1-1-1994

Use of log normal transformation in environmental statistics

Sally L Stewart
University of Nevada, Las Vegas

USE OF LOG NORMAL TRANSFORMATIONS

IN

ENVIRONMENTAL STATISTICS

by

SALLY L. STEWART

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Mathematics

Department of Mathematics
University of Nevada, Las Vegas

December, 1994

UMI Number: 1361103

The Thesis of Sally L. Stewart for the degree of Master of Science in Mathematics is approved.

Ashok K. Singh

Chairperson, Ashok K. Singh, Ph. D.

Malwane Ananda

Examining Committee Member, Malwane Ananda, Ph. D.

Examining Committee Member, Rohan Dalpatadu, Ph. D.

Graduate Faculty Representative, Sidkazem Taghva, Ph. D.

Cheryl L. Bowles

Dean of the Graduate College, Cheryl L. Bowles, Ed. D.

University of Nevada, Las Vegas
December, 1994

ABSTRACT

The log normal transformation is commonly used in the analysis of environmental data. The sample histogram of observed contaminant concentrations from a Superfund site typically appears to be log normal and the concentration data is log-transformed so that the classical statistical methods based on normal distribution can be used. USEPA guidance documents on statistical evaluation of attainment of cleanup standards for soils suggest using the log normal transformation in case the contaminant concentration data appears to be log normal. There are two basic problems with using a transformation in data analysis:

i) interpretation of results, and

ii) in transforming a formula based on the assumption of normality of the data so that it can be applied to transformed data.

The present thesis will address the second problem associated with the log transformation. In addition, the performance of some of the common normal-theory based procedures applied on original concentration data when the data distribution is in fact log normal will be investigated. Real Superfund site characterization data and simulated data will be used to provide examples.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Ashok K. Singh, for suggesting this topic, and supporting me throughout my studies. I am grateful for the opportunities which have come my way related to my area of study.

I would also like to thank Mr. Ken Brown of United States Envorinmental Protection Agency, for his support and help in using real-life examples from past EPA work.

I am thankful for the EBSCoR Women in Science Grant which funded my last year of study. It offered me additional time to devote to research and learning. In addition, it funded my trip to the Conference of Environmetrics in Burlington, Canada where I presented my thesis. I especially want to thank Ellen Jacobson, Director of EBSCoR, for her guidance and advice.

Lastly, I would like to thank John for all his support, encouragement, and patience throughout the year.

CHAPTER 1

INTRODUCTION TO THE LOG NORMAL DISTRIBUTION

HISTORY

The theory of log normal distribution appears to have been first introduced by D. McAlister in his memoir presented to the Royal Society of London in 1879 **[14]** in which he gave expressions for the mean, median, mode and the second moment of the distribution. The memoir was presented by Francis Galton, who originally suggested the study. In his opening remarks **[7]**, Galton expressed the view that in certain cases the geometric mean is to be preferred to the arithmetic mean as a measure of location. His assumption lies at the basis of the well-known law of 'Frequency of Error' which he believes to be incorrectly applied to many social phenomena. In 1903, the next advance was made by J. C. Kapteyn **[10]** in which he established clearer genesis of the distribution and described a machine for generating samples from a log normal population similar to Galton for normal populations.

Kapteyn's theory on the genesis of the log normal distribution is based on the law of proportionate effect which states that a change in the variate at any step of the process is a random proportion of a function $\phi(X_{j-1})$ of the value X_{j-1} already attained **[10]**. In other words, suppose that the variate is initially X_0 and that after the jth step in the process it is X_j; the final value is X_n. Then the general case suggested by Kapteyn is $X_j - X_{j-1} = e_j\phi(X_{j-1})$. However, the special case $\phi(X) = X$ (the change of the variate is a random proportion of the momentary value of the variate) of proportionate effect would reduce to $X_j - X_{j-1} = e_jX_{j-1}$. The connection between this law and the additive form of the central limit theorem is shown in the proof of Theorem 2, Appendix I. Thus, Kapteyn's machine (see Figure 1) is based on the generating model:

$X_j - X_{j-1} = e_jX_{j-1}$ $(j = 1, \ldots, n)$, where e_j is specified by:

$P\{e_j = a\} = 1/2$ and $P\{e_j = -a\} = 1/2$, for all j, and a is a positive constant.

The machine consists of nine rows of wedges encased in a wood and glass frame 104 cm high. The width of the wedges are proportional to the distance of the vertex of the wedge from the left-hand side of the frame. i.e., if X_{j-1} is the distance of a vertex from the left-hand side of the frame, the width of the wedge is $2aX_{j-1}$.

Sand is poured into a funnel directly above the center wedge in the top row. When arriving at the point X_{j-1}, the sand is divided into two equal parts, displaced either to $X_{j-1}(1 + a)$ or $X_{j-1}(1 - a)$. The sand arrives in the receptacles at the bottom of the machine, forming a skewed histogram.

M. J. Van Uven joined J. C. Kapteyn in 1916 to further develop the distribution in which estimation using quantiles was added [11]. Shortly after, the distribution received great criticism from K. Pearson who based his objections on general mistrust of the technique of transformations. Interest in the distribution died down until about 1930 when papers published by Clark [4], Hemmingsen [8], and Bliss [2] indicated that log normal distributions were effective in normalizing distributions in biological studies. With the invention of high-speed computing, sophisticated methods of analysis were developed in order to create tables of characteristics of the log normal distribution. The United States Environmental Protection Agency has currently developed two packages, SCOUT and GEO-EAS [16], which include the log transformation of data. In this thesis, the SCOUT package is used to perform the Kolmogorov-Smirnov test for normality, and the GEO-EAS package to compute sample statistics. The characteristics of the two-parameter log normal distribution are defined below.

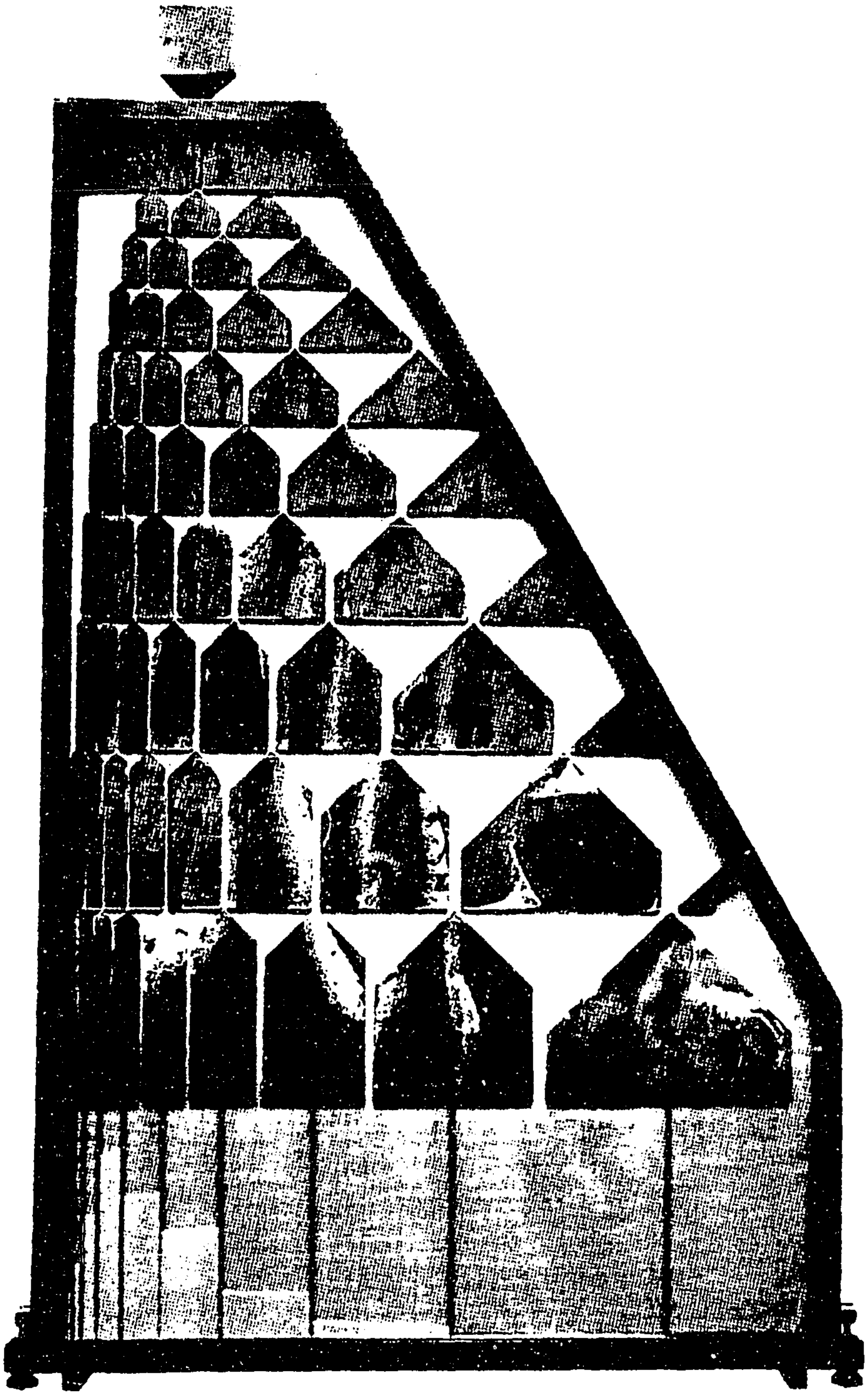

Figure 1. Kapteyn's Analogue Machine for Generating a Skew Frequency

DEFINITION

Consider a positive variate Y ($0 < y < \infty$) such that X = ln Y is normally distributed with mean μ and variance σ^2. Then we say that Y is log normally distributed with parameters μ,σ, and denote it by $Y \sim \Lambda(y \mid \mu,\sigma^2)$ and correspondingly, $X \sim N(x \mid \mu,\sigma^2)$. It is important to note that the distribution of X is completely specified by the two parameters μ,σ. However, Y cannot assume zero values since the transformation X = ln Y is not defined for Y = 0. Figure 2 gives a comparison of the frequency curves of the $N(x \mid \mu=0,\sigma^2=0.5)$ and $\Lambda(y \mid \mu=0,\sigma^2= 0.5)$, showing the positions of the mean, median and mode for the $\Lambda(y \mid \mu=0,\sigma^2= 0.5)$ distribution. Since X and Y are connected by the relationship X = ln Y, the distribution functions are related. Thus, from the properties of the moment generating function of the normal distribution, we can derive the following formulas (see Appendix I) [1]:

Mean of Y = E(Y) = $e^{\mu+.5\sigma^2}$

Median of Y = e^{μ}

Mode of Y = $e^{\mu-\sigma^2}$

Variance of Y = VAR(Y) = $e^{2\mu+\sigma^2}(e^{\sigma^2}-1) = (e^{\mu+.5\sigma^2})^2\eta^2$, where $\eta^2 = (e^{\sigma^2}-1)$

Skewness of Y = $\eta^3 + 3\eta$

Coefficient of Kurtosis of Y = $\eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2$

It is clear from the above formulas that the distribution is positively skewed and that the greater the value of σ^2, the greater the skewness. Also, the distribution has positive kurtosis which increases as σ^2 increases. Figure 3 shows the frequency curves for $\Lambda(x \mid \mu = 0,\sigma^2 = 0.5)$, $\Lambda(x \mid \mu = 0,\sigma^2 = 0.1)$, and $\Lambda(x \mid \mu=0,\sigma^2 = 2)$ from which the flexibility of the distribution may be obtained. An additional remark is that the two-parameter log normal distribution has several properties which are immediate consequences of those for the normal distribution. In particular, it is important to note that $E[\ln(Y)] \neq \ln(E[Y])$.

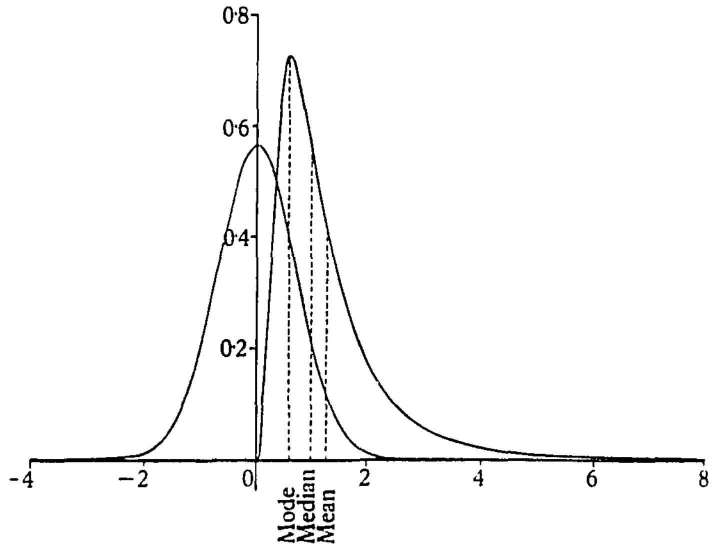

Figure 2. Frequency Curves for $N(x \mid \mu = 0, \sigma^2 = 0.5)$ and $\Lambda(x \mid \mu = 0, \sigma^2 = 0.5)$

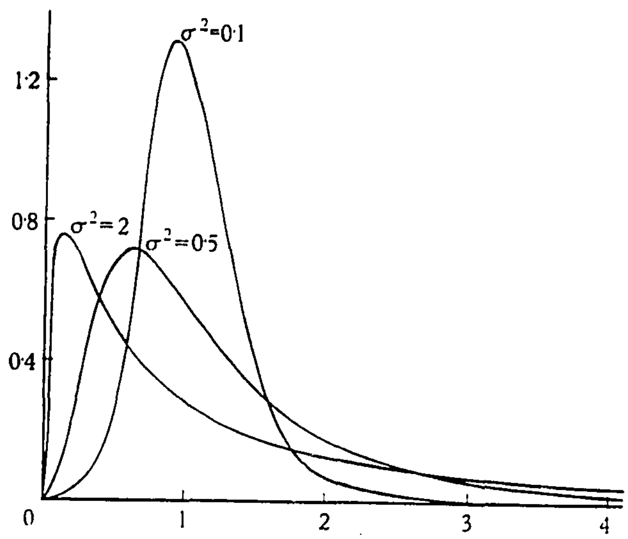

Figure 3. Frequency Curves $\Lambda(x \mid \mu = 0, \sigma^2 = 0.5)$, $\Lambda(x \mid \mu = 0, \sigma^2 = 0.1)$, and $\Lambda(x \mid \mu=0, \sigma^2 = 2)$

MOTIVATION FOR USE

The log normal distribution can be adequately described in natural occurrences of observed distributions in several fields of study, such as Economics, Biology, and Small-particle Statistics. Documented examples of application of log normal theory in these fields follow, with emphasis on Small-particle Statistics.

