SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
Volume 8 No 1
ISSN 2054-7390

Statistical Inference for k $(k \geq 3)$ Lognormal Means from Left Censored Data

Abou El-Makarim A. Aboueissa
Department of Mathematics and Statistics
University of Southern Maine
96 Falmouth Street, P.O. Box 9300
Portland, Maine 04104-9300, USA

February 15, 2020

Abstract

The occurrence of censored data due to less than detectable measurements is a common problem with environmental data such as quality and quantity monitoring applications of water, soil, and air samples. The log-normal distribution is one of the most common distributions used for modeling skewed and positive data. Over the past decades, various methods for comparing the parameters of two lognormal distributions in the presence censored data have been proposed. Some of them are differing in terms of how the statistic test adjust to accept or to reject the null hypothesis. As a model distribution of measured environmental and/or biomedical data, log-normal distribution is considered. Logmormal means can be compared either by confidence intervals or hypothesis testing procedures. In this article, a new test procedure for comparing the means of k $(k >= 3)$ lognormal distributions in the presence of left-censored data is introduced and evaluated. Asymptotic chi-square test is used in the proposed test procedure. A simulation study was performed to examine the power and the size of the proposed test procedure introduced in this article utilizing a computer program written in the R language. We find analytically that the considered test procedure is doing well through comparing the size and power of the statistic test.

Key words: multiple detection limits, left censored data, normal and lognormal distributions, maximum likelihood estimators, expectation maximization algorithm, likelihood ratio test.

1 Introduction

Left-censored data commonly arise in environmental contexts. Left-censored observations (observations reported as leass than a detection limit DL can occur when the substance or attribute being measured is either absent or exists at such low concentrations that the substance is not present above the DL. Data sets containing left-censored observations are referred to as left-censored data. In many environmental applications the distribution of variables such as concentration, inhalation, digestion, and consumption rates are positive and skewed to the right. Hence, censored observations occur between zero and DL. In some instances a log transformation can provide a more natural scale to analyze such measurements. Many environmental data sets are characterized by a small number of high concentrations and a large

DOI: 10.14738/tmlai.81.7896
Publication Date: 4th March, 2020
URL: http://dx.doi.org/10.14738/tmlai.81.7896

number of low concentrations and are often right-skewed (Shumway et al., 1989). The log-normal distribution is positively skewed and hence can incorporate the few unusually high measurements of such environmental data in its long right-hand tail. For this reason the log-normal distribution is often applied to environmental data (Gilbert, 1987). While analyzing environmental and exposure data, a very common phenomenon is the occurrence of non-detects, i.e., observations below an analytical detection limit (DL), resulting in Type I singly left censored samples. Detection limit is the lowest concentration level that can be determined to be statistically different from a blank. The presence of observations below the DL significantly complicates the data analysis. Faced with such data, several strategies have been recommended for data analysis. One approach consists of replacing the below DL values with a constant such as $\frac{DL}{2}$, and using methods available for complete samples. It is easy to demonstrate that the conclusions resulting from this routine practice can be seriously imperfect; in fact, the conclusions may depend on the substitution value used for replacing sample values below the DL. In general, censoring means that observations at one or both tails are not available. Left-censored data commonly arise in environmental contexts. Left-censored data (data reported as less than detection limit) can occur when the substance or attribute being measured is either absent or exists at such low concentrations that the substance is not present above the DL level. Data sets containing left-censored observations are referred to as left-censored data. A sample is multiply censored if there are several detection limits. When more than two distinct detection limits DL_1, DL_2 , ..., DL_{m_c} ($m_c \geq 3$) are reported, the data are said to be multiply-left-censored. Samples to be considered in this paper are those that are Type I multiple-left-censored. Suppose that a sample of n data points is given of which m data points are non-censored (fully measured), and the remaining $m_c = n - m$ observations are left-censored with multiple detection limits DL_1, DL_2, ..., DL_{m_c}. In such Type I censored samples detection limits are fixed, whereas m and m_c are random. It is common to have environmental data contains detection limits. Multiple censoring commonly occurs with environmental data because detection limits can change over time (e.g., because of analytical improvements), or detection limits can depend on the type of sample or the background matrix. Millard and Deverel (1988) give three possible causes for multiple censoring on the left when measuring the concentration of zinc in shallow groundwater. First, there may be more than one method available, and each method may be optimal in different ranges of zinc concentration. A second cause involves the amount of dilution that a lab technician may use. Note that the detection limit depends on the number of dilutions. A third cause may be decreasing detection limits over time as the measurement technique improves. In many environmental applications the distribution of variables such as chemical concentration, inhalation, digestion, and consumption rates are positive and skewed to the right. Hence, censored observations occur between zero and DL. In some instances a log transformation can provide a more natural scale to analyze such measurements.

