544	1.9736	1.8881	0.2452
		DL/2	5,000	2.2877	2.1063	2.0371	0.2246
		EM	5,000	7.5899	6.9624	5.8123	0.3827
	60	None	5,000	1.8776	1.9278	2.0507	0.2501
		DL/2	4,577	0.7052	0.7372	0.7535	0.0977
		EM	4,577	51.9154	36.2887	45.4289	0.6357
	20	None	5,000	28.8288	29.2796	1.9407	0.2444
		DL/2	5,000	25.3158	24.7636	1.8121	0.2206
		EM	5,000	91.2184	91.0792	11.5617	0.3531
1	0	None	5,000	29.5284	29.2439	1.9021	0.2450
		DL/2	4,829	5.5191	5.5557	1.5827	0.1064
		EM	4,829	221.8104	211.3804	22.8690	0.7139
	60	None	5,000	0.0075	0.0084	0.0082	0.0009
		DL/2	5,000	0.0691	0.0723	0.0722	0.0076
		EM	5,000	0.0250	0.0287	0.0306	0.0016
	20	None	5,000	0.0076	0.0078	0.0085	0.0010
		DL/2	4,585	0.0681	0.0662	0.0774	0.0043
		EM	4,585	0.2233	0.2199	0.2237	0.0028
2	0	None	5,000	0.0415	0.0407	0.0011	0.0003
		DL/2	5,000	0.1713	0.1674	0.0079	0.0016
		EM	5,000	0.0681	0.0667	0.0035	0.0007
	60	None	5,000	0.1118	0.1165	0.0078	0.0010
		DL/2	4,847	0.3154	0.3228	0.0650	0.0046
		EM	4,847	0.7365	0.7733	0.1012	0.0026
	20	None	5,000	0.0415	0.0407	0.0011	0.0003
		DL/2	5,000	0.1713	0.1674	0.0079	0.0016
		EM	5,000	0.0681	0.0667	0.0035	0.0007
3	0	None	5,000	0.0415	0.0407	0.0011	0.0003
		DL/2	5,000	0.1713	0.1674	0.0079	0.0016
		EM	5,000	0.0681	0.0667	0.0035	0.0007
	60	None	5,000	0.1118	0.1165	0.0078	0.0010
		DL/2	4,847	0.3154	0.3228	0.0650	0.0046
		EM	4,847	0.7365	0.7733	0.1012	0.0026
	20	None	5,000	0.0415	0.0407	0.0011	0.0003
		DL/2	5,000	0.1713	0.1674	0.0079	0.0016
		EM	5,000	0.0681	0.0667	0.0035	0.0007

Table B-6.1.13. Simulated variance of mean squares with 8 replicates per cell, DL/2 and adjusted MLE (EM) methods.

Distn	Censor	Type	N	Source			
				A	B	A*B	Error
				VarMS	VarMS	VarMS	VarMS
0	0	None	5,000	1.8464	2.2115	2.0753	0.0712
		DL/2	5,000	1.9251	2.3483	2.2520	0.0596
		EM	5,000	2.0947	2.5741	2.4619	0.1007
	60	None	5,000	2.0216	1.8712	2.0786	0.0694
		DL/2	5,000	0.7071	0.7077	0.7437	0.0301
		EM	5,000	31.9932	23.7174	29.8812	0.2539
	1	None	5,000	74.2429	71.4008	1.8835	0.0723
		DL/2	5,000	63.9344	60.7964	1.8117	0.0648
		EM	5,000	123.5552	113.4544	7.9164	0.0996
1	0	None	5,000	73.5367	72.1528	1.9157	0.0721
		DL/2	5,000	13.7503	14.0385	3.4787	0.0312
		EM	5,000	470.4853	509.6565	29.2295	0.2272
	60	None	5,000	0.0079	0.0079	0.0076	0.0003
		DL/2	5,000	0.0685	0.0670	0.0663	0.0022
		EM	5,000	0.0094	0.0095	0.0093	0.0005
	2	None	5,000	0.0080	0.0084	0.0080	0.0003
		DL/2	4,999	0.0740	0.0771	0.0738	0.0009
		EM	4,999	0.1517	0.1547	0.1525	0.0011
2	0	None	5,000	0.2850	0.2860	0.0079	0.0003
		DL/2	5,000	1.1589	1.1547	0.0582	0.0016
		EM	5,000	0.4604	0.4481	0.0234	0.0004
	60	None	5,000	0.2994	0.2849	0.0078	0.0003
		DL/2	5,000	0.8475	0.8333	0.1231	0.0012
		EM	5,000	1.5865	1.4813	0.0890	0.0008
	3	None	5,000	0.2850	0.2860	0.0079	0.0003
		DL/2	5,000	1.1589	1.1547	0.0582	0.0016
		EM	5,000	0.4604	0.4481	0.0234	0.0004

Table B-6.1.14. Simulated relative MSE for the mean squares with 3 replicates per cell, DL/2 and uniform replacement.

SAS

				SOURCE			
				A	B	A*B	ERROR
				RELMSE	RELMSE	RELMSE	RELMSE
DISTN	CENSOR	TYPE	N				
0	0	none	5000	2.02	2.04	1.96	0.25
		DL/2	5000	2.00	2.15	2.07	0.22
	20	UN	5000	2.13	2.31	2.19	0.22
		DL/2	4901	0.85	0.86	0.89	0.25
		UN	4901	0.97	0.99	1.03	0.15
		DL/2	4901	0.97	0.99	1.03	0.15
1	0	none	5000	0.48	0.48	1.96	0.25
		DL/2	5000	0.41	0.41	1.73	0.23
	20	UN	5000	0.43	0.43	1.80	0.22
		DL/2	4976	0.50	0.49	1.47	0.33
		UN	4976	0.47	0.47	1.64	0.20
		DL/2	4976	0.47	0.47	1.64	0.20

Table B-6.1.15. Simulated relative MSE for the mean squares with 8 replicates per cell, DL/2 and LR method.

SAS

				SOURCE			
				A	B	A*B	ERROR
				RELMSE	RELMSE	RELMSE	RELMSE
DISTN	CENSOR	TYPE	N				
0	0	none	1998	1.70	2.07	1.84	0.07
1	0	none	1782	0.20	0.21	1.64	0.07
	20	DL/2	1782	0.18	0.19	1.68	0.07
		LR	1782	0.47	0.50	11.26	0.32

Table B-6.1.16. Simulated relative MSE for the mean squares with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Source			
				A	B	A*B	Error
				RelMSE	RelMSE	RelMSE	RelMSE
0	0	None	5,000	2.1548	1.9732	1.8878	0.2452
		DL/2	5,000	2.2914	2.1072	2.0378	0.2261
		EM	5,000	7.7107	7.0578	5.8901	0.3856
	60	None	5,000	1.8774	1.9275	2.0503	0.2502
		DL/2	4,577	0.8567	0.8821	0.9000	0.2401
		EM	4,577	58.1816	41.7355	51.6010	0.6418
	20	None	5,000	0.4800	0.4874	1.9403	0.2445
		DL/2	5,000	0.4276	0.4169	1.8132	0.2250
		EM	5,000	1.5993	1.6031	12.5023	0.3533
1	0	None	5,000	0.4917	0.4868	1.9027	0.2449
		DL/2	4,829	0.4855	0.4833	1.5903	0.3183
		EM	4,829	4.0042	3.8235	25.6206	0.7138
	60	None	5,000	1.9222	2.1582	2.0988	0.2383
		DL/2	5,000	21.5108	22.6855	22.5607	5.8577
		EM	5,000	6.4882	7.4910	7.9568	0.4242
	20	None	5,000	1.9431	1.9992	2.1686	0.2628
		DL/2	4,585	21.9328	21.1654	24.5263	5.9139
		EM	4,585	78.1965	76.6948	78.1652	0.7269
2	0	None	5,000	0.1853	0.1787	0.6833	0.0723
		DL/2	5,000	1.2469	1.2866	2.0333	1.8327
		EM	5,000	0.2923	0.2851	1.1497	0.2812
	60	None	5,000	0.4766	0.4967	1.9896	0.2618
		DL/2	4,847	1.5611	1.6151	21.9562	2.4570
		EM	4,847	4.0655	4.3341	32.6502	0.6710
	20	None	5,000	0.1853	0.1787	0.6833	0.0723
		DL/2	5,000	1.2469	1.2866	2.0333	1.8327
		EM	5,000	0.2923	0.2851	1.1497	0.2812
3	0	None	5,000	0.1853	0.1787	0.6833	0.0723
		DL/2	5,000	1.2469	1.2866	2.0333	1.8327
		EM	5,000	0.2923	0.2851	1.1497	0.2812
	60	None	5,000	0.4766	0.4967	1.9896	0.2618
		DL/2	4,847	1.5611	1.6151	21.9562	2.4570
		EM	4,847	4.0655	4.3341	32.6502	0.6710
	20	None	5,000	0.1853	0.1787	0.6833	0.0723
		DL/2	5,000	1.2469	1.2866	2.0333	1.8327
		EM	5,000	0.2923	0.2851	1.1497	0.2812