In the field of Economics, distributions of personal income have attracted the greatest attention. The choice of a particular form of the distribution is governed by the statistical description of the model and the criterion specified. Champernowne [3] developed a model which depended on the subdivision of income into discrete ranges. Contrarily, Lorenz [12] developed a model based on the concept of the concentration of incomes. The evidence studied by the authors suggested that the distribution of income is in fact log normal. Moreover, the more homogeneous the group of income recipients is, the more likely the distribution is log normal.

In the field of Biology, Cramer [5] discusses the growth of an organism subject to a number of small independent impulses acting in an ordered sequence. The law of proportionate effect applies if the influence of each impulse is proportionate to the momentary size of the organism. Thus, the final size of the organism will tend to be log normally distributed (as proved prior by Kapteyn).

The log normal distribution is well-established in Small-particle statistics. Many contamination data sets are highly skewed with as much as hundred-fold increase in size from the smallest to the largest. In addition, researchers are often interested in related particle measurements such as diameters, volumes and weights. Extensive research by Matheron [13] showed that the geochemical process of solution and concentration tends to produce log normal distributions for grades in mining applications.

The following two examples demonstrate that contaminant concentration data is typically highly skewed.

EXAMPLE 1: Samples of PCB collected from Superfund Site 1

Statistical analysis was performed on the data samples collected from Site 1 using The Environmental Protection Agency's GEO-EAS package. As we can see from Figure 4(a), the coefficient of skewness is 1.730 which implies that the data is highly skewed. However, by using the transformation X = ln Y, and performing the analysis on the transformed data, we can see from Figure 4(b) the coefficient of skewness is now -0.522 which means the data is just slightly skewed.

EXAMPLE 2: Samples of Chrysene collected from Superfund Site 2

Using the same statistical package (GEO-EAS), Figure 5(a) analysis shows the coefficient of skewness for the original data is 1.848 which implies the data is highly skewed. However, once the data is transformed, we can see in Figure 5(b) that the data is just slightly skewed since the coefficient of skewness is -0.328.

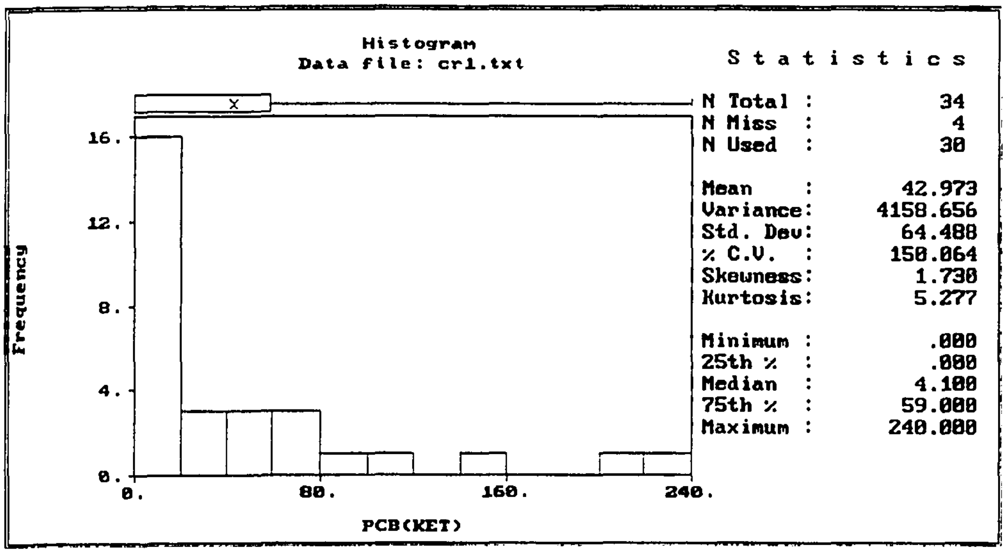

Figure 4(a). Example 1: Histogram of Superfund Site 1 (raw data)

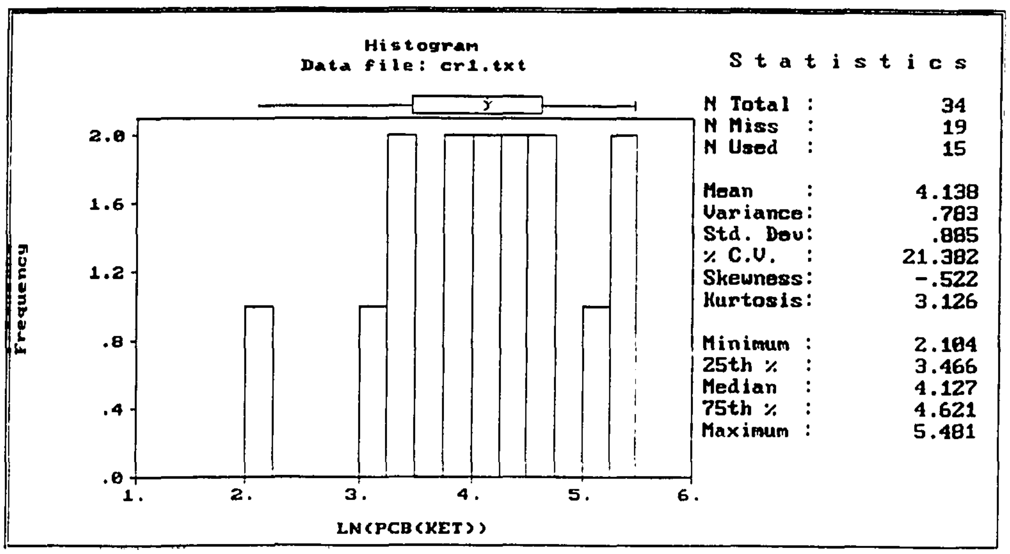

Figure 4(b). Example 1: Histogram of Superfund Site 1 (ln of raw data)

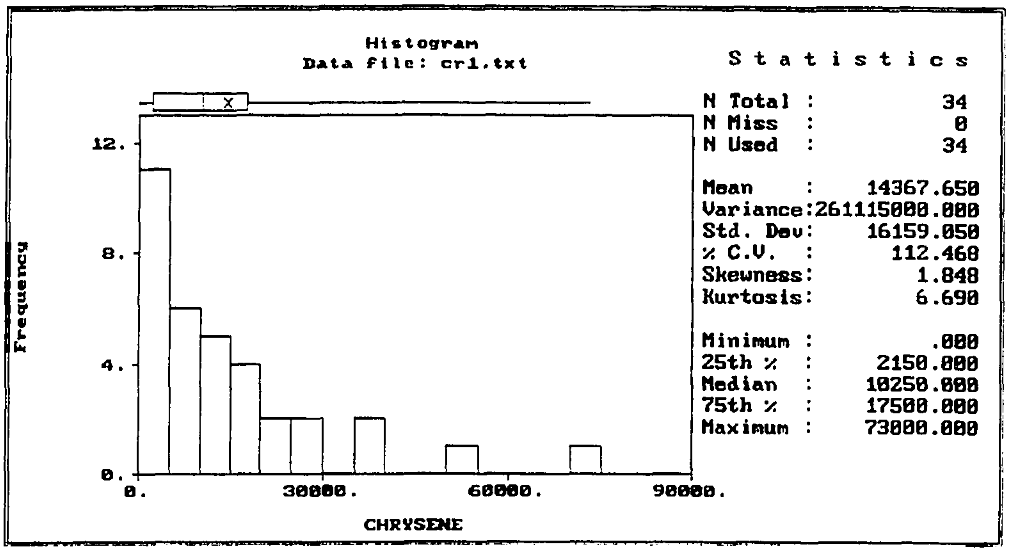

Figure 5(a). Example 2: Histogram of Superfund Site 2 (raw data)

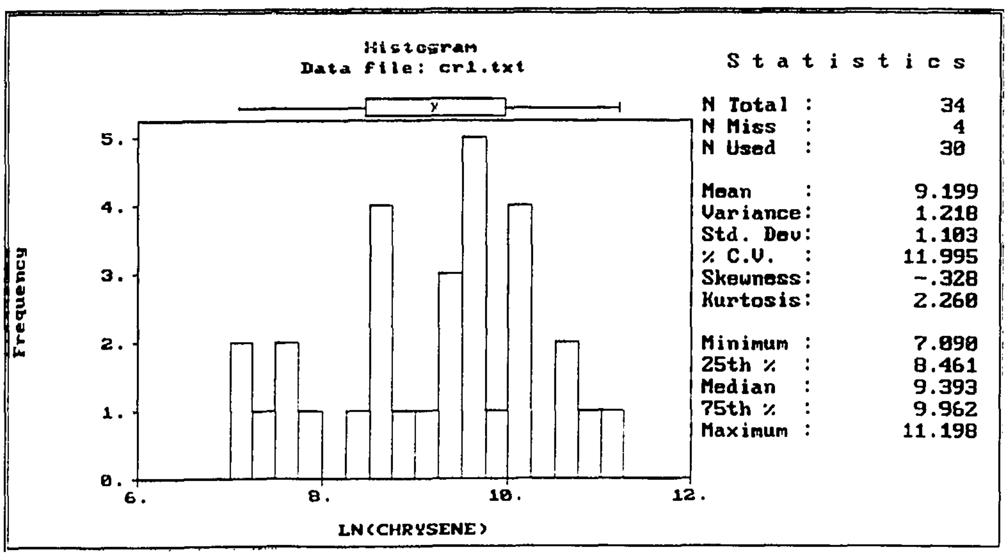

Figure 5(b). Example 2: Histogram of Superfund Site 2 (ln of raw data)

CHAPTER 2

ESTIMATION OF THE MEAN OF A LOG NORMAL DISTRIBUTION

METHOD 1: POINT ESTIMATION

Generally, a random variable X has a probability density function of known form which depends on an unknown parameter θ, $\theta \in \Omega$. Therefore, we have a family of distributions for each value of θ, $\theta \in \Omega$. We denote this family as $\mathcal{F} = \{f(x,\theta) ; \theta \in \Omega\}$. Our goal is to select one member of this family as being the probability density function of X. In other words, we want a point estimate of θ. This estimate of θ is denoted by $\hat{\theta}$. Following are some of the desirable properties which a good estimator must possess.

i) the estimator should be unbiased, i.e., $E(\theta) = \theta$.

ii) the variance of the estimator should be minimized.

Suppose we are given $y_1, y_2, \ldots, y_n$ independent random samples from $\Lambda(y \mid \mu,\sigma^2)$. The sample mean, $\bar{y} = 1/n\sum_{i=1}^{n} y_i$ can be used as an estimate of μ. However, it is known to be an inefficient estimator of μ since it is usually affected by a few large values.

For data that is highly skewed, the arithmetic mean of assays tends to overestimate the mean of the distribution due to the presence of erratic high values. Sichel [15] theoretically derived a better way to estimate the mean by the following relationship: $\mu = e^{\alpha+.5\beta^2}$, where α is the mean of the logarithms and β is their standard deviation. However, this relationship is between parameter values and not their estimates. Sichel developed the Sichel "t-estimator" to overcome this estimation problem. It is defined as follows:

$$t = e^{Y}\psi(s^2/2, n), \text{ where } \psi(u,n) = 1 + \frac{n-1}{n}u + \frac{(n-1)^3}{n^2(n+1)2.1}u^2 + \ldots \text{ and}$$

s^2= MLE of σ^2.

Tables of $\psi(\beta^2, n)$ (Table 1, Appendix II) are given in Sichel [15]. Sichel's estimator is similar in form to the minimum variance unbaised estimator derived by Finney [6].