Nondetect values can cause an especially difficult problem when the goal is to compare $k(k \geq 3)$ different populations. There has been a great deal of literature on the subject of the statistical inference of the parameters of normal and log-normal populations from both fully measured and censored data. Gupta and Li (2006) developed a score test for testing the equality of the means of two independent log-normal populations from fully measured data. Zhou et al (1997) considered two methods for comparing the means of two independent log-normal non-censored samples. Harris (1991) considered two parametric and two non-parametric methods for testing the equality of medians of two independent log-normal distributions when some data are left-censored. Paul and Gary (2007) compare the performance of several methods for statistically analyzing censored data sets when estimating the 95th percentile and the mean of right-skewed occupational exposure data. Krishnamoorthy et al (2014) proposed tests and confidence intervals for the ratio of the two means of two log-normal distributions, based on pivotal quantities involving the maximum likelihood estimators. Other suggested methods for comparing the means of two log-normal distributions are discussed in Krishnamoorthy et al (2014, 2011, 2007, 2006, 2003). Some of these methods are based on the generalized p-value and generalized confidence intervals, and others are based on the generalized test variable. Aboueissa (2015) introduced a test procedure for comparing the means of two independent log-normal populations when data is singly censored. Abdollahnezhad et al (2012) introduced a new method of test for comparing the means of two log- normal populations through the generalized measure of evidence to have against the null hypothesis. Prentice (1978) developed linear rank tests with right censored data. Millard and Deverel (1988) adapted several existing right censored non-parametric procedure so that they can be used in environmental setting with left-censored data. Methods for the estimation of the log-normal parameters for one-sample cases where there may exist left-censored data are discussed by El-Shaarawi (1989). Stoline (1993) extended results first suggested by Harris (1991) and proposed a procedure for comparing medians of two independent log-normal distributions where some data may be left-censored. Stoline (1993)used the Expectation Maximization (EM) algorithm introduced by Dempster et al. (1977) to calculate the maximum likelihood estimates of population parameters μ and σ. Other suggested methods for estimating population parameters from censored samples are discussed in Marco (2005), Jin et al (2011), Gibbons (1994), Gleit (1985), El-Shaarawi and Esterby (1992), Elshaarawi and Dolan (1989), Gilbert (1987), Stavros (2004) and Schneider (1986).

The purpose of this paper is to develop a parametric procedure to test the equality of k $(k \geq 3)$ lognormal distribution means when data are multiply left-censored. This procedure may be used to compare the concentration of a pollutant in shallow groundwater among $k \geq 3$ geological zones found in different geographical areas. For example, the pollutant may be copper in low concentrations (micrograms per liter

of water) in different geographical areas that have different types of soil. The EM algorithm will be used to obtain the maximum likelihood estimates of population parameters under different hypotheses. A simulation study was performed to inspect the size and the power of the proposed test procedure. To facilitate the application of this procedure, a computer program is written in the R language which calculates the maximum likelihood estimates, and asymptotic chi-square test statistics and their p-values.

2 Assumptions and Notations

Assume that there exists k random samples of n_i data values: $y_{i1}, y_{i2}, ..., y_{im_i}$, $y_{im_i+1}, ..., y_{in_i}$ taken from k independent log-normal populations $LN(\mu_i, \sigma_i)$ for $i = 1, 2, ..., k$. Where $LN(\mu, \sigma)$ denotes a log-normally distributed variable y with the probability density function

$$f(y; \mu, \sigma) = \frac{1}{y\ \sigma\sqrt{2\pi}}\ e^{-\frac{(log\ y-\mu)^2}{2\sigma^2}}\ , \quad \text{for} \quad y > 0,$$

where $-\infty < \mu < \infty$ and $\sigma > 0$. For convenience, for each sample i let us assume that the first m_i observations $y_{i1}, y_{i2}, ..., y_{im_i}$ are non-censored (fully measured) and the remaining $m_{c_i} = n_i - m_i$ observations are left-censored for $i = 1, 2, ..., k$. For left censored observations, it is assumed that for each sample i it is only known that $y_{ij} < LDL_{ij}$ for $j = m_i + 1, ..., n_i$ (or $j = 1, 2, ..., m_{c_i}$) and $i = 1, 2, ..., k$. The parameters for the i^{th} log-normal population can be expressed as functions of the parameters μ_i and σ_i as:

$$\begin{aligned}
\textbf{mean:} \quad & \mu_{y_i} = e^{\mu_i + \frac{\sigma_i^2}{2}} \\
\textbf{medain:} \quad & M_{y_i} = e^{\mu_i} \\
\textbf{variance:} \quad & \sigma^2_{y_i} = \gamma_i(\gamma_i - 1)\ e^{2\mu_i} \\
\textbf{skewness:} \quad & s_{y_i} = (\gamma_i + 2)\sqrt{(\gamma_i - 1)}
\end{aligned}$$

where $\gamma_i = e^{\sigma_i^2}$ for $i = 1, 2, ..., k$.

Let

$$x_{ij} = \begin{cases} \log(y_{ij})\ , & \text{for } i = 1, 2, ..., k \text{ and } j = 1, 2, ..., m_i, \\ DL_{ij} = \log(LDL_{ij})\ , & \text{for } i = 1, 2, ..., k \text{ and } j = m_i + 1, m_i + 2, ..., n_i. \end{cases}$$

where LDL_{ij} are the detection limits in the i^{th} log-normal sample and $m_i + m_{c_i} = n_i$ and $j = m_i + 1, m_i + 2, ..., n_i$ for $i = 1, 2, ..., k$.

To simplify the presentation in this paper, the analysis is described and illustrated by reference to the analysis of normally distributed data, though this condition occurs infrequently in typical environmental data analysis. However, it is frequently necessary to transform real environmental data before analysis; typically the logarithmic transformation of $x_{ij} = \log(y_{ij})$ is used, although other transformations are possible. When the logarithmic or other transformation is used prior to censored data set analysis, it is necessary to transform the analysis results back to the original scale of measurement following parameter estimation. For each sample i let

$$\bar{x}_{m_i} = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij} \ , \quad \text{and} \quad s^2_{m_i} = \frac{1}{m_i}\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_{m_i})^2$$

be the sample mean and sample variance of the m_i non-censored observations $x_{i1}, x_{i2}, ..., x_{im_i}$, for $i = 1, 2, ..., k$. Let the functions $\phi(.)$ and $\Phi(.)$ be the *pdf* and *cdf* of the standard unit normal. Define

$$\Phi(\xi_{ij}) = \int_{-\infty}^{\xi_{ij}} \phi(t)dt \ , \quad \text{where} \ \ \xi_{ij} = \frac{DL_{ij} - \mu_i}{\sigma_i} \ ,$$

for $\quad i = 1, 2, ..., k$ and $j = m_i + 1, m_i + 2, ..., n_i$.

We also define

$$W(x) = \frac{\phi(x)}{\Phi(x)} \ \text{ and } \ z_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i} \ \text{ for } \ i = 1, 2, ..., k \ \text{ and } \ j = 1, 2, ..., m_i.$$

The likelihood function of the samples under consideration is given by:

$$\begin{aligned} L(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k) &= \prod_{i=1}^{k}\left(\prod_{j=1}^{m_i}\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x_{ij}-\mu_i}{\sigma_i})^2}\prod_{j=m_i+1}^{n_i}P(x_{ij} < DL_{ij})\right) \\ &= \prod_{i=1}^{k}\left(\prod_{j=1}^{m_i}\frac{1}{\sigma_i}\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{k}(\frac{x_{ij}-\mu_i}{\sigma_i})^2}\prod_{j=m_i+1}^{n_i}P(\frac{x_{ij}-\mu_i}{\sigma_i} < \frac{DL_{ij}-\mu_i}{\sigma_i})\right) \\ &= \prod_{i=1}^{k}\left(\prod_{j=1}^{m_i}\frac{1}{\sigma_i}\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(\frac{x_{ij}-\mu_i}{\sigma_i})^2}\prod_{j=m_i+1}^{n_i}P(Z < \frac{DL_{ij}-\mu_i}{\sigma_i})\right) \end{aligned}$$

which can be written as

$$\begin{aligned} L(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k) &= \prod_{i=1}^{k}\left[\prod_{j=1}^{m_i}\frac{1}{\sigma_i}\phi\left(\frac{x_{ij}-\mu_i}{\sigma_i}\right)\prod_{j=m_i+1}^{n_i}\Phi\left(\frac{DL_{ij}-\mu_i}{\sigma_i}\right)\right] \\ &= \prod_{i=1}^{k}\left[\prod_{j=1}^{m_i}\frac{1}{\sigma_i}\phi\left(z_{ij}\right)\prod_{j=m_i+1}^{n_i}\Phi\left(\xi_{ij}\right)\right] \end{aligned} \tag{2.1}$$

where

$$z_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i} \quad \text{and} \quad \xi_{ij} = \frac{DL_{ij} - \mu_i}{\sigma_i}$$

for $i = 1, 2, ..., k$ and $j = m_i + 1, ..., n_i$.