Table B-6.1.17. Simulated relative MSE for the mean squares with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Source			
				A	B	A*B	Error
				RelMSE	RelMSE	RelMSE	RelMSE
0	0	None	5,000	1.8474	2.2118	2.0768	0.0712
		DL/2	5,000	1.9247	2.3522	2.2579	0.0607
		EM	5,000	2.0947	2.5835	2.4743	0.1013
	60	None	5,000	2.0215	1.8709	2.0785	0.0694
		DL/2	5,000	0.8751	0.8727	0.9014	0.1908
		EM	5,000	33.6863	25.1360	31.5823	0.2539
	1	None	5,000	0.2326	0.2230	1.8833	0.0723
		DL/2	5,000	0.1977	0.1884	1.8133	0.0687
		EM	5,000	0.3920	0.3590	8.0216	0.0996
1	0	None	5,000	0.2298	0.2267	1.9157	0.0721
		DL/2	5,000	0.4380	0.4329	4.4997	0.2463
		EM	5,000	1.9846	2.1572	39.1731	0.2272
	60	None	5,000	2.0204	2.0205	1.9565	0.0715
		DL/2	5,000	21.6992	20.8028	20.6793	4.5292
		EM	5,000	2.4294	2.4440	2.3822	0.1191
	1	None	5,000	2.0361	2.1446	2.0588	0.0708
		DL/2	4,999	23.1870	24.2803	23.1656	4.3880
		EM	4,999	41.9393	42.7655	41.9515	0.2835
2	0	None	5,000	0.2021	0.2028	2.0097	0.0687
		DL/2	5,000	1.5956	1.5932	18.9733	1.7633
		EM	5,000	0.3280	0.3193	6.0894	0.0922
	60	None	5,000	0.2123	0.2020	1.9864	0.0692
		DL/2	5,000	0.7654	0.7574	50.0092	1.5306
		EM	5,000	1.7406	1.6409	35.2113	0.2166
	1	None	5,000	0.2021	0.2028	2.0097	0.0687
		DL/2	5,000	1.5956	1.5932	18.9733	1.7633
		EM	5,000	0.3280	0.3193	6.0894	0.0922
3	0	None	5,000	0.2021	0.2028	2.0097	0.0687
		DL/2	5,000	1.5956	1.5932	18.9733	1.7633
		EM	5,000	0.3280	0.3193	6.0894	0.0922
	60	None	5,000	0.2123	0.2020	1.9864	0.0692
		DL/2	5,000	0.7654	0.7574	50.0092	1.5306
		EM	5,000	1.7406	1.6409	35.2113	0.2166
	1	None	5,000	0.2021	0.2028	2.0097	0.0687
		DL/2	5,000	1.5956	1.5932	18.9733	1.7633
		EM	5,000	0.3280	0.3193	6.0894	0.0922

Table B-6.1.10 gives the variance of mean squares for simulations performed using 3 replicates and the DL/s and uniform replacement methods. A similar pattern as that described for relative bias is observed, with values for 60% censoring being distinctly different from the others. The variance of mean squares is probably lower for 60% censoring because more values are being replaced with the same or similar values. The variance of a DL/2 mean square in general is less than the variance from uniform replacement for the same reason. The variance of mean squares for main effects for the Distribution 1 is quite high compared to other values. Table B-6.1.11 shows similar results for the second set of simulations, with 8 replicates per cell and the DL/2 and LR methods. Tables B-6.1.12 and B-6.1.13 show similar comparisons for the third set of simulations, with 3 or 8 replicates per cell, and the DL/2 and adjusted MLE methods. The large variances of the treatment mean squares based on the adjusted MLE under 60% censoring are noteworthy.

Relative mean squared error for the mean squares for the three simulation sets are given in Tables B-6.1.14 through B-6.1.17, respectively. The estimates are dominated by the variance of mean squares and, for main effects for the second distribution, by the *ems* factor.

Parameter Estimation and Cell Means. Tables B-6.1.18 and B-6.1.19 show the bias in the estimate of the model parameters, and Tables B-6.1.20 and B-6.1.21 show the mean squared errors of these parameter estimates. Tables B-6.1.22 through B-6.1.25 show the same quantities for the cell means rather than the model parameters. For all these tables, the distributions are defined in Table B-6.1.1, Censor is the average percent censored, Type denotes one of the methods described earlier, and N is the number of data sets used. For each quantity tabulated, the standard error is shown in parentheses. The true cell means are defined in Table B-6.1.1.

Table B-6.1.18. Simulated bias of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter					
				μ		τ_2		β_2	
				Bias(SE)		Bias(SE)		Bias(SE)	$(\tau\beta)_{22}$
0	0	none	5,000	0.0002(0.0082)	0.0005(0.0117)	0.0025(0.0114)	0.0017(0.0162)	-0.0310(0.0035)	σ
	20	DL/2	5,000	-0.0289(0.0084)	0.0004(0.0119)	0.0017(0.0116)	0.0042(0.0166)	-0.0071(0.0033)	
		EM	5,000	-0.0300(0.0095)	0.0009(0.0134)	0.0070(0.0129)	0.0001(0.0185)	-0.0135(0.0040)	
	0	none	5,000	0.0085(0.0081)	-0.0249(0.0114)	-0.0010(0.0115)	0.0198(0.0163)	-0.0363(0.0035)	
	60	DL/2	4,577	0.1596(0.0066)	-0.0177(0.0095)	-0.0025(0.0096)	0.0161(0.0134)	-0.2354(0.0029)	
		EM	4,577	-0.1027(0.0161)	-0.0289(0.0230)	0.0160(0.0222)	0.0306(0.0319)	-0.1126(0.0054)	
1	0	none	5,000	0.0122(0.0081)	-0.0129(0.0116)	-0.0034(0.0115)	0.0081(0.0163)	-0.0344(0.0034)	
	20	DL/2	5,000	0.1209(0.0064)	-0.1438(0.0108)	-0.1347(0.0106)	0.1604(0.0159)	-0.0643(0.0034)	
		EM	5,000	-0.2796(0.0172)	0.2666(0.0193)	0.2763(0.0192)	-0.2601(0.0226)	-0.0332(0.0040)	
	0	none	5,000	-0.0146(0.0082)	0.0168(0.0115)	0.0026(0.0116)	-0.0104(0.0161)	-0.0299(0.0035)	
	60	DL/2	4,829	1.2217(0.0016)	-1.0426(0.0063)	-1.0422(0.0063)	0.8575(0.0122)	-0.3005(0.0032)	
		EM	4,829	-0.3771(0.0172)	0.1373(0.0219)	0.1421(0.0216)	0.1011(0.0271)	-0.0810(0.0056)	

Table B-6.1.18 (contd). Simulated bias of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter					
				μ		τ_2		$(\tau\beta)_{22}$	
				Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)
2	0	none	5,000	-0.0012(0.0020)	0.0022(0.0029)	0.0041(0.0029)	-0.0057(0.0041)	-0.0085(0.0009)	
	20	DL/2	5,000	-0.1138(0.0035)	0.0015(0.0050)	0.0071(0.0050)	-0.0062(0.0071)	0.1673(0.0015)	
		EM	5,000	-0.0092(0.0023)	0.0011(0.0034)	0.0052(0.0033)	-0.0072(0.0048)	-0.0028(0.0010)	
	0	none	5,000	0.0007(0.0021)	-0.0049(0.0029)	-0.0028(0.0029)	0.0089(0.0041)	-0.0057(0.0009)	
	60	DL/2	4,585	-0.2604(0.0037)	-0.0068(0.0054)	-0.0016(0.0053)	0.0107(0.0076)	0.1899(0.0012)	
		EM	4,585	-0.0754(0.0048)	-0.0104(0.0072)	-0.0005(0.0070)	0.0152(0.0101)	-0.0170(0.0014)	
3	0	none	5,000	-0.0003(0.0013)	-0.0034(0.0018)	0.0010(0.0018)	0.0016(0.0025)	-0.0019(0.0005)	
	20	DL/2	5,000	-0.2788(0.0022)	0.2136(0.0029)	0.2202(0.0029)	-0.1579(0.0036)	0.1161(0.0008)	
		EM	5,000	-0.0124(0.0019)	0.0077(0.0023)	0.0120(0.0023)	-0.0085(0.0029)	0.0349(0.0006)	
	0	none	5,000	0.0002(0.0021)	-0.0015(0.0029)	0.0023(0.0029)	-0.0002(0.0040)	-0.0077(0.0009)	
	60	DL/2	4,847	-0.1951(0.0012)	-0.0953(0.0037)	-0.0903(0.0036)	0.3189(0.0060)	0.0998(0.0015)	
		EM	4,847	-0.2350(0.0033)	0.1248(0.0059)	0.1356(0.0058)	-0.0241(0.0079)	-0.0187(0.0013)	