To demonstrate the use of Sichel's Table, suppose we have a sample of size 10 randomly drawn from a log normal distribution. The logarithmic mean, $\alpha = 5.298$ and the variance, $\beta^2 = 0.8$. Then the mean of $\overline{X}$ is given as:

$$\overline{X} = e^{\alpha}\psi(\beta^2, n), where \beta^2 = 0.8, n = 10$$

$= e^{5.298}(1.472)$ (from Table 1, Appendix II)

$= 294.3$

METHOD 2: CONFIDENCE INTERVAL ESTIMATION

The confidence interval utilizes the information in the sample to arrive at two numbers that are intended to enclose the parameter, θ, of interest. Ideally, we would like the interval to have two properties: (i) the interval should contain the target parameter, θ; and (ii) the interval should be relatively narrow. The probability that a confidence interval will enclose θ is called the confidence coefficient, denoted by $1 - \alpha$. This confidence coefficient gives the fraction of the time, in repeated sampling, that the interval constructed will contain the target parameter θ.

Suppose that θ_L and θ_U denote the lower and upper confidence limits, respectively, for a parameter θ. If $P(\theta_L < \theta < \theta_U) = 1 - \alpha$, then the probability, $(1 - \alpha)$, is the confidence coefficient. The random interval, (θ_L, θ_U), is called a two-sided confidence interval.

For example, let $\alpha = .05$. If we performed repeated sampling, say 100 times, then 95% of the

time, our confidence interval would contain our target parameter θ. In other words, in the long run, about 95 out of the 100 confidence intervals constructed would contain θ.

NORMAL THEORY

Since we rarely know the form of the population frequency distribution, we make the assumption that the samples have been randomly selected from a normal population, where μ and σ^2 are unknown. Also, we know that $T = \frac{\overline{Y} - \mu}{s / \sqrt{n}}$ has a t-distribution with (n-1) degrees of freedom.

Then we can form a confidence interval for μ. The resulting confidence interval is: $\overline{Y} \pm t_{.5\alpha} \frac{s}{\sqrt{n}}$.

SICHEL'S METHOD

Sichel calculated multiplying factors to compute a central 90% confidence interval for the mean of the log normal distribution. Table 2 and 3 in Appendix II is the lower and upper 5% limits of error of the t-estimator, respectively.

As an illustration of the use of the tables to compute a 90% confidence interval, suppose we have a sample of size 10 randomly drawn from a log normal distribution. The logarithmic mean, $\alpha = 5.298$ and the variance, $\hat{\beta}^2 = 0.8$. Then the 90% confidence interval for the mean is given as:

$$C_L = e^{\mu} t_L(\hat{\beta}^2, n), where \hat{\beta}^2 = 0.8, n = 10$$

$= e^{5.298}(0.93)$ (from Table 2, Appendix II)

$= 186.0$

$$C_U = e^{\mu} t_U(\hat{\beta}^2, n), where \hat{\beta}^2 = 0.8, n = 10$$

$= e^{5.298}(3.55)$ (from Table 3, Appendix II)

$= 709.8$

CHAPTER 3

ONE PROBLEM WITH THE USE OF LOG NORMAL DISTRIBUTIONS

The United States Environmental Protection Agency (USEPA) Guidance document [17] provides detailed information on statistical methods applicable to the problem of deciding, on the basis of a random sample collected from the site, whether the site meets the cleanup standards or not. In particular, this guidance document provides the following formula for determining the number of samples required to obtain a specified confidence level:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(C_s - \mu_1)^2} \sigma^2, \text{ where } C_s = \text{cleanup standard for the site}$$

σ^2 = variance (estimated)

μ_1 = mean under the alternative hypothesis ($< C_s$)

$z_{1-\alpha}$ = upper $100(1-\alpha)\%$ point of standard normal distribution

$z_{1-\beta}$ = upper $100(1-\beta)\%$ point of standard normal distribution

An evaluation of an EPA Superfund site was requested to determine the number of boring locations necessary to characterize chemical concentrations of sediments in the study area. The above-referenced USEPA evaluation method utilizing the expected variability based on historic data was used.

Preliminary evaluation of the results from the analysis of the USACE samples indicated that two distinct 2,3,7,8-TCDD concentration distributions were present. Thus, the site was divided into two areas: A and B, respectively. The following historical data was used in the computations:

AREA A (ppt)	49	180	290	220	230	110	57	120	94	120	210
AREA B (ppt)	760	480	1500	68	380	6300	20	820	960		

The evaluation of the data was performed by first computing summary statistics and secondly, evaluating the underlying distribution of the data. The summary statistics are as follows:

	$\bar{y}$	s	MIN	MAX
AREA A	153	78	49	290
AREA B	1734	2242	20	6300

The Kolmogorov-Smirnov test was used for Area A (see Figure 6) and did not reject normality for Distribution A. Thus, Area A's underlying distribution is indicated to be normal. On the other hand, the Kolmogorov-Smirnov test rejected normality for Distribution B (see Figure 7(a)). However, by converting the samples of Area B to log-scale, the Kolmogorov-Smirnov test did not reject normality for the log-scale data (see Figure 7(b)). Thus, the underlying distribution for Area B is indicated to be log normal.

Using the above-mentioned formula for the number of boring locations needed and given the following criteria:

AREA A: $\alpha = 0.05$ $\beta = 0.20$ $C_s - \mu_1 = 80$ ppt

AREA B: $\alpha = 0.05$ $\beta = 0.20$ $C_s - \mu_1 = 100$ ppt

the contractor suggested that a total of eight samples per mile would be adequate to characterize 2378-TCDD sediment concentrations for Area A. Additionally, seventeen samples per mile would be adequate for Area B. However, this estimation appeared to be incorrect because the formula for the number of samples depends on the standard deviation, which is considerably higher for Area B. Consequently, the number of samples for Area B was calculated using the standard deviation of the real scale values of the samples and it was determined that 3054 samples would be adequate.

This huge discrepancy in the two sample sizes forced us to look at the usage of the sample-size formula on log-transformed data. In this example, the contractor replaced the error limit of 100 ppt by ln(100). The problem with this approach is that the difference of ln(100) on a log-scale does not

translate to a difference of 100 in the means of the concentrations on the real scale. This is shown clearly by the following two graphs. In Figure 8, the X-axis represents the difference of mean of log-transformed variables ($\mu_1 - \mu_0$) when $\mu_0 = 1$. The Y-axis represents the difference in means of the original variables. If we look at a change of log(100) from μ_0 in the X-direction, this corresponds to a change of 100 in the Y-direction. However, in Figure 9, a change of log(100) in the X-direction corresponds to a change of nearly 8000 in the Y-direction.

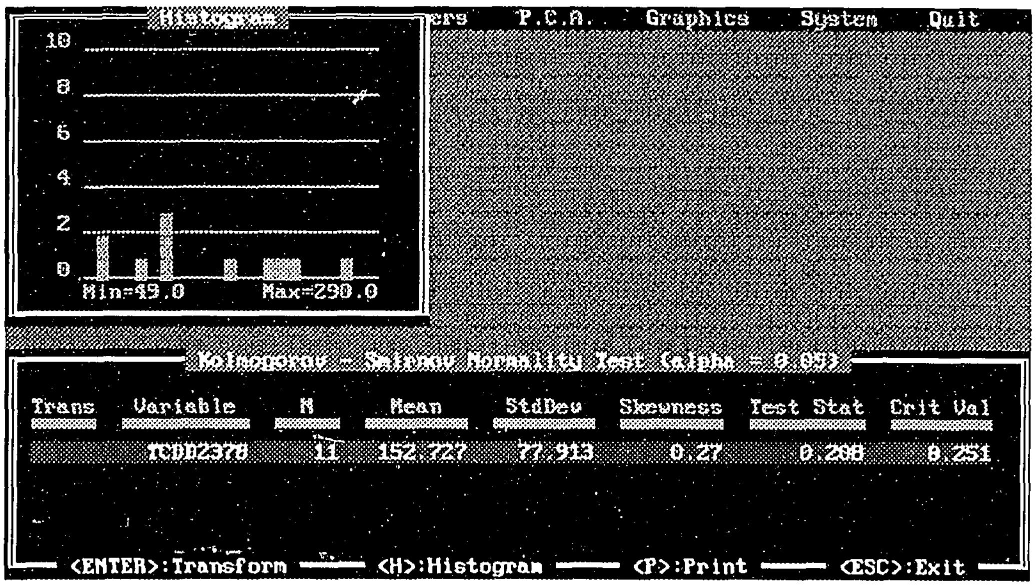

Figure 6. Kolmogorov-Smirnov Test for Normality for Area A (raw data)

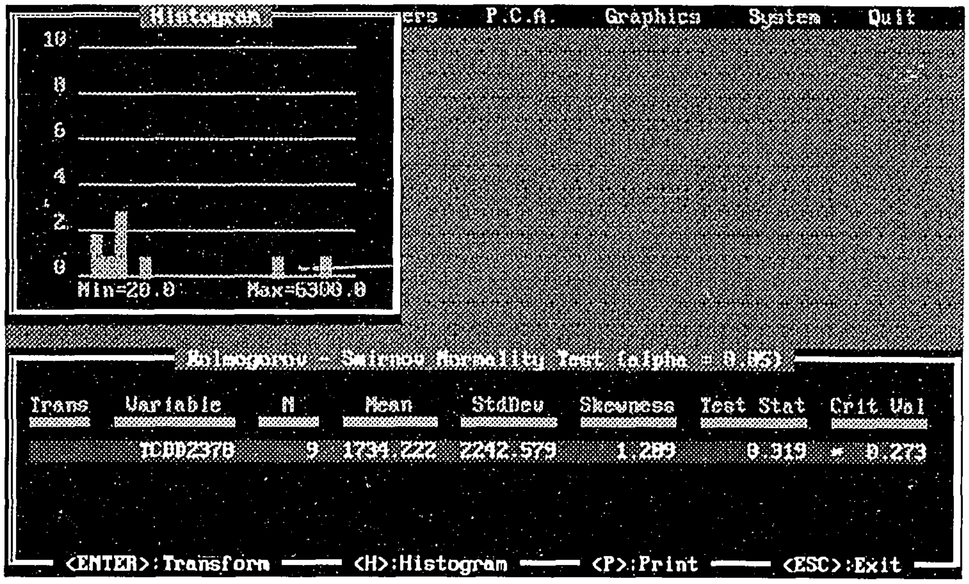

Figure 7(a). Kolmogorov-Smirnov test for normality for Area B (raw data)

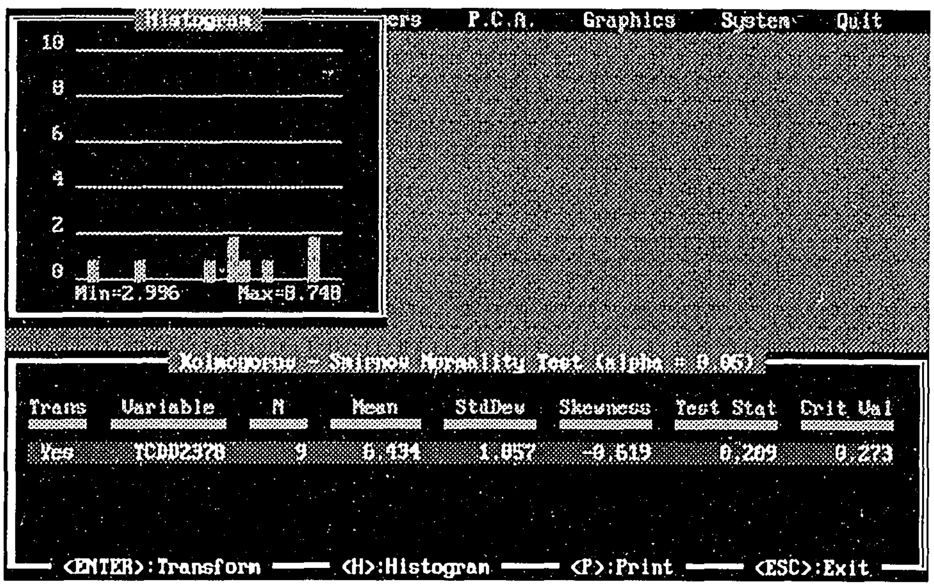

Figure 7(b). Kolmogorov-Smirnov Test for Normality for Area B (ln of raw data)

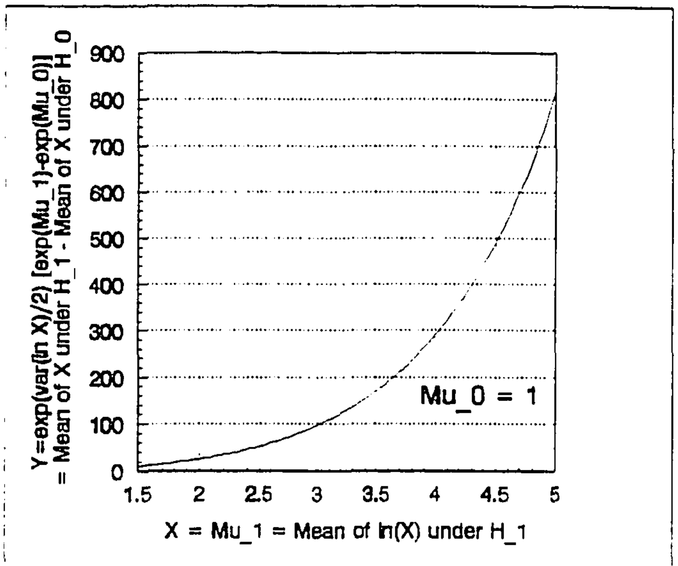

Figure 8. Graph of Mean Differences in Original Variables vs. Transformed Variables ($\mu_0 = 1$)

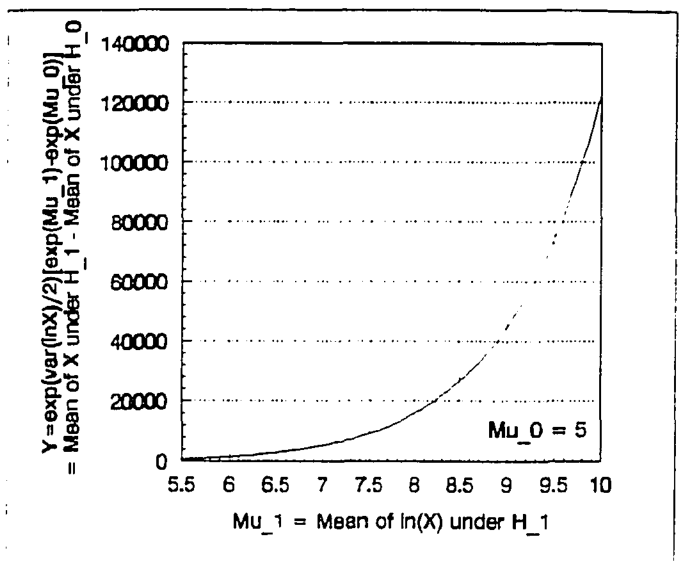

Figure 9. Graph of Mean Differences in Original Variables vs. Transformed Variables ($\mu_0 = 5$)

CHAPTER 4

SIMULATION EXPERIMENT

As we have shown, contaminant concentration data is typically highly skewed. The statistician normally will compute the Kolmogorov-Smirnov test for normality to determine if the sample data fit a normal distribution. If not, the sample data will be log transformed. The Kolmogorov-Smirnov test will then be computed on the transformed data to determine if it fits a normal distribution. If so, "normal" theory-based formulas will be applied for the analysis of the sample data.