Four hypotheses are possible:

$\underline{H_{0N}}$: $\mu_1 = \mu_2 = ... = \mu_k = \mu$ and $\sigma_1 = \sigma_2 = ... = \sigma_k = \sigma$; (overall homogeneity).
$\underline{H_{A1N}}$: $\mu_i \neq \mu_j$ and $\sigma_i \neq \sigma_j$ for all $i \neq j$; (overall heterogeneity),
$\underline{H_{A2N}}$: $\mu_i \neq \mu_j$ and $\sigma_1 = \sigma_2 = ... = \sigma_k = \sigma$ for all $i \neq j$; (mean heterogeneity, variance homogeneity),
$\underline{H_{A3N}}$: $\mu_1 = \mu_2 = ... = \mu_k = \mu$ and $\sigma_i \neq \sigma_j$ for all $i \neq j$; (mean homogeneity, variance heterogeneity)

The k log-normal population means are confirmed equal whenever no evidence was available to reject the null hypothesis $H_{0LN} : \mu_{y_1} = \mu_{y_2} = ... = \mu_{y_k}$ in favor of the alternative hypothesis $H_{ALN} : \mu_{y_1} \neq \mu_{y_2} \neq ... \neq \mu_{y_k}$. Equivalently the k log-normal population means are confirmed equal whenever the null hypothesis H_{0N} is accepted in favor of one of the hypotheses H_{A1N}, H_{A2N} or H_{A3N}.

Two tests are considered in this article. Test 1: H_{0N} versus H_{A1N}, overall homogeneity versus overall heterogeneity, and Test 2: H_{0N} versus H_{A2N}, overall homogeneity versus mean heterogeneity and variance homogeneity. The other tests H_{A2N} versus H_{A1N} and H_{A3N} versus H_{A1N} can be similarly derived and computed.

3 Maximum Likelihood Estimates of Population Parameters

In this section the maximum likelihood estimates of population parameters μ_i and σ_i, for $i = 1$ and 2, are derived under each of the hypotheses H_{0N}, H_{A1N} and H_{A2N}. The derivations of these estimates are now described.

3.1 Maximum Likelihood Estimates under H_{0N}

Under the hypothesis H_{0N}, x_{ij}, for $i = 1, 2, ..., k$ and $j = 1, 2, ..., n_i$, are assumed to be normally distributed with mean μ and standard deviation σ. That is, it is assumed that there exists a random sample of $n = n_1 + n_2 + ... + n_k$ data values taken from a normal population with mean μ and standard deviation σ. For convenience, for each sample i let us assume that the first m_i observations $x_{i1}, x_{i2}, ..., x_{im_i}$ are non-censored

(fully measured) and the remaining $m_{c_i} = n_i - m_i$ observations are left-censored for $i = 1, 2, ..., k$. For left censored observations, it is assumed that for each sample i it is only known that $x_{ij} < DL_{ij}$ for $j = m_i+1, ..., n_i$ (or $j = 1, 2, ..., m_{c_i}$) and $i = 1, 2, ..., k$.

The likelihood function $L_{H_{0N}}(\mu, \sigma)$ under H_{0N} is given by:

$$L_{H_{0N}}(\mu, \sigma) = \prod_{i=1}^{k} \left(\sigma^{-m_i} (\frac{1}{\sqrt{2\pi}})^{m_i} \; e^{-\frac{1}{2}\sum_{j=1}^{m_i}(\frac{x_{ij}-\mu}{\sigma})} \prod_{j=m_i+1}^{n_i} \Phi\left(\frac{DL_{ij}-\mu}{\sigma}\right) \right) \tag{3.1}$$

Hence, the corresponding log-likelihood function $\ell_{H_{0N}}(\mu, \sigma) = \log(L_{H_{0N}}(\mu, \sigma))$ of (3.1) is given by:

$$\begin{aligned} \ell_{H_{0N}}(\mu, \sigma) = \sum_{i=1}^{k} \left(-m_i \log \sigma + m_i \log(\frac{1}{\sqrt{2\pi}}) - \frac{1}{2} \sum_{j=1}^{m_i} (\frac{x_{ij}-\mu}{\sigma}) \right) \\ + \sum_{i=1}^{k} \sum_{j=m_i+1}^{n_i} \log \left(\Phi\left(\frac{DL_{ij}-\mu}{\sigma}\right) \right) \end{aligned} \tag{3.2}$$