Table B-6.1.19. Simulated bias of parameter estimators with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter								
				μ		τ_2		β_2		$(\tau\beta)_{22}$		σ
				Bias(SE)		Bias(SE)		Bias(SE)		Bias(SE)		Bias(SE)
0	0	none	5,000	0.0070(0.0050)	-0.0073(0.0070)	-0.0066(0.0072)	0.0082(0.0102)	-0.0085(0.0019)				
	20	DL/2	5,000	-0.0180(0.0050)	-0.0091(0.0071)	-0.0081(0.0074)	0.0092(0.0104)	0.0095(0.0017)				
		EM	5,000	0.0015(0.0051)	-0.0100(0.0072)	-0.0101(0.0075)	0.0113(0.0106)	0.0005(0.0022)				
	0	none	5,000	0.0004(0.0049)	0.0000(0.0071)	0.0037(0.0071)	-0.0072(0.0101)	-0.0111(0.0019)				
	60	DL/2	5,000	0.1218(0.0038)	-0.0026(0.0055)	0.0001(0.0055)	-0.0027(0.0078)	-0.2340(0.0016)				
		EM	5,000	-0.0274(0.0078)	-0.0098(0.0109)	-0.0030(0.0105)	0.0039(0.0152)	-0.0287(0.0034)				
1	0	none	5,000	-0.0032(0.0050)	0.0139(0.0070)	0.0096(0.0070)	-0.0225(0.0099)	-0.0086(0.0019)				
	20	DL/2	5,000	0.1131(0.0039)	-0.1249(0.0064)	-0.1295(0.0064)	0.1376(0.0096)	-0.0407(0.0019)				
		EM	5,000	-0.0421(0.0075)	0.0494(0.0091)	0.0448(0.0090)	-0.0548(0.0115)	-0.0083(0.0022)				
	0	none	5,000	-0.0008(0.0049)	0.0014(0.0071)	0.0105(0.0070)	-0.0069(0.0099)	-0.0103(0.0019)				
	60	DL/2	5,000	1.2218(0.0010)	-1.0537(0.0037)	-1.0476(0.0038)	0.8588(0.0073)	-0.2778(0.0017)				
		EM	5,000	-0.7114(0.0152)	0.6522(0.0163)	0.6675(0.0163)	-0.6119(0.0185)	-0.0245(0.0032)				

Table B-6.1.19 (contd). Simulated bias of parameter estimators with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Parameter												
Distn	Censor	Type	N	μ		τ_2		β_2		$(\tau\beta)_{22}$		σ
				Bias(SE)		Bias(SE)		Bias(SE)		Bias(SE)		
2	0	none	5,000	-0.0012(0.0013)		0.0021(0.0018)		0.0014(0.0018)		-0.0037(0.0025)		-0.0019(0.0005)
	20	DL/2	5,000	-0.1113(0.0022)		0.0016(0.0030)		0.0001(0.0030)		-0.0049(0.0043)		0.1788(0.0008)
		EM	5,000	-0.0044(0.0013)		0.0011(0.0018)		0.0005(0.0018)		-0.0026(0.0026)		0.0034(0.0006)
	0	none	5,000	-0.0019(0.0013)		0.0029(0.0018)		0.0006(0.0018)		0.0017(0.0025)		-0.0023(0.0005)
	60	DL/2	4,999	-0.2846(0.0022)		0.0050(0.0031)		0.0017(0.0031)		0.0036(0.0044)		0.1845(0.0005)
		EM	4,999	-0.0231(0.0022)		0.0023(0.0030)		0.0020(0.0030)		0.0040(0.0041)		0.0019(0.0008)
3	0	none	5,000	0.0011(0.0013)		-0.0012(0.0018)		-0.0006(0.0017)		0.0004(0.0025)		-0.0024(0.0005)
	20	DL/2	5,000	-0.2765(0.0022)		0.2188(0.0029)		0.2189(0.0029)		-0.1635(0.0037)		0.1141(0.0008)
		EM	5,000	-0.0097(0.0019)		0.0089(0.0023)		0.0092(0.0022)		-0.0087(0.0029)		-0.0019(0.0005)
	0	none	5,000	0.0012(0.0012)		-0.0007(0.0018)		-0.0007(0.0018)		0.0014(0.0025)		-0.0030(0.0005)
	60	DL/2	5,000	-0.1994(0.0007)		-0.0952(0.0022)		-0.0945(0.0022)		0.3158(0.0036)		0.1093(0.0007)
		EM	5,000	-0.2239(0.0033)		0.2062(0.0039)		0.2053(0.0038)		-0.1888(0.0046)		-0.0033(0.0008)

Table B-6.1.20. Simulated mean squared error (MSE) of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter					
				μ	τ_2	β_2	$(\tau\beta)_{22}$	σ	
				MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)
0	0	none	5,000	0.3383(0.0066)	0.6799(0.0141)	0.6532(0.0126)	1.3199(0.0259)	0.0607(0.0012)	
	20	DL/2	5,000	0.3524(0.0066)	0.7048(0.0143)	0.6780(0.0128)	1.3779(0.0269)	0.0543(0.0011)	
		EM	5,000	0.4483(0.0196)	0.8973(0.0317)	0.8354(0.0261)	1.7079(0.0455)	0.0816(0.0018)	
	0	none	5,000	0.3282(0.0067)	0.6534(0.0132)	0.6624(0.0136)	1.3349(0.0270)	0.0621(0.0012)	
	60	DL/2	4,577	0.2251(0.0057)	0.4094(0.0086)	0.4183(0.0089)	0.8228(0.0171)	0.0935(0.0015)	
		EM	4,577	1.1997(0.0529)	2.4261(0.0739)	2.2657(0.0684)	4.6485(0.1328)	0.1466(0.0034)	
1	0	none	5,000	0.3282(0.0064)	0.6751(0.0136)	0.6632(0.0132)	1.3250(0.0263)	0.0607(0.0011)	
	20	DL/2	5,000	0.2177(0.0049)	0.6000(0.0126)	0.5775(0.0116)	1.2822(0.0254)	0.0622(0.0012)	
		EM	5,000	1.5508(0.0479)	1.9327(0.0530)	1.9133(0.0531)	2.6280(0.0641)	0.0802(0.0017)	
	0	none	5,000	0.3387(0.0069)	0.6612(0.0130)	0.6726(0.0135)	1.2921(0.0260)	0.0608(0.0012)	
	60	DL/2	4,829	1.5047(0.0046)	1.2766(0.0121)	1.2770(0.0121)	1.4527(0.0241)	0.1405(0.0021)	
		EM	4,829	1.5694(0.0393)	2.3438(0.0504)	2.2642(0.0461)	3.5469(0.0917)	0.1572(0.0032)	