We know that for a small number of sample data, it is very difficult to determine what type of distribution the data may have. On the other hand, for a large number of samples, we can apply the Central Limit Theorem. The question that arises, then, is: "For a small number of sample data, is it necessary to log transform the sample data and apply Sichel's theory to determine an estimate for the mean, or rather, use Normal theory estimates on the real-scale sample points?" As discussed prior, there is misinterpretation of the use of this log transformed sample data. The goal of the simulation experiment was to compare these two methods of confidence interval estimation for the mean in hopes to answer the question proposed.

A FORTRAN program was written and included in Appendix III for the simulation experiment in which Monte Carlo simulation was used to generate sample data. The simulation experiment is fully described below.

DESCRIPTION - PART 1

The first step in the simulation experiment was to generate a random sample of size n from a log normal distribution with parameters μ, σ [denoted $\Lambda(y \mid \mu,\sigma^2)$]. This was accomplished using IMSL STAT/LIBRARY [9] subroutine RNLNL. Once our samples were generated, a 90% confidence

interval was computed using "normal" theory. Then each sample was log transformed. A 90% confidence interval was then computed on the transformed samples using "Sichel's" theory. Since Sichel's t-tables were limited, we determined whether or not the confidence interval could be computed due to missing table values. This is indicated on the output as "Number of Misses on Table". Once the confidence interval for each method was computed, a determination of whether the true mean was contained in each interval was noted. This is indicated as "Number of Hits for Normal Theory" and "Number of Hits for Log Normal Theory", for each method respectively. Finally, the average interval length and its standard deviation was computed.

In Examples 1, 2 and 3 that follow, Step 1 of the simulation experiment in demonstrated. Note that the average length of the intervals for each theory is the actual interval length which has a standard deviation value as zero since we have only computed one interval for each method.

In Example 1, we have requested 10 samples from a $\Lambda(y \mid \mu = 1.1, \sigma^2 = .16)$. First, the output shows what we have requested and lists the generated data points. The 90% confidence interval is computed. The data points are then log transformed and listed. The 90% confidence interval under Sichel's theory is computed. The results of Example 1 show that the true mean is contained in each interval.

Similarly, in Example 2, we have requested 10 samples from a $\Lambda(y \mid \mu = 1.1, \sigma^2 = .25)$. The results are the same as in Example 1.

Example 3 was generated from a $\Lambda(y \mid \mu = 1.1, \sigma^2 = .04)$. The results for this example, show that an interval under Sichel's theory was not computed due to missing table value. Therefore, the interval defaults to [0,0] and the indicator for number of misses on table is now 1. The true mean is contained in the interval computed under normal theory.

EXAMPLE 1

REQUIRED INPUT: Number of samples, Number of simulation runs, Mean, Standard deviation, T-value, T-column

PROGRAM OUTPUT: The program returns the following output.

```
YOU HAVE REQUESTED 10 DATA POINTS
GENERATED FROM A LN SAMPLE WITH MEAN:          1.10000
                  AND A STANDARD DEVIATION:    0.40000

THE EXACT MEAN BASED ON LOG NORMAL THEORY FOR THIS DATA IS:  3.254375

THE GENERATED LOG NORMAL DATA POINTS ARE:

   3.157331    2.747720    5.457298    5.348978    1.618213
   3.098199    3.111766    1.753176    5.046471    4.516624

THE 90% CONFIDENCE INTERVAL BASED ON NORMAL THEORY IS:
[2.803816, 4.367339]

THE LN OF THE GENERATED DATA POINTS ARE:

   1.149727    1.010772    1.696954    1.676905    0.4813222
   1.130821    1.135190    0.5614293   1.618689    1.507765

THE 90% CONFIDENCE INTERVAL BASED ON SICHEL'S THEORY IS:
[2.945927, 5.329149]

THE NUMBER OF HITS FOR NORMAL THEORY IS:          1
AVERAGE LENGTH:         1.5635        S.D:    0.0000

THE NUMBER OF HITS FOR LN THEORY IS:              1
AVERAGE LENGTH:         2.3832        S.D:    0.0000
   NUMBER OF MISSES ON TABLE IS:    0

THIS IS THE END OF THE RUN
```

EXAMPLE 2

REQUIRED INPUT: Number of samples, Number of simulation runs, Mean, Standard deviation, T-value, T-column

PROGRAM OUTPUT: The program returns the following output.

```
YOU HAVE REQUESTED 10 DATA POINTS
GENERATED FROM A LN SAMPLE WITH MEAN:          1.10000
                AND A STANDARD DEVIATION:      0.50000

THE EXACT MEAN BASED ON LOG NORMAL THEORY FOR THIS DATA IS:  3.404166

THE GENERATED LOG NORMAL DATA POINTS ARE:

   4.882174    2.535231    1.617043    2.480418    1.570380
   11.17985    12.99840    6.926432    1.563053    3.284885

THE 90% CONFIDENCE INTERVAL BASED ON NORMAL THEORY IS:
[ 2.613038, 7.194535]

THE LN OF THE GENERATED DATA POINTS ARE:

   1.585591    0.9302849   0.4805991   0.9084271   0.4513175
   2.414113    2.564826    1.935345    0.4466408   1.189332

THE 90% CONFIDENCE INTERVAL BASED ON SICHEL'S THEORY IS:
[ 3.271626, 10.06934]

THE NUMBER OF HITS FOR NORMAL THEORY IS:          1
AVERAGE LENGTH:        4.5815        S.D:    0.0000

THE NUMBER OF HITS FOR LN THEORY IS:              1
AVERAGE LENGTH:        6.7977        S.D:    0.0000
   NUMBER OF MISSES ON TABLE IS:   0

THIS IS THE END OF THE RUN
```

EXAMPLE 3

REQUIRED INPUT: Number of samples, Number of simulation runs, Mean, Standard deviation, T-value, T-column

PROGRAM OUTPUT: The program returns the following output.

```
YOU HAVE REQUESTED 10 DATA POINTS
GENERATED FROM A LN SAMPLE WITH MEAN:          1.10000
                  AND STANDARD DEVIATION:      0.20000

THE EXACT MEAN BASED ON LOG NORMAL THEORY FOR THIS DATA IS:  3.064854

THE GENERATED LOG NORMAL DATA POINTS ARE:

   3.526819     4.589726     2.952541     2.681173     2.466662
   2.262289     3.188480     3.437556     3.396833     3.197326

THE 90% CONFIDENCE INTERVAL BASED ON NORMAL THEORY  IS:
[ 2,798257, 3.530624]

THE LN OF THE GENERATED DATA POINTS ARE:

   1.260396     1.523820     1.082666     0.9862543    0.9028659
   0.8163772    1.159544     1.234761     1.222844     1.162315

THE 90% CONFIDENCE INTERVAL BASED ON SICHEL'S THEORY IS:
[ 0.0000, 0.0000]

THE NUMBER OF HITS FOR NORMAL THEORY IS:          1
AVERAGE LENGTH:       0.7214         S.D:    0.0000

THE NUMBER OF HITS FOR LN THEORY IS:              0
AVERAGE LENGTH:       0.0000         S.D:    0.0000
   NUMBER OF MISSES ON TABLE IS:     1

THIS IS THE END OF THE RUN
```

DESCRIPTION - PART 2

Recall that a specified confidence of 90% means that for repeated samplings, about 90% of the estimated intervals should contain the true mean. Consequently, for the second part of the simulation experiment, Step 1 was repeatedly run the desired number of times. The output has been modified to reflect only the computed confidence intervals for each simulation run without listing the actual simulated data points for each run. Therefore, the indicators for the "Number of Hits for Normal Theory" and "Number of Hits for Log Normal Theory", for each method respectively, reflect the total for all the simulation runs along with the average interval length and its standard deviation.

Examples 4, 5, and 6, correspond to Examples 1, 2 and 3, respectively, with the number of simulations increased to ten. Note that the first interval in Examples 4, 5 and 6, are exactly the same as in their corresponding example.

The results of Example 4 indicates that the true mean was contained in all 10 of the intervals computed for both methods.

In Example 5, we see that 9 out of 10 of the intervals contained the true mean for Sichel's theory, whereas all 10 did for normal theory.

Example 6 shows that under Sichel's theory, 9 intervals were not computed due to missing table values, but the remaining one did contain the true mean. Under normal theory, all 10 intervals did contain the true mean.

EXAMPLE 4

REQUIRED INPUT: Number of samples, Number of simulation runs, Mean, Standard deviation, T-value, T-column

PROGRAM OUTPUT: The program returns the following output.

```
YOU HAVE REQUESTED 10 DATA POINTS
GENERATED FROM A LN SAMPLE WITH MEAN:          1.10000
                  AND STANDARD DEVIATION:      0.40000

THE EXACT MEAN BASED ON LOG NORMAL THEORY FOR THIS DATA IS:  3.254375

                  THE 90% CONFIDENCE INTERVAL
                              FOR
   NORMAL THEORY                          SICHEL'S THEORY

   [ 2.8038, 4.3673]                      [ 2.9459, 5.3291]
   [ 2.3935, 3.6491]                      [ 2.5408, 3.8395]
   [ 2.1225, 3.5190]                      [ 2.3449, 3.5434]
   [ 2.5602, 3.6278]                      [ 2.6635, 4.0249]
   [ 2.7610, 4.7020]                      [ 3.0642, 5.5432]
   [ 2.2100, 3.2571]                      [ 2.3485, 3.5488]
   [ 2.5450, 3.6758]                      [ 2.5860, 4.6780]
   [ 3.0108, 3.9093]                      [ 3.0353, 4.5866]
   [ 2.6643, 4.1545]                      [ 2.8094, 5.0822]
   [ 2.7008, 4.6110]                      [ 2.9391, 5.3168]

THE NUMBER OF HITS FOR NORMAL THEORY IS:       10
AVERAGE LENGTH:       1.3701        S.D:    0.3404

THE NUMBER OF HITS FOR LN THEORY IS:           10
AVERAGE LENGTH:       1.8215        S.D:    0.5163
   NUMBER OF MISSES ON TABLE IS:   0
```

THIS IS THE END OF THE RUN

EXAMPLE 5

REQUIRED INPUT: Number of samples, Number of simulation runs, Mean, Standard deviation, T-value, T-column

PROGRAM OUTPUT: The program returns the following output.

```
YOU HAVE REQUESTED 10 DATA POINTS
GENERATED FROM A LN SAMPLE WITH MEAN:          1.10000
                 AND STANDARD DEVIATION:        0.50000

THE EXACT MEAN BASED ON LOG NORMAL THEORY FOR THIS DATA IS:  3.404166

                THE 90% CONFIDENCE INTERVAL
                            FOR
   NORMAL THEORY                          SICHEL'S THEORY

   [ 2.6130, 7.1945]                      [ 3.2716, 10.0693]
   [ 2.5770, 3.9915]                      [ 2.7570,  4.1661]
   [ 2.7860, 4.7012]                      [ 3.0390,  5.4975]
   [ 2.1868, 4.6069]                      [ 2.5350,  5.4975]
   [ 2.2167, 3.4240]                      [ 2.3241,  4.2042]
   [ 2.8889, 4.2500]                      [ 2.9700,  5.3728]
   [ 2.9972, 5.0091]                      [ 3.2026,  5.7935]
   [ 2.6777, 4.5829]                      [ 2.9260,  5.2931]
   [ 2.8945, 5.2829]                      [ 3.9403,  9.4302]
   [ 2.8945, 5.2829]                      [ 3.2350,  5.8520]

THE NUMBER OF HITS FOR NORMAL THEORY IS:          10
AVERAGE LENGTH:        2.3715        S.D:    1.1535

THE NUMBER OF HITS FOR LN THEORY IS:              9
AVERAGE LENGTH:        3.0836        S.D:    1.6041
   NUMBER OF MISSES ON TABLE IS:   0
```