Table B-6.1.20 (contd). Simulated mean squared error (MSE) of parameter estimators with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter								
				μ		τ_2		β_2		$(\tau\beta)_{22}$		σ
				MSE(SE)		MSE(SE)		MSE(SE)		MSE(SE)		MSE(SE)
2	0	none	5,000	0.0208(0.0004)	0.0414(0.0008)	0.0428(0.0009)	0.0851(0.0017)	0.0037(0.0001)				
	20	DL/2	5,000	0.0758(0.0016)	0.1253(0.0025)	0.1262(0.0025)	0.2515(0.0051)	0.0399(0.0005)				
		EM	5,000	0.0276(0.0011)	0.0567(0.0019)	0.0551(0.0018)	0.1139(0.0033)	0.0052(0.0001)				
	0	none	5,000	0.0214(0.0004)	0.0426(0.0009)	0.0418(0.0009)	0.0853(0.0017)	0.0039(0.0001)				
	60	DL/2	4,585	0.1295(0.0020)	0.1326(0.0026)	0.1298(0.0025)	0.2642(0.0055)	0.0422(0.0004)				
		EM	4,585	0.1134(0.0033)	0.2379(0.0050)	0.2279(0.0049)	0.4644(0.0093)	0.0091(0.0002)				
3	0	none	5,000	0.0078(0.0002)	0.0157(0.0003)	0.0153(0.0003)	0.0307(0.0006)	0.0011(0.0000)				
	20	DL/2	5,000	0.1018(0.0013)	0.0882(0.0014)	0.0894(0.0014)	0.0913(0.0017)	0.0165(0.0002)				
		EM	5,000	0.0184(0.0009)	0.0262(0.0009)	0.0256(0.0009)	0.0411(0.0011)	0.0032(0.0001)				
	0	none	5,000	0.0214(0.0004)	0.0424(0.0008)	0.0416(0.0008)	0.0817(0.0017)	0.0040(0.0001)				
	60	DL/2	4,847	0.0452(0.0002)	0.0746(0.0013)	0.0710(0.0013)	0.2754(0.0049)	0.0207(0.0003)				
		EM	4,847	0.1071(0.0015)	0.1862(0.0028)	0.1815(0.0026)	0.2999(0.0061)	0.0091(0.0002)				

Table B-6.1.21. Simulated mean square error (MSE) of parameter estimators with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter								
				μ		τ_2		β_2		$(\tau\beta)_{22}$		σ
				MSE(SE)		MSE(SE)		MSE(SE)		MSE(SE)		MSE(SE)
0	0	none	5,000	0.1234(0.0024)	0.2473(0.0050)	0.2621(0.0053)	0.5219(0.0102)	0.0178(0.0004)				
	20	DL/2	5,000	0.1269(0.0025)	0.2534(0.0051)	0.2704(0.0054)	0.5398(0.0106)	0.0145(0.0003)				
		EM	5,000	0.1291(0.0026)	0.2608(0.0053)	0.2777(0.0057)	0.5566(0.0111)	0.0235(0.0005)				
	0	none	5,000	0.1211(0.0024)	0.2508(0.0051)	0.2501(0.0051)	0.5090(0.0102)	0.0173(0.0003)				
	60	DL/2	5,000	0.0882(0.0020)	0.1502(0.0031)	0.1511(0.0030)	0.3013(0.0061)	0.0673(0.0008)				
		EM	5,000	0.3014(0.0187)	0.5907(0.0266)	0.5518(0.0230)	1.1533(0.0386)	0.0571(0.0013)				
1	0	none	5,000	0.1248(0.0024)	0.2465(0.0048)	0.2456(0.0048)	0.4934(0.0097)	0.0179(0.0003)				
	20	DL/2	5,000	0.0881(0.0019)	0.2213(0.0044)	0.2207(0.0044)	0.4778(0.0095)	0.0188(0.0004)				
		EM	5,000	0.2819(0.0147)	0.4132(0.0165)	0.4050(0.0159)	0.6634(0.0199)	0.0240(0.0005)				
	0	none	5,000	0.1208(0.0024)	0.2495(0.0050)	0.2416(0.0047)	0.4900(0.0098)	0.0179(0.0004)				
	60	DL/2	5,000	1.4975(0.0025)	1.1799(0.0076)	1.1679(0.0076)	1.0054(0.0132)	0.0919(0.0010)				
		EM	5,000	1.6577(0.0297)	1.7544(0.0302)	1.7775(0.0295)	2.0771(0.0382)	0.0517(0.0011)				

Table B-6.1.21 (contd). Simulated mean square error (MSE) of parameter estimators with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Censor	Type	N	Parameter					
				μ		τ_2		β_2	
				MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)
2	0	none	5,000	0.0079(0.0002)	0.0156(0.0003)	0.0154(0.0003)	0.0308(0.0006)	0.0011(0.0000)	
	20	DL/2	5,000	0.0356(0.0007)	0.0462(0.0009)	0.0459(0.0009)	0.0914(0.0018)	0.0351(0.0003)	
		EM	5,000	0.0085(0.0002)	0.0168(0.0003)	0.0168(0.0004)	0.0333(0.0007)	0.0016(0.0000)	
	0	none	5,000	0.0081(0.0002)	0.0159(0.0003)	0.0159(0.0003)	0.0313(0.0006)	0.0011(0.0000)	
	60	DL/2	4,999	0.1054(0.0013)	0.0493(0.0010)	0.0483(0.0010)	0.0958(0.0019)	0.0353(0.0002)	
		EM	4,999	0.0253(0.0014)	0.0457(0.0018)	0.0443(0.0018)	0.0847(0.0028)	0.0036(0.0001)	
3	0	none	5,000	0.0079(0.0002)	0.0157(0.0003)	0.0151(0.0003)	0.0309(0.0006)	0.0011(0.0000)	
	20	DL/2	5,000	0.1012(0.0012)	0.0900(0.0014)	0.0893(0.0014)	0.0944(0.0017)	0.0160(0.0002)	
		EM	5,000	0.0173(0.0008)	0.0254(0.0009)	0.0248(0.0009)	0.0409(0.0011)	0.0014(0.0000)	
	0	none	5,000	0.0077(0.0002)	0.0163(0.0003)	0.0155(0.0003)	0.0320(0.0006)	0.0011(0.0000)	
	60	DL/2	5,000	0.0420(0.0002)	0.0338(0.0006)	0.0332(0.0006)	0.1657(0.0025)	0.0144(0.0001)	
		EM	5,000	0.1051(0.0009)	0.1182(0.0014)	0.1158(0.0013)	0.1414(0.0021)	0.0031(0.0001)	

Table B-6.1.22. Simulated bias of estimators of cell means, with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell					
				11		12		21	
				Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)
0	0	None	5,000	0.0002(0.0082)	0.0027(0.0082)	0.0007(0.0082)	0.0049(0.0082)		
	20	DL/2	5,000	-0.0289(0.0084)	-0.0271(0.0084)	-0.0284(0.0083)	-0.0225(0.0084)		
		EM	5,000	-0.0300(0.0095)	-0.0230(0.0093)	-0.0291(0.0095)	-0.0221(0.0094)		
	0	None	5,000	0.0085(0.0081)	0.0075(0.0082)	-0.0164(0.0081)	0.0024(0.0081)		
	60	DL/2	4,577	0.1596(0.0066)	0.1571(0.0066)	0.1419(0.0064)	0.1556(0.0065)		
		EM	4,577	-0.1027(0.0161)	-0.0867(0.0152)	-0.1316(0.0159)	-0.0850(0.0151)		
1	0	None	5,000	0.0122(0.0081)	0.0088(0.0083)	-0.0007(0.0082)	0.0040(0.0081)		
	20	DL/2	5,000	0.1209(0.0064)	-0.0138(0.0085)	-0.0229(0.0085)	0.0027(0.0081)		
		EM	5,000	-0.2796(0.0172)	-0.0033(0.0087)	-0.0130(0.0087)	0.0032(0.0081)		
	0	None	5,000	-0.0146(0.0082)	-0.0121(0.0082)	0.0022(0.0081)	-0.0056(0.0082)		
	60	DL/2	4,829	1.2217(0.0016)	0.1795(0.0060)	0.1791(0.0060)	-0.0055(0.0084)		
		EM	4,829	-0.3771(0.0172)	-0.2350(0.0172)	-0.2398(0.0172)	0.0034(0.0084)		

Table B-6.1.22 (contd). Simulated bias of estimators of cell means, with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell					
				11		12		21	
				Bias(SE)		Bias(SE)		Bias(SE)	
2	0	None	5,000	-0.0012(0.0020)		0.0028(0.0021)		0.0009(0.0021)	
	20	DL/2	5,000	-0.1138(0.0035)		-0.1067(0.0035)		-0.1123(0.0036)	
		EM	5,000	-0.0092(0.0023)		-0.0039(0.0023)		-0.0081(0.0024)	
	0	None	5,000	0.0007(0.0021)		-0.0021(0.0020)		-0.0041(0.0021)	
	60	DL/2	4,585	-0.2604(0.0037)		-0.2620(0.0036)		-0.2672(0.0036)	
		EM	4,585	-0.0754(0.0048)		-0.0758(0.0048)		-0.0857(0.0050)	
3	0	None	5,000	-0.0003(0.0013)		0.0007(0.0012)		-0.0037(0.0012)	
	20	DL/2	5,000	-0.2788(0.0022)		-0.0586(0.0019)		-0.0652(0.0019)	
		EM	5,000	-0.0124(0.0019)		-0.0004(0.0013)		-0.0047(0.0013)	
	0	None	5,000	0.0002(0.0021)		0.0025(0.0020)		-0.0013(0.0020)	
	60	DL/2	4,847	-0.1951(0.0012)		-0.2854(0.0034)		-0.2904(0.0034)	
		EM	4,847	-0.2350(0.0033)		-0.0994(0.0048)		-0.1102(0.0049)	

Table B-6.1.23. Simulated bias of estimators of cell means, with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell							
				11		12		21		22	
				Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)		
0	0	None	5,000	0.0070(0.0050)	0.0004(0.0051)	-0.0003(0.0050)	0.0013(0.0050)				
	20	DL/2	5,000	-0.0180(0.0050)	-0.0261(0.0052)	-0.0271(0.0051)	-0.0260(0.0051)				
		EM	5,000	0.0015(0.0051)	-0.0085(0.0053)	-0.0085(0.0052)	-0.0072(0.0051)				
	0	None	5,000	0.0004(0.0049)	0.0042(0.0050)	0.0005(0.0050)	-0.0030(0.0050)				
	60	DL/2	5,000	0.1218(0.0038)	0.1219(0.0039)	0.1192(0.0039)	0.1166(0.0039)				
		EM	5,000	-0.0274(0.0078)	-0.0305(0.0078)	-0.0373(0.0081)	-0.0365(0.0077)				
1	0	None	5,000	-0.0032(0.0050)	0.0063(0.0050)	0.0107(0.0049)	-0.0023(0.0050)				
	20	DL/2	5,000	0.1131(0.0039)	-0.0163(0.0051)	-0.0118(0.0051)	-0.0036(0.0050)				
		EM	5,000	-0.0421(0.0075)	0.0027(0.0050)	0.0073(0.0050)	-0.0027(0.0050)				
	0	None	5,000	-0.0008(0.0049)	0.0097(0.0049)	0.0006(0.0050)	0.0042(0.0050)				
	60	DL/2	5,000	1.2218(0.0010)	0.1742(0.0036)	0.1682(0.0036)	-0.0207(0.0052)				
		EM	5,000	-0.7114(0.0152)	-0.0439(0.0086)	-0.0592(0.0091)	-0.0036(0.0051)				

Table B-6.1.23 (contd). Simulated bias of estimators of cell means, with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell		
				11	12	21
				Bias(SE)	Bias(SE)	Bias(SE)
2	0	None	5,000	-0.0012(0.0013)	0.0002(0.0012)	0.0009(0.0012)
	20	DL/2	5,000	-0.1113(0.0022)	-0.1112(0.0022)	-0.1097(0.0021)
		EM	5,000	-0.0044(0.0013)	-0.0039(0.0013)	-0.0033(0.0013)
	0	None	5,000	-0.0019(0.0013)	-0.0013(0.0012)	0.0010(0.0013)
	60	DL/2	4,999	-0.2846(0.0022)	-0.2829(0.0021)	-0.2796(0.0022)
		EM	4,999	-0.0231(0.0022)	-0.0210(0.0021)	-0.0208(0.0022)
3	0	None	5,000	0.0011(0.0013)	0.0005(0.0012)	-0.0000(0.0013)
	20	DL/2	5,000	-0.2765(0.0022)	-0.0576(0.0019)	-0.0577(0.0019)
		EM	5,000	-0.0097(0.0019)	-0.0005(0.0012)	-0.0008(0.0013)
	0	None	5,000	0.0012(0.0012)	0.0005(0.0013)	0.0005(0.0013)
	60	DL/2	5,000	-0.1994(0.0007)	-0.2938(0.0021)	-0.2946(0.0021)
		EM	5,000	-0.2239(0.0033)	-0.0187(0.0023)	-0.0177(0.0022)

Table B-6.1.24. Simulated mean squared error (MSE) of estimators of cell means with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell							
				11		12		21		22	
				MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)
0	0	None	5,000	0.3383(0.0066)	0.3383(0.0067)	0.3355(0.0068)	0.3363(0.0066)				
	20	DL/2	5,000	0.3524(0.0066)	0.3531(0.0066)	0.3485(0.0068)	0.3495(0.0065)				
		EM	5,000	0.4483(0.0196)	0.4296(0.0172)	0.4507(0.0222)	0.4434(0.0196)				
	0	None	5,000	0.3282(0.0067)	0.3328(0.0069)	0.3311(0.0066)	0.3248(0.0063)				
	60	DL/2	4,577	0.2251(0.0057)	0.2220(0.0055)	0.2085(0.0049)	0.2152(0.0050)				
		EM	4,577	1.1997(0.0529)	1.0608(0.0426)	1.1774(0.0488)	1.0511(0.0418)				
1	0	None	5,000	0.3282(0.0064)	0.3442(0.0069)	0.3384(0.0069)	0.3286(0.0066)				
	20	DL/2	5,000	0.2177(0.0049)	0.3654(0.0073)	0.3632(0.0073)	0.3312(0.0067)				
		EM	5,000	1.5508(0.0479)	0.3820(0.0135)	0.3754(0.0126)	0.3301(0.0066)				
	0	None	5,000	0.3387(0.0069)	0.3378(0.0068)	0.3285(0.0064)	0.3355(0.0066)				
	60	DL/2	4,829	1.5047(0.0046)	0.2071(0.0053)	0.2077(0.0051)	0.3425(0.0067)				
		EM	4,829	1.5694(0.0393)	1.4865(0.0507)	1.4864(0.0515)	0.3389(0.0076)				

Table B-6.1.24 (contd). Simulated mean squared error (MSE) of estimators of cell means with 3 replicates per cell, DL/2 and adjusted MLE (EM).

Cell											
				11		12		21		22	
Distn	Cens	Type	N	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)	MSE(SE)
2	0	None	5,000	0.0208(0.0004)	0.0216(0.0004)	0.0215(0.0004)	0.0208(0.0004)	0.0215(0.0004)	0.0208(0.0004)	0.0208(0.0004)	0.0208(0.0004)
	20	DL/2	5,000	0.0758(0.0016)	0.0740(0.0016)	0.0761(0.0017)	0.0749(0.0017)	0.0761(0.0017)	0.0749(0.0017)	0.0749(0.0017)	0.0749(0.0017)
		EM	5,000	0.0276(0.0011)	0.0273(0.0012)	0.0299(0.0013)	0.0293(0.0014)	0.0299(0.0013)	0.0293(0.0014)	0.0293(0.0014)	0.0293(0.0014)
	0	None	5,000	0.0214(0.0004)	0.0200(0.0004)	0.0211(0.0004)	0.0206(0.0004)	0.0211(0.0004)	0.0206(0.0004)	0.0206(0.0004)	0.0206(0.0004)
	60	DL/2	4,585	0.1295(0.0020)	0.1280(0.0020)	0.1324(0.0020)	0.1272(0.0020)	0.1324(0.0020)	0.1272(0.0020)	0.1272(0.0020)	0.1272(0.0020)
		EM	4,585	0.1134(0.0033)	0.1115(0.0033)	0.1201(0.0034)	0.1094(0.0031)	0.1201(0.0034)	0.1094(0.0031)	0.1094(0.0031)	0.1094(0.0031)
3	0	None	5,000	0.0078(0.0002)	0.0078(0.0002)	0.0078(0.0002)	0.0079(0.0002)	0.0078(0.0002)	0.0079(0.0002)	0.0079(0.0002)	0.0079(0.0002)
	20	DL/2	5,000	0.1018(0.0013)	0.0215(0.0005)	0.0225(0.0005)	0.0083(0.0002)	0.0225(0.0005)	0.0083(0.0002)	0.0083(0.0002)	0.0083(0.0002)
		EM	5,000	0.0184(0.0009)	0.0081(0.0002)	0.0081(0.0002)	0.0079(0.0002)	0.0081(0.0002)	0.0079(0.0002)	0.0079(0.0002)	0.0079(0.0002)
	0	None	5,000	0.0214(0.0004)	0.0203(0.0004)	0.0205(0.0004)	0.0205(0.0004)	0.0205(0.0004)	0.0205(0.0004)	0.0205(0.0004)	0.0205(0.0004)
	60	DL/2	4,847	0.0452(0.0002)	0.1380(0.0019)	0.1417(0.0019)	0.0515(0.0012)	0.1417(0.0019)	0.0515(0.0012)	0.0515(0.0012)	0.0515(0.0012)
		EM	4,847	0.1071(0.0015)	0.1233(0.0029)	0.1293(0.0030)	0.0214(0.0006)	0.1293(0.0030)	0.0214(0.0006)	0.0214(0.0006)	0.0214(0.0006)