THIS IS THE END OF THE RUN

EXAMPLE 6

REQUIRED INPUT: Number of samples, Number of simulation runs, Mean, Standard deviation, T-value, T-column

PROGRAM OUTPUT: The program returns the following output.

```
YOU HAVE REQUESTED 10 DATA POINTS
GENERATED FROM A LN SAMPLE WITH MEAN:          1.10000
                AND STANDARD DEVIATION:        0.20000

THE EXACT MEAN BASED ON LOG NORMAL THEORY FOR THIS DATA IS:  3.064854

                 THE 90% CONFIDENCE INTERVAL
                            FOR
     NORMAL THEORY                        SICHEL'S THEORY

     [ 2.8093, 3.5306]                    [ 0.0000, 0.0000]
     [ 2.8119, 3.3804]                    [ 0.0000, 0.0000]
     [ 2.6725, 3.2991]                    [ 0.0000, 0.0000]
     [ 2.7759, 3.3546]                    [ 0.0000, 0.0000]
     [ 2.8161, 3.1976]                    [ 0.0000, 0.0000]
     [ 2.6470, 3.3807]                    [ 0.0000, 0.0000]
     [ 3.0041, 3.3246]                    [ 0.0000, 0.0000]
     [ 2.5416, 3.1179]                    [ 0.0000, 0.0000]
     [ 2.9843, 3.7493]                    [ 0.0000, 0.0000]
     [ 2.8375, 3.7257]                    [ 2.8767, 4.3470]

THE NUMBER OF HITS FOR NORMAL THEORY IS:        10
AVERAGE LENGTH:        0.6160        S.D:    0.1640

THE NUMBER OF HITS FOR LN THEORY IS:             1
AVERAGE LENGTH:        1.4703        S.D:    0.0000
   NUMBER OF MISSES ON TABLE IS:   9
```

THIS IS THE END OF THE RUN

RESULTS OF SIMULATION EXPERIMENT

The simulation experiment was repeated 100, 1000, and 10000 times for the requested number of data samples [5, 10, 15, 20, 25, 30, 50, 100]. Table 4 in Appendix II is a summary of the simulation experiment results. The columns are divided by the number of data samples requested and subdivided by the number of simulations runs. The rows are divided by standard deviation and subdivided by the two estimation methods. Also, the number of uncalculated estimated intervals for Sickle's theory was noted. Each cell entry represents either the number of times the estimated interval contained the true mean under the appropriate theory or the number of uncalculated intervals for Sichel's theory. For example, the results state for a standard deviation of .4 and 5 data samples requested, 86 out of 100 simulation runs contained the true mean for "normal" theory, whereas 71 out of 100 simulation runs contained the true mean with 24 intervals not computed for Sichel's theory.

The results of our extensive simulation experiment indicates that the approximate "normal" theory based confidence interval compares quite well in comparison to Sichel's theory based confidence interval for the mean of the log normal distribution. Overall, the average interval length for normal theory was smaller (see Table 5, Appendix II) with a higher confidence percentage (see Table 4, Appendix II).

For small n, it is not easy to test for normality or log-normality. By computing the Kolmogorov-Smirnov test for normality on the sample points in Example 1, the results indicate that the sample points fit a normal distribution (see Figure 10(a)). In addition, Figure 10(b) shows that the log transformation of the sample points also fit a normal distribution. Likewise, the same results follow with the sample data in Example 2 (see Figure 11(a),and (b)).

Since the approximate "normal" theory based confidence interval performs quite well even for small n, we might want to just use the approximate normal theory based methods.

For large n, the approximate "normal" theory based confidence interval is almost identical to Sichel's theory based confidence interval and hence, it does not pay to log-transform data to attain normality.

If log transformation must be used, then care should be taken in transforming formulas based on normal theory.

The simulation experiment was modified to generate data samples from a normal distribution with parameters μ,σ, for varying values of the parameters. In each case, the normal theory-based intervals were shorter in length and contained the true mean more times than that of Sichel's theory.

In conclusion, if the experimenter knows that the underlying distribution is log normal, and the parameters of interest are the natural log normal parameters (μ,σ^2), then log normal theory should be used. However, if the interest is in the mean of the distribution, then it is not necessary to log transform the sample data set when estimating the true mean of the sample distribution.

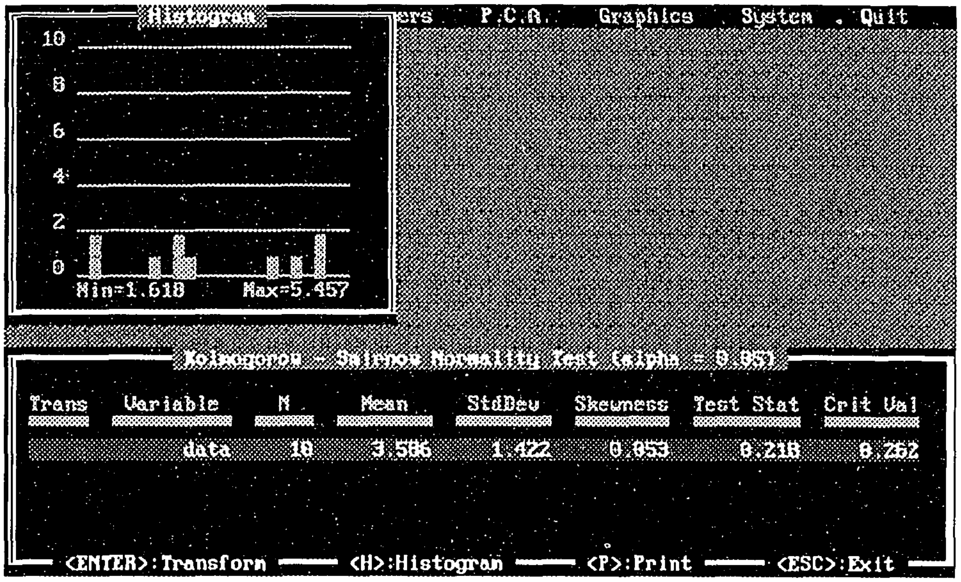

Figure 10(a). Kolmogorov-Smirnov Test for Normality for Example 1 (raw data)

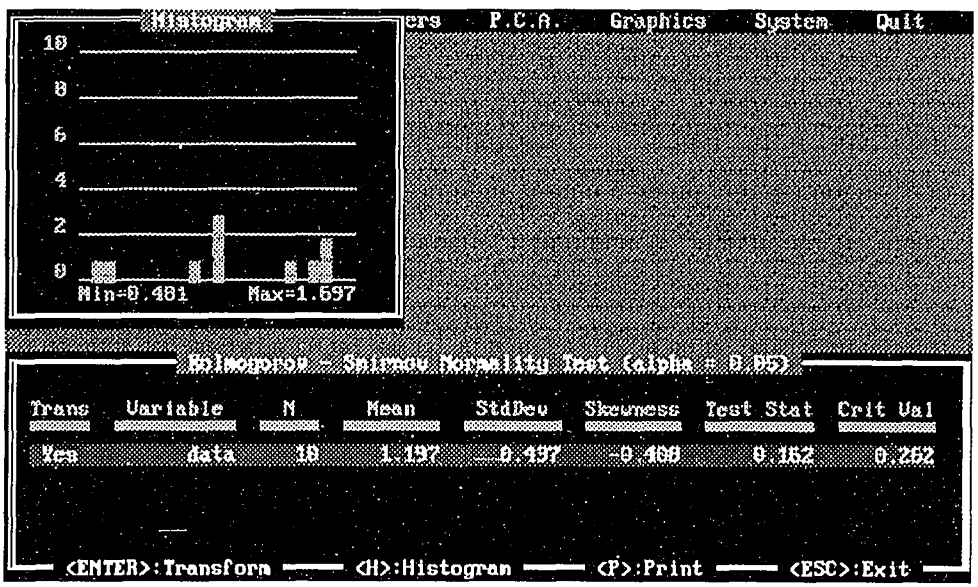

Figure 10(b). Kolmogorov-Smirnov Test for Normality for Example 1 (ln of raw data)

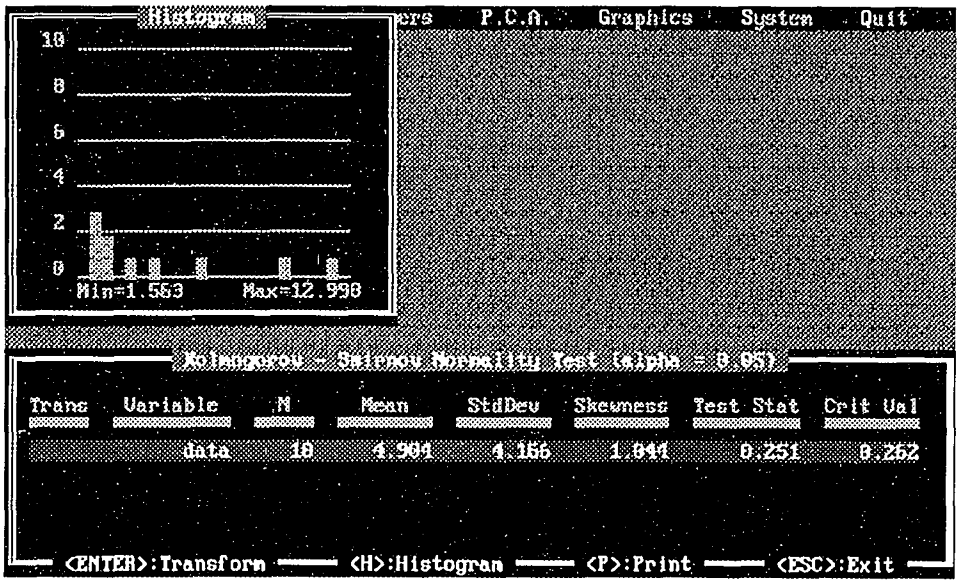

Figure 11(a). Kolmogorov-Smirnov Test for Normality for Example 2 (raw data)

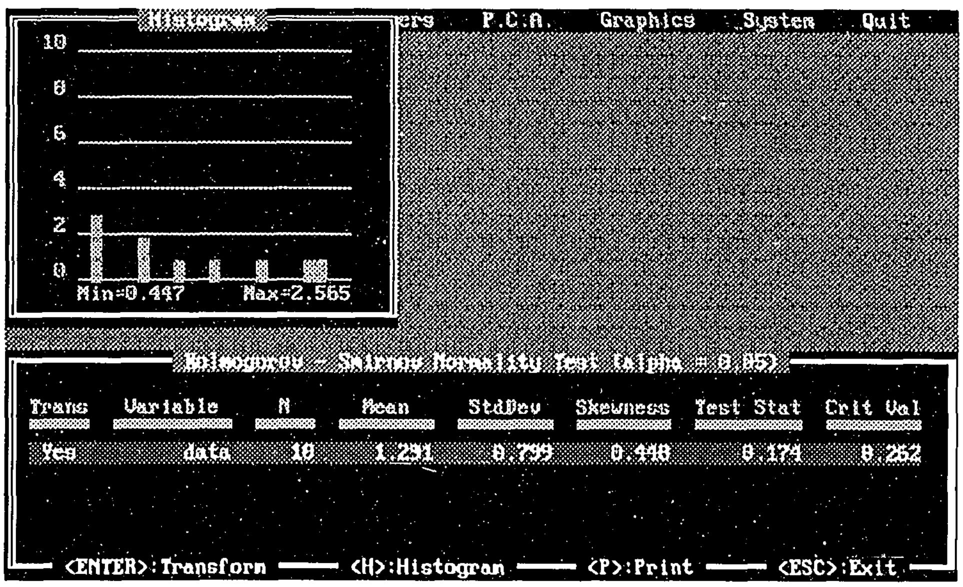

Figure 11(b). Kolmogorov-Smirnov Test for Normality for Example 2 (ln of raw data)

CHAPTER 5

CORRECT METHOD OF APPLYING SAMPLE SIZE FORMULA

From Chapter 3, the formula to determine the number of samples is:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \, \sigma^2}{(C_s - \mu_1)^2}, \text{ where } C_s = \text{cleanup standard for the site}$$

σ^2 = variance (estimated)

μ_1 = mean under the alternative hypothesis ($< C_s$)

$z_{1-\alpha}$ = upper 100(1-α)% point of standard normal distribution

$z_{1-\beta}$ = upper 100(1-β)% point of standard normal distribution

We determined that the error limit, ($C_s - \mu_1$), cannot be replaced by simply just taking the log transformation of the error limit. Recall that our test to determine whether a site is clean is given by:

H_0: $E(Y) = C_s$ vs H_a: $E(Y) = C_s - \Delta$

We wish to detect the difference Δ in the mean of the y-values. The difference is then:

$\Delta = C_s - E(Y \mid H_1) = e^{\mu_0 + .5\sigma^2} - e^{\mu_1 + .5\sigma^2}$, where σ^2 = sample variance of $X = \ln(Y)$.

Next, from the clean-up standard, $C_s = e^{\mu_0 + .5\sigma^2}$, we have:

$\mu_0 = \ln(C_s) - 0.5\sigma^2$; and from $E(Y \mid H_a) = M_1 = e^{\mu_1 + .5\sigma^2}$, we have:

$\mu_1 = \ln(M_1) - 0.5\sigma^2$

Thus, $\mu_0 - \mu_1 = \ln(C_s) - 0.5\sigma^2 - [\ln(M_1) - 0.5\sigma^2] = \ln\left(\frac{C_s}{M_1}\right) \neq \ln(\Delta)$

Therefore, the corrected formula for the number of samples for log transformed data is given by:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \, \sigma^2}{\ln\left(C_s/M_1\right)^2}, \text{ where all variables are defined as above.}$$

One point that need be clarified is that it is not enough to specify the Δ along, the clean-up standard must be specified. If we review the requirements for the sample design given in the problem of Chapter 3, the error limit for Area B was 100 ppt. That is $C_S - \mu_1 = 100$ ppt.

Case 1: Let $C_S = 200$, $M_1 = 100$

then the number of samples required is given by: $$n = \frac{(2.487)^2(3.447)}{\ln\left(200/100\right)^2} \approx 45$$

Case 2: Let $C_S = 5000$, $M_1 = 4900$

then the number of samples required is given by: $$n = \frac{(2.487)^2(3.447)}{\ln\left(5000/4900\right)^2} \approx 52{,}237$$

The error limit in each case is 100 ppt, but the number of samples required is dramatically different.

APPENDIX I

PROOFS OF LOG NORMAL FORMULAS

Theorem 1 [1]: *Let $\{X_j\}$ be a sequence of independent, positive variates such that*

$E\{\log X_j\} = \mu_j$, $D^2\{\log X_j\} = \sigma_j^2$, and $E\{|\log X_j - \mu_j|^3\} = \omega_j^3$ all exist for every j. Then if

$$\mu_{(n)} = \sum_j^n \mu_j \ , \ \sigma_{(n)}^2 = \sum_j^n \sigma_j^2 \ , \text{and } \omega_{(n)}^3 = \sum_j^n \omega_j^3$$

then the product $\prod_{j=1}^n X_j$ is asymptotically distributed as $\Lambda(\mu_{(n)}, \sigma_{(n)}^2)$, provided

$\omega_{(n)} / \sigma_{(n)} \to 0$, as $n \to \infty$.

Theorem 2 [1]: *A variate subject to the law of proportionate effect tends, for large n, to be distributed as a two-parameter $\Lambda - \text{var}\,iate$, provided that the sequence $X_0, 1+\varepsilon_1, 1+\varepsilon_2, \ldots$, satisfies the conditions of Theorem 1.*

Proof: Given: $X_j - X_{j-1} = \varepsilon_j X_{j-1}$ the law of proportionate effect. We can rewrite this as:

$$\frac{X_j - X_{j-1}}{X_{j-1}} = \varepsilon_j \text{ so that } \sum_{j=1}^n \frac{X_j - X_{j-1}}{X_{j-1}} = \sum_{j=1}^n \varepsilon_j .$$

If the effect at each step is small, we have

$$\sum_{j=1}^n \frac{X_j - X_{j-1}}{X_{j-1}} \approx \int_{X_0}^{X_n} \frac{dX}{X} = \log(X_n) - \log(X_0)$$

giving $\log X_n = \log X_0 + \varepsilon_1 + \ldots + \varepsilon_n$. By the additive form of the central-limit theorem $\log X_n$ is asymptotically normally distributed and hence X_n is asymptotically log normally distributed in a two-parameter form.

PROOF OF LOG NORMAL PROPERTIES

Since X and Y are related by X = ln Y, the distribution function of X and Y are related by:

$\Lambda(y) = N(\ln y)$, where (y > 0) and $\Lambda(y) = 0$, (y ≤ 0).

Thus, the distribution function of Y is :

$$d\Lambda(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2\right)dy, \; (y > 0)$$

The distribution possesses moments of any order, the jth moment about the origin is denote by :

$$\lambda_j' = \int_0^{\infty} y^j d\Lambda(y) = \int_{-\infty}^{\infty} e^{jx} dN(x) = e^{j\mu + .5j^2\sigma^2}.$$

Therefore, the mean = E(Y) is derived using the first moment:

$$E(Y) = \lambda_1(1) = e^{u+.5\sigma^2}$$

Similarly, we can derive the variance, coefficient of skewness, and coefficient of kurtosis.

APPENDIX II

TABLE 1: SICHEL'S T-ESTIMATOR OF THE MEAN $\psi(\beta^2,n)$

# of samples/var ln(x)	5	10	15	20	25	30	50	100
0.1	1.051	1.051	1.051	1.051	1.051	1.051	1.051	1.051
0.2	1.103	1.104	1.104	1.105	1.105	1.105	1.105	1.105
0.3	1.158	1.16	1.6	1.161	1.161	1.161	1.161	1.162
0.4	1.214	1.217	1.218	1.219	1.22	1.22	1.22	1.221
0.5	1.272	1.277	1.279	1.28	1.281	1.281	1.202	1.283
0.6	1.332	1.339	1.343	1.344	1.345	1.346	1.348	1.349
0.7	1.393	1.404	1.409	1.411	1.413	1.414	1.416	1.417
0.8	1.457	1.472	1.478	1.481	1.483	1.484	1.487	1.489
0.9	1.523	1.542	1.55	1.554	1.557	1.558	1.562	1.565
1	1.591	1.615	1.625	1.63	1.634	1.636	1.641	1.645
1.1	1.661	1.691	1.703	1.71	1.714	1.717	1.723	1.728
1.2	1.733	1.77	1.785	1.793	1.798	1.802	1.81	1.816
1.3	1.807	1.851	1.87	1.88	1.886	1.891	1.9	1.908
1.4	1.8884	1.937	1.959	1.971	1.978	1.984	1.995	2.004
1.5	1.963	2.025	2.051	2.065	2.075	2.081	2.095	2.105
1.6	2.044	2.116	2.147	2.164	2.175	2.183	2.199	2.212
1.7	2.128	2.212	2.247	2.267	2.28	2.289	2.3	2.323
1.8	2.214	2.31	2.352	2.375	2.39	2.4	2.422	2.44
1.9	2.303	2.413	2.46	2.487	3.505	2.517	2.542	2.563
2	2.395	2.519	2.574	2.604	2.624	2.63	2.668	2.692
2.1	2.409	2.63	2.691	2.727	2.749	2.765	2.8	2.827
2.2	2.506	2.744	2.814	2.854	2.88	2.89	2.937	2.969
2.3	2.606	2.863	2.942	2.987	3.016	3.037	3.082	3.11
2.4	2.788	2.986	3.074	3.125	3.159	3.182	3.233	3.275
2.5	2.894	3.113	3.212	3.27	3.307	3.334	3.391	3.438
2.6	3.003	3.245	3.356	3.42	3.462	3.492	3.557	3.61
2.7	3.114	3.382	3.505	3.577	3.624	3.658	3.73	3.791
2.8	3.229	3.524	3.661	3.74	3.793	3.831	3.912	3.98
2.9	3.347	3.671	3.822	3.911	3.969	4.011	4.102	4.178
3	3.469	3.824	3.99	4.088	4.153	4.2	4.301	4.387

TABLE 2: SICHEL'S T-ESTIMATOR (t_L) FOR LOWER 5% CONFIDENCE INTERVAL

# of samples/var ln(x)	5	10	15	20	25	30	50	100
0.1	0.83	0.90	0.93	0.94	0.95	0.96	0.98	1.00
0.2	0.79	0.89	0.92	0.95	0.96	0.97	1.00	1.03
0.3	0.76	0.88	0.93	0.96	0.97	0.98	1.02	1.06
0.4	0.75	0.89	0.94	0.97	0.99	1.10	1.05	1.10
0.5	0.74	0.89	0.95	0.99	1.01	1.03	1.09	1.14
0.6	0.74	0.90	0.98	1.01	1.04	1.06	1.02	1.18
0.7	0.74	0.91	0.99	1.04	1.07	1.10	1.16	1.22
0.8	0.73	0.93	1.01	1.06	1.09	1.12	1.19	1.27
0.9	0.74	0.94	1.03	1.09	1.13	1.16	1.23	1.32
1.0	0.74	0.96	1.05	1.11	1.15	1.19	1.28	1.37
1.1	0.74	0.97	1.08	1.14	1.19	1.22	1.32	1.42
1.2	0.74	0.99	1.10	1.17	1.22	1.26	1.36	1.48
1.3	0.75	1.01	1.13	1.20	1.26	1.30	1.41	1.53
1.4	0.76	1.03	1.15	1.23	1.29	1.34	1.46	1.59
1.5	0.77	1.05	1.18	1.27	1.33	1.38	1.51	1.66

TABLE 3: SICHEL'S T-ESTIMATOR (t_U) FOR UPPER 5% CONFIDENCE INTERVAL

#of samples/var ln(x)	5	10	15	20	25	30	50	100
0.1	1.74	1.36	1.27	1.23	1.21	1.17	1.15	1.11
0.2	2.30	1.61	1.45	1.38	1.34	1.31	1.25	1.20
0.3	2.94	1.86	1.64	1.54	1.48	1.44	1.36	1.29
0.4	3.67	2.13	1.84	1.70	1.63	1.57	1.47	1.39
0.5	4.54	2.43	2.05	1.88	1.78	1.72	1.59	1.48
0.6	5.57	2.77	2.26	2.07	1.96	1.88	1.72	1.59
0.7	6.80	3.13	2.53	2.27	2.12	2.03	1.85	1.69
0.8	8.30	3.55	2.81	2.50	2.31	2.20	1.99	1.81
0.9	10.00	4.00	3.11	2.74	2.54	2.40	2.14	1.93
1.0	12.10	4.52	3.44	3.00	2.77	2.61	2.30	2.06
1.1	14.60	5.09	3.80	3.28	3.00	2.84	2.48	2.20
1.2	17.60	5.70	4.20	3.59	3.24	3.05	2.66	2.35
1.3	21.20	6.40	4.60	3.93	3.50	3.17	2.86	2.51
1.4	25.50	7.30	5.10	4.30	3.80	3.60	3.07	2.67
1.5	30.60	8.20	5.60	4.70	4.10	3.90	3.30	2.85

TABLE 4: RESULTS OF SIMULATION EXPERIMENT (DATA POINTS = [5, 10, 15, 20])

		5			10			15			20		
		100 / 1,000 / 10,000			100 / 1,000 / 10,000			100 / 1,000 / 10,000			100 / 1,000 / 10,000		
m = 1.1	NORMAL	90	875	8642	82	878	8821	83	874	8841	94	888	8881
sd = .2	LN	20	178	1746	14	198	1850	11	158	1721	11	151	1506
	LN MISSES	80	820	8234	85	801	8098	88	837	8213	89	844	8456
m = 1.1	NORMAL	89	837	8578	86	878	8763	83	879	8804	86	893	8853
sd = .3	LN	67	552	5788	79	723	7427	80	836	8121	84	869	8589
	LN MISSES	29	423	3986	16	227	2119	14	115	1320	10	77	820
m = 1.1	NORMAL	86	845	8514	96	876	8644	87	871	8741	91	882	8890
sd = .4	LN	71	749	7707	97	873	8749	89	880	8909	89	878	8901
	LN MISSES	24	203	1816	0	44	386	0	15	97	0	4	26
m = 1.1	NORMAL	83	840	8278	80	867	8605	87	883	8654	86	878	8727
sd = .5	LN	83	861	8375	81	892	8926	86	898	8943	86	893	8917
	LN MISSES	12	81	955	2	11	85	0	3	14	0	0	2
m = 1.1	NORMAL	80	818	8204	93	829	8459	87	851	8554	86	880	8622
sd = .6	LN	85	859	8733	93	887	8979	92	903	8953	87	887	8945
	LN MISSES	2	49	476	0	3	23	0	0	1	0	1	0
m = 1.1	NORMAL	75	813	8062	84	850	8328	88	837	8428	85	863	8552
sd = .7	LN	90	886	8904	89	905	8976	89	908	9009	88	898	8979
	LN MISSES	4	26	308	0	2	12	0	0	0	0	0	0
m = 1.1	NORMAL	79	785	7800	81	808	8101	72	812	8320	85	842	8430
sd = .8	LN	90	894	8792	86	881	8943	82	892	9002	95	893	8973
	LN MISSES	3	25	348	0	6	46	1	5	9	0	1	3
m = 1.1	NORMAL	78	765	7528	85	812	7986	77	821	8135	82	837	8329
sd = .9	LN	84	869	8474	85	885	8806	88	897	8930	89	894	8978
	LN MISSES	4	53	624	3	16	233	0	16	105	1	8	67
m = 1.1	NORMAL	71	725	7396	78	761	7745	75	805	8007	76	813	8121
sd = 1.0	LN	84	808	8054	82	836	8327	83	840	8549	91	878	8618
	LN MISSES	7	100	1056	4	79	748	7	64	558	1	39	404
m = 1.1	NORMAL	55	453	4629	50	518	5129	55	520	5559	54	597	5743
sd = 2.0	LN	16	148	1572	1	30	292	1	3	44	0	1	13
	LN MISSES	73	764	7440	91	924	9160	97	974	9739	99	992	9876

TABLE 4: RESULTS OF SIMULATION EXPERIMENT (DATA POINTS = [25, 30, 50, 100])