Table B-6.1.25. Simulated mean squared error (MSE) of estimators of cell means with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell					
				11		12		21	
				MSE(SE)		MSE(SE)		MSE(SE)	
0	0	None	5,000	0.1234(0.0024)		0.1286(0.0025)		0.1272(0.0026)	
	20	DL/2	5,000	0.1269(0.0025)		0.1340(0.0026)		0.1322(0.0026)	
		EM	5,000	0.1291(0.0026)		0.1390(0.0029)		0.1369(0.0029)	
	0	None	5,000	0.1211(0.0024)		0.1243(0.0025)		0.1270(0.0025)	
	60	DL/2	5,000	0.0882(0.0020)		0.0894(0.0020)		0.0890(0.0020)	
		EM	5,000	0.3014(0.0187)		0.3054(0.0165)		0.3305(0.0206)	
1	0	None	5,000	0.1248(0.0024)		0.1226(0.0025)		0.1205(0.0024)	
	20	DL/2	5,000	0.0881(0.0019)		0.1306(0.0026)		0.1284(0.0026)	
		EM	5,000	0.2819(0.0147)		0.1258(0.0025)		0.1239(0.0026)	
	0	None	5,000	0.1208(0.0024)		0.1214(0.0024)		0.1263(0.0025)	
	60	DL/2	5,000	1.4975(0.0025)		0.0955(0.0021)		0.0940(0.0021)	
		EM	5,000	1.6577(0.0297)		0.3716(0.0196)		0.4136(0.0221)	
	0	None	5,000						
	20	DL/2	5,000						
		EM	5,000						

Table B-6.1.25 (cont'd). Simulated mean squared error (MSE) of estimators of cell means with 8 replicates per cell, DL/2 and adjusted MLE (EM).

Distn	Cens	Type	N	Cell					
				11		12		21	
				MSE(SE)	N	MSE(SE)	N	MSE(SE)	N
2	0	None	5,000	0.0079(0.0002)		0.0077(0.0002)		0.0078(0.0002)	
	20	DL/2	5,000	0.0356(0.0007)		0.0357(0.0007)		0.0351(0.0007)	
		EM	5,000	0.0085(0.0002)		0.0085(0.0002)		0.0085(0.0002)	
	0	None	5,000	0.0081(0.0002)		0.0078(0.0002)		0.0078(0.0002)	
	60	DL/2	4,999	0.1054(0.0013)		0.1031(0.0012)		0.1026(0.0013)	
		EM	4,999	0.0253(0.0014)		0.0224(0.0012)		0.0249(0.0013)	
3	0	None	5,000	0.0079(0.0002)		0.0075(0.0001)		0.0078(0.0002)	
	20	DL/2	5,000	0.1012(0.0012)		0.0207(0.0005)		0.0209(0.0005)	
		EM	5,000	0.0173(0.0008)		0.0078(0.0002)		0.0080(0.0002)	
	0	None	5,000	0.0077(0.0002)		0.0079(0.0002)		0.0081(0.0002)	
	60	DL/2	5,000	0.0420(0.0002)		0.1088(0.0012)		0.1091(0.0012)	
		EM	5,000	0.1051(0.0009)		0.0260(0.0013)		0.0254(0.0012)	

Under the null distributions (Distributions 0 and 2), the following observations hold.

- There is no bias for the (zero-valued) treatment parameters, and the biases for the overall mean μ are consistent with the discussion of the one-sample problem in Section B-4. These conclusions can be seen by looking at the bias for μ in Tables B-6.1.18 and B-6.1.19, or at the biases for the (theoretically equal) cell means in Tables B-6.1.22 and B-6.1.23.
- It was noted earlier that the adjusted MLE does not maintain a Type I error of 0.05. This now is seen to result not from biases of the parameter estimators, but from inflated MSEs of the parameter estimators, resulting from inflated variances of the parameter estimators, which in turn result from the cell-mean estimators having variances greater than $\sigma^2/(\text{no. of replicates})$. The usual ANOVA formulas do not account for the greater variances resulting from censoring.

Under the Distributions 1 and 3, the following observations can be made.

- Consider possible reasons for the false rejection of $(\tau\beta)_{22} = 0$. The DL/2 method biases any estimated cell mean depending on the fraction censored for the cell and on the within-cell standard deviation. Cell 11 has the smallest true mean, and therefore the largest fraction censored, while Cell 22 has the smallest fraction censored. The different censoring of the cells can introduce a curvature into the estimated response surface, and a bias in the estimate of $(\tau\beta)_{22}$. At high censoring with many replicates, the bias becomes an important contributor to the rejection of $(\tau\beta)_{22} = 0$ for the DL/2 test; this conclusion is based on a comparison of the tabulated bias² and the MSE of the estimators of $(\tau\beta)_{22}$. For the adjusted MLE, the comparison of the bias² and the MSE of the estimators of $(\tau\beta)_{22}$ shows that the bias is never an important contributor to the rejection of $(\tau\beta)_{22} = 0$.

- At 20% censoring there is very little bias, and the two methods perform equally well. At 60% censoring, the MLE parameter estimators tend to be less biased than those from the DL/2 method, although the DL/2 estimators of the main effect parameters are less biased than the MLE estimators when $\sigma = 0.25$.
- Consider comparisons of the methods based on the MSEs of the cell mean estimators. The DL/2 method is superior to the MLE, except when $\sigma = 1$ and there are 8 replicates per cell.
- For estimating cell means, the adjusted MLE tends to have negative bias, while the bias of the DL/2 method depends on σ .

B-6.1.4 Conclusions

The LR method cannot be used if any cell contains only censored data. This is a serious restriction, making the LR method unusable for many data sets. The MLE requires substantial programming. A simple replacement method such as DL/2 or uniform replacement is therefore preferable to the LR or adjusted MLE, if it performs acceptably. The comparisons below consider the DL/2 method, uniform replacement, and the adjusted MLE.

If the purpose of the analysis is to test for effects, as is usual in analysis of variance analyses, power (including Type I error) should be the basis of comparisons. The DL/2 method was the only method that maintained the correct Type I error of 0.05 under censoring as high as 60%. Therefore, DL/2 is the method of choice.

When main effects are present, but no interaction is present, all the methods considered tend to falsely reject the no-interaction hypothesis with probability greater than 0.05. This is apparently an inherent feature of the two-way layout with censoring.

If the purpose of the analysis is to estimate cell means rather than to test for effects, the comparisons should be based primarily on the MSEs of the

cell mean estimators. Under the null distributions, so that all the cells are censored equally, the DL/2 method is better than the MLE when $\sigma = 1.0$, and worse when $\sigma = 0.25$. Under the alternate distributions, there is no clear better method: the method that is better for one cell can be worse for a different cell, and the relative performances depend on σ . In terms of bias rather than MSE, the MLE typically has smaller bias than the DL/2 estimate.

B-6.1.5 SAS Programs for Data Generation and ANOVA Simulation

The following SAS programs were used to generate the data sets and produce the results tabulated above.