		25			30			50			100		
		100 / 1,000 / 10,000			100 / 1,000 / 10,000			100 / 1,000 / 10,000			100 / 1,000 / 10,000		
m = 1.1	NORMAL	85	883	8883	91	900	8932	90	891	8946	89	911	8982
sd = .2	LN	13	142	1489	9	133	1244	10	91	811	2	34	319
	LN MISSES	87	848	8458	90	864	8704	90	908	9123	98	964	9620
m = 1.1	NORMAL	88	890	8835	88	887	8885	96	907	8956	88	901	8938
sd = .3	LN	90	906	8871	89	880	8887	98	913	9212	88	914	9086
	LN MISSES	0	46	505	3	32	353	0	6	46	0	0	2
m = 1.1	NORMAL	90	871	8808	94	890	8891	88	867	8896	90	901	8988
sd = .4	LN	90	862	8996	92	868	8713	88	852	8687	84	827	8220
	LN MISSES	0	0	11	0	1	1	0	0	1	0	0	0
m = 1.1	NORMAL	87	868	8761	86	891	8738	90	895	8931	89	899	8913
sd = .5	LN	89	901	9018	87	893	8930	90	896	8983	87	875	8677
	LN MISSES	0	1	0	0	0	0	0	0	0	0	0	0
m = 1.2	NORMAL	91	892	8748	85	861	8717	88	875	8794	92	895	8957
sd = .6	LN	95	914	9090	90	914	9000	88	814	8218	95	898	8877
	LN MISSES	0	0	0	0	0	0	0	0	0	0	0	0
m = 1.2	NORMAL	84	856	8682	92	836	8638	90	872	8791	88	880	8835
sd = .7	LN	90	898	9085	95	882	8973	58	564	5597	87	886	8852
	LN MISSES	0	0	0	0	0	0	0	0	0	0	0	0
m = 1.1	NORMAL	84	845	8491	87	856	8544	87	880	8686	84	871	8854
sd = .8	LN	89	902	9015	90	898	9010	81	725	7147	88	887	8923
	LN MISSES	0	0	0	0	0	0	0	0	0	0	0	0
m = 1.1	NORMAL	89	824	8433	79	815	8491	84	871	8643	92	871	8787
sd = .9	LN	90	896	9004	89	872	8959	85	867	8713	87	891	8961
	LN MISSES	0	5	27	1	0	18	0	0	0	0	0	0
m = 1.1	NORMAL	85	820	8247	83	808	8310	85	844	8463	83	855	8700
sd = 1	LN	89	870	8740	86	888	8850	87	891	8909	90	880	8974
	LN MISSES	61	565	339	3	18	206	3	5	56	0	0	1
m = 1.1	NORMAL	2	34	5927	58	643	6137	63	648	6442	67	676	6903
sd = 2	LN	0	0	2	0	0	0	0	0	0	0	0	0
	LN MISSES	99	999	9956	100	999	9989	100	1000	9999	100	1000	10000

TABLE 5: RESULTS OF SIMULATION EXPERIMENT INTERVAL LENGTHS (DATA POINTS = [5, 10, 15, 20])

		5			10			15			20		
		100	1000	10000	100	1000	10000	100	1000	10000	100	1000	10000
m = 1.1	NORMAL	1.0077	0.9770	0.9810	0.6498	0.6622	0.6601	0.5296	0.5260	0.5332	0.4525	0.4561	0.4572
sd = .2	LN	2.6902	2.7438	2.7541	1.3958	1.3839	1.3875	1.0217	1.0238	1.0234	0.8710	0.8701	0.8711
m = 1.1	NORMAL	1.5795	1.4876	1.5185	1.0265	1.0141	1.0176	0.8018	0.8269	0.8201	0.7083	0.7116	0.7110
sd = .3	LN	2.8896	3.0440	3.0313	1.4308	1.4437	1.4432	1.0326	1.0499	1.0458	0.8778	0.8782	0.8837
m = 1.1	NORMAL	2.0132	2.0503	2.0951	1.5000	1.4107	1.4161	1.1477	1.1436	1.1512	0.9425	0.9909	0.9921
sd = .4	LN	3.7058	3.8619	3.8730	1.8643	1.8016	1.7940	1.2974	1.3055	1.3203	1.0444	1.0878	1.0923
m = 1.1	NORMAL	2.6692	2.7672	2.7125	1.7929	1.8842	1.8659	1.4725	1.5124	1.5140	1.3205	1.3166	1.3177
sd = .5	LN	5.3711	5.3378	5.2539	2.3556	2.4438	2.4139	1.7223	1.8064	1.7840	1.4842	1.4624	1.4500
m = 1.1	NORMAL	3.4073	3.3734	3.4759	2.3166	2.3871	3.3869	2.0399	1.9777	1.9517	1.6650	1.7079	1.6944
sd = .6	LN	6.8158	7.2363	7.6283	3.1378	3.2506	3.2712	2.5187	2.4079	2.3890	1.9641	1.9676	1.9578
m = 1.1	NORMAL	3.8486	4.4185	4.3358	3.1353	3.0110	3.0149	2.3369	2.4697	2.4659	2.0970	2.1330	2.1561
sd = .7	LN	10.0617	10.9590	11.1395	4.5182	4.3891	4.4139	3.0337	3.1259	3.1512	2.4337	2.5620	2.5822
m = 1.1	NORMAL	5.7079	5.2663	5.2920	3.7848	3.6928	3.7031	3.1755	3.1125	3.0795	2.5959	2.6747	2.6946
sd = .8	LN	16.0244	15.4744	15.3794	5.7061	5.7627	5.8857	4.1991	4.1595	4.1421	3.1733	3.3222	3.3562
m = 1.1	NORMAL	6.4124	6.4314	6.4812	5.0251	4.7893	4.6215	3.6639	3.9484	3.8227	3.3037	3.4102	3.3837
sd = .9	LN	20.3915	19.8514	20.2320	8.1269	7.8860	7.6426	5.3039	5.4446	5.3927	4.2137	4.3689	4.3515
m = 1.1	NORMAL	7.2010	7.9367	7.9080	5.5584	5.2855	5.5937	4.4905	4.7961	4.7313	3.7116	4.1066	4.2006
sd = 1.0	LN	23.3612	26.0645	24.6604	9.8636	9.2855	9.3460	6.5697	6.7339	6.6216	4.9998	5.4014	5.4649
m = 1.1	NORMAL	65.1430	70.0162	64.9246	56.4434	53.1410	51.7584	38.1653	45.6931	46.2411	41.4227	45.2396	42.9902
sd = 2.0	LN	47.5946	61.0446	57.3523	12.4607	19.0244	18.0229	12.3036	10.1656	11.5832	11.6785	8.3862	9.5757

TABLE 5: RESULTS OF SIMULATION EXPERIMENT INTERVAL LENGTHS (DATA POINTS = [25, 30, 50, 100])

		25			30			50			100		
		100	1000	10000	100	1000	10000	100	1000	10000	100	1000	10000
m = 1.1	NORMAL	0.4053	0.4128	0.4097	0.3731	0.3749	0.3733	0.2841	0.2858	0.2877	0.2042	0.2029	0.2036
sd = .2	LN	0.7758	0.7815	0.7817	0.6347	0.6283	0.6304	0.5073	0.5075	0.5110	0.3292	0.3283	0.3307
m = 1.1	NORMAL	0.6399	0.6347	0.6371	0.6024	0.5899	0.5772	0.4617	0.4471	0.4472	0.3135	0.3162	0.3168
sd = .3	LN	0.7844	0.7813	0.7883	0.6429	0.6418	0.6352	0.5099	0.5133	0.5117	0.3305	0.3298	0.3305
m = 1.1	NORMAL	0.8500	0.8833	0.8844	0.8180	0.8228	0.8102	0.6328	0.6279	0.6254	0.4411	0.4448	0.4450
sd = .4	LN	0.9379	0.9635	0.9712	0.8399	0.8596	0.8381	0.6426	0.6496	0.6464	0.4413	0.4429	0.4452
m = 1.1	NORMAL	1.1624	1.2095	1.1746	1.0483	1.0829	1.0732	0.8466	0.8421	0.8349	0.5948	0.5899	0.5932
sd = .5	LN	1.2877	1.3379	1.3061	1.1517	1.1703	1.1680	0.8631	0.8800	0.8740	0.5965	0.5930	0.5931
m = 1.1	NORMAL	1.5733	1.5408	1.5176	1.3867	1.3883	1.3932	1.0790	1.0740	1.0805	0.7738	0.7693	0.7716
sd = .6	LN	1.7976	1.7509	1.7243	1.5499	1.5316	1.5427	0.9518	0.9099	0.9387	0.7753	0.7908	0.7901
m = 1.1	NORMAL	1.9064	1.9407	1.9309	1.7943	1.7401	1.7660	1.3718	1.3842	1.3806	0.9846	0.9932	0.9867
sd = .7	LN	2.2013	2.2171	2.2382	1.9639	1.9539	1.9891	0.6828	0.6148	0.6102	0.9702	0.9827	0.9801
m = 1.1	NORMAL	2.6435	2.3765	2.4314	2.2975	2.2434	2.2325	1.7773	1.7502	1.7535	1.2697	1.2436	1.2527
sd = .8	LN	3.0597	2.7841	2.9015	2.5526	2.5994	2.5648	1.8219	1.5171	1.4625	1.2922	1.2726	1.2839
m = 1.1	NORMAL	3.1491	2.9615	3.0628	2.7515	2.7163	2.7920	2.1159	2.2376	2.2080	1.5583	1.5732	1.5939
sd = .9	LN	3.9341	3.6424	3.7326	3.1432	3.2235	3.3008	2.3537	2.4246	2.3528	1.6480	1.6329	1.6403
m = 1.1	NORMAL	3.9433	3.6881	3.8336	3.5646	3.4366	3.5036	2.8411	2.8634	2.7771	2.0479	2.0389	2.0352
sd = 1.0	LN	4.6142	4.6422	4.6353	4.2148	4.0324	4.1614	3.1266	3.1201	3.0885	2.0282	2.0704	2.0877
m = 1.1	NORMAL	41.6366	61.4185	42.3737	48.1140	40.6428	39.7120	34.6950	34.6623	33.5489	31.0025	27.4787	27.9365
sd = 2.0	LN	3.8951	5.1911	8.0587	0.0000	6.1233	5.9729	0.0000	0.0000	4.9740	0.0000	0.0000	0.0000