PROGRAM FOR DATA GENERATION AND ANOVA SIMULATION

```
*****
*****
Variables changed and sections modified for simulation runs are indicated in
bold italicized print. Comments are preceded with asterisks.
*****
*****
```

*Generate data:

```
  k 2x2 data sets with r reps
  cell means t, u, and v
  xnn is censoring cutoff value and represents the nth
    percentile such that sum censored from N(0,1), N(t,1),
    N(u,1), and N(v,1) is n/100
  screen data sets using 'drop2.sas' (by cell)
    or 'drop.sas' (by data set)
```

```
;
```

```
libname erdp 'c:\ltderdp\test';
```

```
data gendat;
```

```
  r=8; k=2000; dat=1; flag=1; %let seed=160589;
  t=1.5; u=3; v=1.5;
  do i=1 to k*r;
    y1=rannor(&seed);
    y2=t+rannor(&seed);
    y3=u+rannor(&seed);
    y4=v+rannor(&seed);
    if i/r<=r*dat then dat=flag;
    if mod(i,r)=0 then flag=flag+1;
```

```

        output; end;
drop k flag t u v i r;
data one; set gendat;
        keep y well depth dat;
        well='a'; depth=5; y=y1;

data two; set gendat;
        keep y well depth dat;
        well='a'; depth=10; y=y2;

data three; set gendat;
        keep y well depth dat;
        well='b'; depth=10; y=y3;

data four; set gendat;
        keep y well depth dat;
        well='b'; depth=5; y=y4;

data gendat; set one two three four;
        xnn=0.2533;

%include 'cellset2.sas';

%include 'clear.sas';
*eliminate data sets with <2 uncensored values in each cell;
proc sort data=gendat; by dat cell;
proc means data=gendat noprint n;var y; by dat;
        output out=allcnt n=allcnt;
data drop;set gendat;if y>=xnn then flag=1;else flag=0;
        %include 'drop2.sas';
data gendat;merge gendat build;if dat ne drop;
        drop _TYPE_ _FREQ_ drop;

data build;set clear;

*Perform censored value replacement:
        replacement section changes according to replacement technique simulated;

data gendat; set gendat;cut=xnn;drop xnn;
        %include 'nr.sas';
        data gendat;set built one;
        drop nscore cut;

data one;set clear;
data built;set clear;

*Produce anova statistics:
        output file names and ems change according to simulation;

```

```

proc sort data=gendat;
  by dat ;
proc anova outstat=simulate noprint;
  class well depth; by dat;
  model y=well depth well*depth;run;
data gendat;set clear;
data simulate;set simulate;
  ms=ss/df;
  if prob<=0.05 then signif=1;else signif=0;
  if prob=. then signif=0;
  drop _NAME_ _TYPE_ prob ss f df;

proc sort data=simulate;by _SOURCE_;
proc means noprint sum;var signif;by _SOURCE_; output out=erdp.sig6nr20
  sum=count;
proc univariate data=simulate noprint ; var ms; by _source_;
  output out=erdp.std6nr20 n=count mean=meanms std=stdms var=varms;

data erdp.sts6nr20;set erdp.sts6nr20;
  if _source_ ='DEPTH' then ems=18;
  if _source_ ='WELL' then ems=18;
  if _source_ ='WELL*DEPTH' then ems=1;
  if _source_ ='ERROR' then ems=1;
  relbias=(meanms-ems)/ems;
  relmse=(varms+(meanms-ems)**2)/(ems**2);

```

PROGRAMS CALLED

*Program: cellset2.sas;

```
if well='a' and depth=5 then cell=1;
if well='a' and depth=10 then cell=2;
if well='b' and depth=10 then cell=3;
if well='b' and depth=5 then cell=4;
```

*Program: drop2.sas;

```
*count number uncensored within cells (from drop);
proc means noprint sum; var flag; by dat cell;
  output out=cnt sum=cntuncen;
data drop;set clear;
*indicate if cell uncensored count is sufficient;
data cnt;set cnt;
  if cntuncen<2 then drop=1;else drop=0;
*indicate data set with one or more insufficient cells -
  first - count the number of cells with insufficient uncensored
          counts within data sets;
proc means noprint sum; var drop; by dat;
  output out=sumdrop sum=cellcnt;
  *second - mark the data set if one or more cells within a data set
          are insufficient;
data sumdrop;set sumdrop;
  if cellcnt > 0 then drop = dat;
*build up (by number of cells) to gendat size for drop;
data build;merge sumdrop allcnt; by dat;
  do i=1 to allcnt;
    output; end;
  drop allcnt cellcnt i;
run;
```

*Program: drop.sas;

```
proc means noprint n; var y; by dat;
  output out=cnt n=cntuncen;
data build;merge cnt allcnt; by dat;
  do i=1 to allcnt;
    if cntuncen<2 then drop=dat;else drop=.;
    output; end;
  drop allcnt cntuncen i;
```

*Program: clear.sas;

```
data clear;input x;cards;
1
;
data one;set clear;
data two;set clear;
data three;set clear;
data four;set clear;
```

*Program: nr.sas;

```
proc rank out=gendat normal=vw; var y; ranks nscore;
by dat cell;
data one;set gendat; if y>=cut;
data two; set gendat; if y lt cut;
data gendat;set clear;
proc reg data=one outest=est noprint; by dat cell;
model y=nscore;
data est; set est; beta=nscore; alpha=intercep;
keep beta alpha dat cell;
proc means data=two noprint n ; var y; by dat cell;
output out=number n=censnumb;
data new; merge number est; by dat cell;
do i=1 to censnumb; output; end;
drop i censnumb;
data number;set clear;
data est;set clear;
data built; merge two new;
y=alpha +(beta*nscore);
drop alpha beta _FREQ_ _TYPE_;
data two;set clear;
data new;set clear;
*dm 'log;*clear';
```

*Program: uniform.sas;

```
proc means noprint n; var y; by cell dat;
output out=replace n=ltdcnt;
proc means data=censored noprint mean; var dl;by cell dat;
output out=dlkeep mean=dl;
data replace;merge replace dlkeep;by cell dat;flag=1;dl=exp(dl);
do i=1 to ltdcnt;
yr=(dl*flag)/(ltdcnt + 1);output;
flag=flag+1;
end;
data replace;merge censored replace;by cell dat;
y=log(yr);
drop ltdcnt flag i yr _TYPE_ _FREQ_ dl;
run;
```


PROGRAM TO COMBINE AND TABULATE RESULTS

Permanent SAS data sets and ems values modified as required to process results. Comments are preceded with asterisks.

*Program: combrsts.sas;

libname erdp 'c:\ltderdp\test';

*calculate alpha;

data erdp.signif0;set erdp.signif0;alpha=count/_freq_;

data erdp.signif6;set erdp.signif6;alpha=count/_freq_;

data erdp.sig0un60;set erdp.sig0un60;alpha=count/_freq_;

data erdp.sig6un60;set erdp.sig6un60;alpha=count/_freq_;

data erdp.sig6hf60;set erdp.sig6hf60;alpha=count/_freq_;

data erdp.sig0un20;set erdp.sig0un20;alpha=count/_freq_;

data erdp.sig6un20;set erdp.sig6un20;alpha=count/_freq_;

data erdp.sig6hf20;set erdp.sig6hf20;alpha=count/_freq_;

data erdp.sig0hf20;set erdp.sig0hf20;alpha=count/_freq_;

*create data sets for distn 2;

data rst6none;merge erdp.statsno6 erdp.signif6;by _source_;

distn=2;

type='none';

censor=0;

data rst6un60;merge erdp.sts6un60 erdp.sig6un60;by _source_;

distn=2;

type='un';

censor=60;

data rst6un20;merge erdp.sts6un20 erdp.sig6un20;by _source_;

distn=2;

type='un';

censor=20;

data rst6hf60;merge erdp.sts6hf60 erdp.sig6hf60;by _source_;

distn=2;

type='hf';

censor=60;

data rst6hf20;merge erdp.sts6hf20 erdp.sig6hf20;by _source_;

distn=2;

type='hf';

censor=20;

*create data sets for distn 0;

data rst0none;merge erdp.statsno0 erdp.signif0;by _source_;

distn=0;

type='none';

censor=0;

data rst0un60;merge erdp.sts0un60 erdp.sig0un60;by _source_;

distn=0;

type='un';

```

    censor=60;
data rst0un20;merge erdp.sts0un20 erdp.sig0un20;by _source_;
    distn=0;
    type='un';
    censor=20;
data rst0hf60;merge erdp.sts0hf60 erdp.sig0hf60;by _source_;
    distn=0;
    type='hf';
    censor=60;
data rst0hf20;merge erdp.sts0hf20 erdp.sig0hf20;by _source_;
    distn=0;
    type='hf';
    censor=20;

*combine all data sets, calculate additional statistics, and set up formats;
data erdp.all;set
    rst6none
    rst6un60
    rst6un20
    rst6hf60
    rst6hf20
    rst0none
    rst0un60
    rst0un20
    rst0hf60
    rst0hf20
;
    stderr=stdms/sqrt(_freq_);
    stdalpha=sqrt(alpha*(1-alpha)/_freq_);
if _source_='ERROR' then alpha=.;
if _source_='ERROR' then _source_='Z';
if _source_='WELL*DEPTH' then _source_='X';

if distn=2 and _source_='WELL' then ems=7.75;
if distn=2 and _source_='DEPTH' then ems=7.75;
    else ems=1;
rberror=stderr/((sqrt(_freq_))*ems);
rbupci=relbias + (2*rberror);
rblowci=relbias - (2*rberror);

options nonumber nodate ps=44 ls=155;
proc sort;by _source_;
proc format;
    value $_source_ 'DEPTH'='A'
                    'WELL'='B'
                    'Z'='ERROR'
                    'X'='A*B';

*tabulate results;
proc tabulate data=erdp.all;
    class censor type _freq_ _source_ distn;
    var alpha relbias stderr varms stdalpha rberror rbupci rblowci relmse;