APPENDIX III

FORTRAN PROGRAM FOR SIMULATION EXPERIMENT

```
****************************************************************
* PROGRAM NAME:  RANDLNSM.F
*
* THIS PROGRAM GENERATES THE DESIRED NUMBER OF DATA SAMPLES
FROM
* A LOGNORMAL DISTRIBUTION WITH THE SPECIFIED MEAN AND STANDARD
* DEVIATION.  THE EXACT MEAN IS COMPUTED (EXP(XM + 1/2 S^2).  THE
* MEAN OF THE SAMPLE IS THEN ESTIMATED WITH A SPECIFIED
* CONFIDENCE INTERVAL USING "NORMAL" THEORY.
* THE SAMPLE POINTS ARE THEN LOG TRANSFORMED.  THE MEAN OF THE
* TRANSFORMED POINTS IS ESTIMATED USING SICHEL'S THEORY.
* THE TOTAL NUMBER OF TIMES THE EXACT MEAN IS CONTAINED IN EACH
* EACH INTERVAL IS RECORDED.
*
****************************************************************

      PROGRAM MAINLN

** INITIALIZE VARIABLES

      REAL XM, S, R(100), NRSQ, SLNM, XMEAN, XS, TVAL, XLNM, SS
      REAL SSUM, XSD, LNR(100), LNXS, LNXMEAN, LNSS, LNSSUM
      REAL LNXVAR, LOW5TAB(16,8), HI5TAB(16,8), CLNLO, CLNUP
      REAL TLO, TUP, INTLN(10000), INTLLN(10000)
      REAL NOSIMSQ, LNS, LLNS, XNLEN, XLNLEN, SS1, SS2
      REAL SS1SUM, SS2SUM, XSDNL, XSDLNL, MISSSQ
         INTEGER NR, INSEED, NOUT, CI, NOSIM, TCOL, TROW
      INTEGER NHITS, LNHITS, MISSES
         EXTERNAL RNLNL, RNSET, UMACH


** LOAD TABLES - lower 5% then upper 5%

      DATA LOW5TAB/ .83,.79,.76,.75,.74,.74,.74,.73,.74,.74,.74,
    /     .75,.76,.77,.77,0,
    /     .90,.89,.88,.89,.89,.90,.91,.93,.94,.96,.97,
    /     .99,1.01,1.03,1.05,0,
    /     .93,.92,.93,.94,.95,.98,.99,1.01,1.03,1.05,1.08,
    /     1.10,1.13,1.15,1.18,0,
    /     .94,.95,.96,.97,.99,1.01,1.04,1.06,1.09,1.11,1.14,
    /     1.17,1.20,1.23,1.27,0,
```

```
     /   .95,.96,.97,.99,1.01,1.04,1.07,1.09,1.13,1.15,1.19,
     /   1.22,1.26,1.29,1.33,0,
     /   .96,.97,.98,1.01,1.03,1.06,1.10,1.12,1.16,1.19,1.22,
     /   1.26,1.30,1.34,1.38,0,
     /   .98,1.00,1.02,1.05,1.90,1.02,1.16,1.19,1.23,1.28,
     /   1.32,1.36,1.41,1.46,1.51,0,
     /   1.00,1.03,1.06,1.10,1.14,1.18,1.22,1.27,1.32,1.37,
     /   1.42,1.48,1.53,1.59,1.66,0/

      DATA HI5TAB/ 1.74,2.30,2.94,3.67,4.54,5.57,6.80,8.30,
     /   10.0,12.1,14.6,17.6,21.2,25.5,30.6,0,
     /   1.36,1.61,1.86,2.13,2.43,2.77,3.13,3.55,4.00,4.52,
     /   5.09,5.70,6.40,7.30,8.20,0,
     /   1.27,1.45,1.64,1.84,2.05,2.26,2.53,2.81,3.11,3.44,
     /   3.80,4.20,4.60,5.10,5.60,0,
     /   1.23,1.38,1.54,1.70,1.88,2.07,2.27,2.50,2.74,3.00,
     /   3.28,3.59,3.93,4.30,4.70,0,
     /   1.21,1.34,1.48,1.63,1.78,1.96,2.12,2.31,2.54,2.77,
     /   3.00,3.24,3.50,3.80,4.10,0,
     /   1.17,1.31,1.44,1.57,1.72,1.88,2.03,2.20,2.40,2.61,
     /   2.84,3.05,3.17,3.60,3.90,0,
     /   1.15,1.25,1.36,1.47,1.59,1.72,1.85,1.99,2.14,2.30,
     /   2.48,2.66,2.86,3.07,3.30,0,
     /   1.11,1.20,1.29,1.39,1.46,1.59,1.69,1.81,1.93,2.06,
     /   2.20,2.35,2.51,2.67,2.85,0/

** SET INPUT DATA
*
      INSEED = 52334
      NR = 10
      NOSIM = 10
      XM = 1.1
      S = 0.2
      CI = 90
      TVAL =  1.833
      TCOL = 2
      NHITS = 0
      LNHITS = 0
      MISSES = 0
      NRSQ = NR
      NRSQ = SQRT (NRSQ)
      NOSIMSQ = NOSIM
      NOSIMSQ = SQRT (NOSIMSQ)

** COMPUTE AND DISPLAY ENTERED INPUT AND EXACT MEAN
      XLNM = EXP(XM + 0.5*S*S)
         PRINT 99990, NR
99990 FORMAT (' ','YOU HAVE REQUESTED ',I3,' DATA POINTS')
```

```
      PRINT *, 'GENERATED FROM A LN SAMPLE WITH MEAN:  ', XM
      PRINT *, '             AND STANDARD DEVIATION:  ', S
      PRINT *, ' '
      PRINT *, 'THE EXACT MEAN BASED ON LOGNORMAL THEORY FOR
     / THIS DATA IS:  ', XLNM
      PRINT *, ' '
**
* USE THIS PRINT IF YOU WANT TO PRINT BOTH CI RESULTS TOGETHER
      PRINT *, '            THE 90% CONFIDENCE INTERVAL '
      PRINT *, '                      FOR'
      PRINT *, '  NORMAL THEORY                    LN THEORY'
*      PRINT *, ' '


      DO 5 K = 1, NOSIM
** RESET COUNTERS
      XS = 0
      SSUM = 0
      LNXS = 0
      LNSSUM = 0
      LNS = 0
      LLNS = 0
      SS1SUM = 0
      SS2SUM = 0

**** GENERATE DATA POINTS

         CALL UMACH (2,NOUT)
         CALL RNSET (INSEED)
         CALL RNLNL (NR, XM, S, R)

* USING THIS PRINT IF YOU WANT TO PRINT GENERATED DATA POINTS
*      PRINT *, 'THE GENERATED LOGNORMAL DATA POINTS ARE: '
*      PRINT *, ' '
*      PRINT *, (R(I), I=1,NR)

*** GENERATE A SPECIFIED CONFIDENCE INTERVAL USING "NORMAL"
THEORY

** COMPUTE MEAN AND STANDARD DEVIATION

      DO 15 I = 1, NR
       XS = XS + R(I)
 15   CONTINUE
      XMEAN = XS / NR

      DO 20 I = 1, NR
       SS = (R(I) - XMEAN)**2
```

```
      SSUM = SSUM + SS
 20   CONTINUE
     XSD = SQRT (SSUM) / NRSQ

** COMPUTE AND PRINT THE CONFIDENCE INTERVAL

     CUP = XMEAN + TVAL*(XSD / NRSQ)
     CLO = XMEAN - TVAL*(XSD / NRSQ)
     INTLN(K) = CUP - CLO

* USE THIS PRINT IF YOU WANT TO DISPLAY CI OF NORMAL W/GENERATED
POINTS
*      PRINT 99995, CI
*99995  FORMAT (' ','THE ',I2,'% CONFIDENCE INTERVAL YOU REQUESTED IS:')
*      PRINT *, '[' , CLO, ',' , CUP , ']'


*** COMPUTE CI BASES ON LOGNORMAL THEORY

* TAKE LOG OF DATA

     DO 30 I = 1, NR
       LNR(I) = LOG(R(I))
 30   CONTINUE

* USE THIS PRINT IF YOU WANT TO DISPLAY TRANSOFMRED DATA POINTS
*      PRINT *, 'THE LN OF THE GENERATED DATA POINTS ARE: '
*      PRINT *, ' '
*      PRINT *, (LNR(I), I = 1, NR)

** COMPUTE MEAN AND S.D. FOR LN DATA

     DO 40 I = 1, NR
       LNXS = LNXS + LNR(I)
 40   CONTINUE
     LNXMEAN = LNXS / NR

     DO 50 I = 1, NR
       LNSS = (LNR(I) - LNXMEAN)**2
       LNSSUM = LNSSUM + LNSS
 50   CONTINUE
     LNXVAR = LNSSUM / NR

*** COMPUTE C.I. USING LN THEORY

*** LOOK UP T FACTORS FOR UPPER/LOWER CI IN EACH TABLE

     IF (LNXVAR .LE. 0.05) THEN
```

```
*        PRINT *, 'LNXVAR SMALLER THAN TABLE'
        MISSES = MISSES + 1
        TROW = 16
      ELSEIF (LNXVAR .LE. 0.15) THEN
        TROW = 1
      ELSEIF (LNXVAR .LE. 0.25) THEN
        TROW = 2
      ELSEIF (LNXVAR .LE. 0.35) THEN
        TROW = 3
      ELSEIF (LNXVAR .LE. 0.45) THEN
        TROW = 4
      ELSEIF (LNXVAR .LE. 0.55) THEN
        TROW = 5
      ELSEIF (LNXVAR .LE. 0.65) THEN
        TROW = 6
      ELSEIF (LNXVAR .LE. 0.75) THEN
        TROW = 7
      ELSEIF (LNXVAR .LE. 0.85) THEN
        TROW = 8
      ELSEIF (LNXVAR .LE. 0.95) THEN
        TROW = 9
      ELSEIF (LNXVAR .LE. 1.05) THEN
        TROW = 10
      ELSEIF (LNXVAR .LE. 1.15) THEN
        TROW = 11
      ELSEIF (LNXVAR .LE. 1.25) THEN
        TROW = 12
      ELSEIF (LNXVAR .LE. 1.35) THEN
        TROW = 13
      ELSEIF (LNXVAR .LE. 1.45) THEN
        TROW = 14
      ELSEIF (LNXVAR .LE. 1.55) THEN
        TROW = 15
      ELSE
*        PRINT *, 'LNXVAR LARGER THAN TABLE'
        MISSES = MISSES + 1
        TROW = 16
      ENDIF
      TUP = HI5TAB(TROW,TCOL)
      TLO = LOW5TAB(TROW,TCOL)
      CLNUP = EXP(LNXMEAN) * TUP
      CLNLO = EXP(LNXMEAN) * TLO
      INTLLN(K) = CLNUP - CLNLO

* USE THIS PRINT IF YOU WANT TO PRINT CI FOR BOTH THEORY TOGETHER
*     PRINT *, ' '
*     PRINT 99997, CLO,CUP,CLNLO,CLNUP
*99997  FORMAT(' ','[',F8.4,',',F8.4,']',20X,'[',F8.4,',',F8.4,']')
```

```
*      PRINT 99996, CI
*99996  FORMAT (' ','THE ',I2,'% CONFIDENCE INTERVAL BASED ON LN
*   /      THEORY IS:')
*      PRINT *, '[' , CLNLO, ',' , CLNUP , ']'

** COUNT NUMBER OF HITS FOR EXACT MEAN

      IF (XLNM .GE. CLO) THEN
        IF (XLNM .LE. CUP) THEN
          NHITS = NHITS + 1
        ENDIF
      ENDIF

      IF (XLNM .GE. CLNLO) THEN
        IF (XLNM .LE. CLNUP) THEN
          LNHITS = LNHITS + 1
        ENDIF
      ENDIF

      INSEED = INSEED + 10112
** REPEAT SIMULATION RUN THE DESIRED NUMBER OF TIMES
 5    CONTINUE

*** STEP 2 -- COMPUTE FINAL RESULTS

** COMPUTE MEAN AND STANDARD DEV OF INTERVAL LENGTHS

      DO 60 I = 1, NOSIM
        LNS = LNS + INTLN(I)
        LLNS = LLNS + INTLLN(I)
 60   CONTINUE

*** MODIFY NUMBER OF VALID INTERVALS FOR LOGNORMAL THEORY

      MISSSQ = NOSIM - MISSES
      MISSSQ = SQRT (MISSSQ)
***********

      XNLEN = LNS / NOSIM
      IF (NOSIM .GT. MISSES) THEN
        XLNLEN = LLNS / (NOSIM - MISSES)
        DO 70 I = 1, NOSIM
          SS1 = (INTLN(I) - XNLEN)**2
          SS1SUM = SS1SUM + SS1
          IF (INTLLN(I) .GT. 0) THEN
            SS2 = (INTLLN(I) - XLNLEN)**2
            SS2SUM = SS2SUM + SS2
          ENDIF
```

```
 70     CONTINUE

        XSDNL = SQRT (SS1SUM) / NOSIMSQ
        XSDLNL = SQRT (SS2SUM) / MISSSQ
      ELSE
        XSDLNL = 0
      ENDIF

** PRINT OUT FINAL RESULTS
************
      PRINT *, '  '
      PRINT *, 'THE NUMBER OF HITS FOR NORMAL THEORY IS:  ', NHITS
      PRINT 99998, XNLEN, XSDNL
99998  FORMAT(' ','AVERAGE LENGTH:  ',F8.4,10X,'S.D:  ',F8.4)
      PRINT*, '  '
      PRINT *, 'THE NUMBER OF HITS FOR LN THEORY IS:      ', LNHITS
      PRINT 99998, XLNLEN, XSDLNL
      PRINT *, '  '
      PRINT *, '  NUMBER OF MISSES ON TABLE IS:  ', MISSES
      PRINT *, ' '
      PRINT *, 'THIS IS THE END OF THE RUN'
         END

*************************************************************************
```

REFERENCES

[1] Aitchison, J. and Brown, J. A. C. (1976). The Lognormal Distribution. Cambridge University Press.

[2] Bliss, C. I. (1935). The Calculation of the Dosage-Mortality curve. Ann. Appl. Biol. V 22. p 134.

[3] Champernowne, D. G. (1954). A Model of Income Distribution. Econ. Journal. V 63. p 318.

[4] Clark, A. J. (1933). The Mode of Action of Drugs upon Cells. London: Edward Arnold.

[5] Cramer, H. (1946). Mathematical Methods of Statistics. Princeton Mathematical Series, No 9. Princeton University Press.

[6] Finney, D. J. (1949). On the distribution of a Variate whose Logarithm is Normally Distributed. Supplement of the Journal of the Royal Statistical Society, 7. V 2. pp 155-61.

[7] Galton, F. (1879). The Geometric Mean in Vital and Social Statistics. Proceedings of the Royal Society. V. 29. p 365.

[8] Hemmingsen, A. M. (1933). A statistical Analysis of the Differences in body size of related species. Vidensk. Medd. Dansk. naturh. Foren. Kbh. 98, 125.

[9] IMSL/STAT LIBRARY Reference Manual. (1987).

[10] Kapteyn, J. C. (1903). Skew Frequency Curves in Biology and Statistics. Astronomical Laboratory, Groningen: Noordhoff.

[11] Kapteyn, J. C. and Van Uven, M. J. (1916). Skew Frequency Curves in Biology and Statistics. Groningen: Hoitsema Bros.

[12] Lorenz, M. O. (1905). Methods of Measuring the Concentration of Wealth. Publications of the American Statistical Association. New Series, V 70. p 209.

[13] Matheron, G. F. (1970). Estimer et choisir, Fascicule 7. Les Cahiers du Dentre de Morphologie Mathematique de Fontainebleau, Ecole Superieure des Mines de Paris.

[14] McAlister, D. (1879). The Law of the Geometric Mean. Proceedings of the Royal Society. V. 29. p 367.

[15] Sichel, H. S. (1966). The Estimation of Means and Associated Confidence Limits for Small Samples from Lognormal Populations. Proceedings of the Symposium on Mathematical Statistics and Computer Applications in Ore Valuation. South African Institute of Mining and Metallurgy, Johannesburg. pp 106-123.

[16] United States Environmental Protection Agency. (1991). GEO-EAS 1.2.1 User's Guide

[17] United States Environmental Protection Agency. (1987). Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Soil Media (EPA230/02-89/042).