```

```
table distn*censor* type*_freq_,_source_*(alpha stdalpha) *f=8.4;  
table distn*censor* type*_freq_,_source_*(relbias rberror)*f=8.4;  
table distn*censor* type*_freq_,_source_*(rbupci rblowci)*f=8.4;  
table distn*censor* type*_freq_,_source_*(stderr varms);  
table distn*censor* type*_freq_,_source_*(relmse);  
keylabel sum='';  
label _freq_ = 'N';  
label stdalpha = 'STDERROR';  
label alpha = 'POWER';  
label rbupci = 'UPPER LIMIT';  
label rblowci = 'LOWER LIMIT';  
format _source_ $_source_.;  
run;
```

REFERENCES

- J. Aitchison and J. A. C. Brown, *The Lognormal Distribution with Special Reference to Its Use in Economics*, Cambridge: Cambridge University Press, 1957.
- M. Aitkin, "A Note on the Regression Analysis of Censored Data," *Technometrics*, **23**, 1981, pp. 161-163.
- M. M. Bernarie, "The Validity of the Log-normal Distribution of Pollutant Concentrations," Paper SU-18D, *Proceedings of the 2nd International Clean Air Conference*, (CE-Trans--7589), 1971.
- Larry G. Blackwood, "Estimating Population Means and Standard Deviations from Censored Samples: An Evaluation of Gupta's Method," unreleased report, EG&G Idaho Inc.: Idaho Falls, ID, 1989.
- Larry G. Blackwood, *The Lognormal Distribution, Environmental Data, and Radiological Monitoring*, EGG-SARE-9243, September 1990.
- L. G. Blackwood, "The Quality of Mean and Variance Estimates for Normal and Lognormal Data when the Underlying Distribution is Misspecified," *Journal of Chemometrics* (to appear, 1991).
- C. Andrew Clayton, John W. Hines, and Phyllis D. Elkins, "Detection Limits with Specified Assurance Probabilities," *Analytical Chemistry*, **59**, 20, 1987, pp. 2506-2514.
- Timothy A. Cohn, "Adjusted Maximum Likelihood Estimation of the Moments of Lognormal Populations from Type 1 Censored Samples," *U. S. Geol. Survey Open File Rep. 88-350*, 1988.
- E. L. Crow, "Applications in Atmospheric Sciences," Chapter 13 in *Lognormal Distributions: Theory and Applications*, (E. L. Crow and K. Shimizu, eds.), New York: Marcel Dekker, Inc., 1988.
- E. L. Crow and K. Shimizu (eds.), *Lognormal Distributions: Theory and Applications*, New York: Marcel Dekker, Inc., 1988.
- Lloyd A. Currie, "Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry," *Analytical Chemistry*, **40**, 3, 1968, pp. 586-593.
- Lloyd A. Currie, *Detection in Analytical Chemistry: Importance, Theory and Practice*, Washington, DC: American Chemical Society, 1988.
- A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," and discussion, *Journal of the Royal Statistical Society, Series B*, **39**, 1977, pp. 1-38.

B. Dennis and G. P. Patil, "Applications in Ecology," Chapter 12 in *Lognormal Distributions: Theory and Applications*, (E. L. Crow and K. Shimizu, eds.), New York: Marcel Dekker, Inc., 1988.

L. L. Eberhardt and R. O. Gilbert, *Gamma and lognormal distributions as models in studying food-chain kinetics*, Battelle, Pacific Northwest Laboratories, BNWL-1747, 1973.

L. L. Eberhardt and R. O. Gilbert, "Statistics and Sampling in Transuranic Studies," in *Transuranic Elements in the Environment* (W. C. Hanson, ed.), DOE/TIC-22800 NTIS, 1980, pp. 173-186.

P. G. Georgopoulos and J. H. Seinfeld, "Statistical Distributions of Air Quality Concentrations," *Environmental Science and Technology*, 16, 1982, 401-416A.

R. O. Gilbert, *Statistical Methods for Environmental Pollution Monitoring*, New York: Van Nostrand Reinhold Co., 1987.

Robert J. Gilliom and Dennis R. Helsel, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data: 1. Estimation Techniques," *Water Resources Research*, 22, 1986, pp. 147-155.

John A. Glaser, Denis L. Foerst, Gerald D. McKee, Stephan A. Quave, and William L. Budde, "Trace Analyses for Wastewaters," *Environmental Science & Technology*, 15, 12, 1981, pp. 1426-1435.

A. K. Gupta, "Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample," *Biometrika*, 39, 1952, pp. 260-273.

Myles Hollander and Douglas A. Wolfe, *Nonparametric Statistical Methods*, New York: John Wiley & Sons, 1973.

T. R. Johnson and M. J. Symons, "Extreme Values of Weibull and Lognormal Distributions Fitted to Ambient Air Quality Data," presented at the 73rd Annual Meeting of the Air Pollution Control Association, Montreal, Quebec, 1980.

William J. Kennedy, Jr., and James E. Gentle, *Statistical Computing*, New York: Marcel Dekker, Inc., 1980.

G. S. Koch and R. F. Link, *Statistical Analysis of Geological Data*, New York: Dover, 1981.

Diane Lambert, Bruce Peterson, and Irma Terpenning, "Nondetects, Detection Limits, and the Probability of Detection," *Journal of the American Statistical Association*, 86, 414, 1991, pp. 266-277.

S. P. Millard and S. J. Deveral, "Nonparametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Nondetect Limits," *Water Resources Research*, 24, 12, 1988, pp. 2087-2098.

John Neter and William Wasserman, 1974, *Applied Linear Statistical Models*, Homewood, IL: Richard D. Irwin, Inc., 842 pp.

A. E. Sarhan and B. G. Greenberg, *Contributions to Order Statistics*, New York: John Wiley and Sons, 1962.

SAS Institute Inc., *SAS Language Guide for Personal Computers, Release 6.03 Edition*, Cary, NC: SAS Institute Inc., 1988a.

SAS Institute Inc., *SAS Procedures Guide, Release 6.03 Edition*, Cary, NC: SAS Institute Inc., 1988b.

SAS Institute Inc., *SAS/STAT User's Guide, Release 6.03 Edition*, Cary, NC: SAS Institute Inc., 1988c.

J. Schmee and G. J. Hahn, "A Simple Method for Regression Analysis with Censored Data," *Technometrics*, 21, 4, 1979. pp. 417-432.

K. Shimizu, K., "Point Estimation," Chapter 2 in *Lognormal Distributions: Theory and Applications*, (E. L. Crow and K. Shimizu, eds.), New York: Marcel Dekker, Inc., 1988.

K. Shimizu and E. L. Crow, "History, Genesis, and Properties," Chapter 1 in *Lognormal Distributions: Theory and Applications*, (E. L. Crow and K. Shimizu, eds.), New York: Marcel Dekker, Inc., 1988.

D. R. Speer and D. A. Waite, *Statistical Distributions as Applied to Environmental Surveillance Data*, Battelle, Pacific Northwest Laboratories, BNWL-SA-5482, 1975.

Statistical Sciences, Inc., *S-PLUS*, Seattle, WA, 1990.

R. G. D. Steel and J. H. Torrie, *Principles and Procedures of Statistics, Second Edition*, New York: McGraw-Hill, 1980.

Technology Partners, Inc., *S/PC*, Orem, UT, 1988.

M. S. Wolynetz, "Algorithm AS 138. Maximum Likelihood Estimation from Confined and Censored Normal Data," *Applied Statistics*, 28, 1979a, pp. 185-195.

M. S. Wolynetz, "Algorithm AS 139. Maximum Likelihood Estimation in a Linear Model from Confined and Censored Normal Data," *Applied Statistics*, 28, 1979b, pp. 195-214.

END

**DATE
FILMED**

01/106/92