For convenience, define $h = \frac{m_c}{n}$, $h_i = \frac{m_{c_i}}{n}$, and $\frac{m_{c_i}}{m} = \frac{h_i}{1-h}$, for $i = 1, 2, ..., k$. For the pooled sample let

$$\bar{x}_m = \frac{1}{m} \sum_{i=1}^{k} \sum_{j=1}^{m} x_{ij} \; , \quad and \quad s_m^2 = \frac{1}{m} \sum_{i=1}^{k} \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_m)^2$$

be the sample mean and sample variance of the $m = \sum_{i=1}^{k} m_i$ non-censored observations, respectively.

The maximum likelihood estimates $\hat{\mu}$ and $\hat{\sigma}$ of μ and σ are the solutions to equations (3.3) and (3.4), the partial derivatives for the log-likelihood equation with respect to μ and σ:

$$\frac{\partial \ell_{H_{0N}}(\mu, \sigma)}{\partial \mu} = \sum_{i=1}^{k} \sum_{j=1}^{m_i} (\frac{x_{ij}-\mu}{\sigma}) - \sum_{i=1}^{k} \sum_{j=m_i+1}^{n_i} W(\xi_{ij}) = 0 \tag{3.3}$$

$$\frac{\partial \ell_{H_{0N}}(\mu, \sigma)}{\partial \sigma} = -m + \sum_{i=1}^{k} \sum_{j=1}^{m_i} (\frac{x_{ij}-\mu}{\sigma})^2 - \sum_{i=1}^{k} \sum_{j=m_i+1}^{n_i} W(\xi_{ij}) \; \xi_{ij} = 0 \tag{3.4}$$

where $m = \sum_{i=1}^{k} m_i$, $W(\xi_{ij}) = \frac{\phi(\xi_{ij})}{\Phi(\xi_{ij})}$ and $\xi_{ij} = \frac{DL_{ij}-\mu}{\sigma}$.

The expectation maximization (EM) algorithm will be used iteratively to obtain the solutions $\hat{\mu}$ and $\hat{\sigma}$ to the maximum likelihood equations (3.3) and (3.4). The EM algorithm was proposed by Dempster et. al. (1977) for calculating the maximum

likelihood estimated from censored samples. The procedure consists of alternately estimating the censored observations from the current parameter estimates and estimating the parameters from the actual and estimated observations. The EM algorithm can be used to calculate the maximum likelihood estimates for the mean μ and standard deviation σ of a normal distribution from both singly- and multiply-censored samples. A brief description for the EM algorithm is given here.

At step 0 of the EM algorithm all non-censored observations are used to calculate the initial estimates of μ and σ as follows:

$$\hat{\mu}_0 = \bar{x}_m = \frac{1}{m}\sum_{i=1}^{k}\sum_{j=1}^{m} x_{ij} \ , \quad \text{and} \quad \hat{\sigma}_0^2 = s_m^2 = \frac{1}{m}\sum_{i=1}^{k}\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_m)^2$$

Let $\hat{\mu}_s$ and $\hat{\sigma}_s$ be the maximum likelihood estimates of μ and σ at step s of this procedure. At step $s+1$, each censored observation x_{ij} (where $i = 1, 2, ..., k$; $j = m_i + 1, 2, ..., n_i$) is replaced by an estimate of $\hat{\mu}_s - \hat{\sigma}_s W(\frac{DL_{ij} - \hat{\mu}_s}{\hat{\sigma}_s})$.

Let the values u_{ij} be calculated at step $s+1$ as follows:

$$u_{ij} = \begin{cases} x_{ij} \ , & \text{for } \ i = 1, 2, ..., k \text{ and } j = 1, ..., m_i \\ \\ \hat{\mu}_s - \hat{\sigma}_s W(\frac{DL_{ij} - \hat{\mu}_s}{\hat{\sigma}_s}) & \text{for } i = 1, 2, ..., k \text{ and } j = m_i + 1, ..., n_i \end{cases}$$

So the updated estimates $\hat{\mu}_{s+1}$ and $\hat{\sigma}_{s+1}$ of μ and σ are given by

$$\hat{\mu}_{s+1} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m_i} u_{ij} + \sum_{i=1}^{k}\sum_{j=m_i+1}^{n_i} u_{ij}}{n}$$

and

$$\hat{\sigma}^2_{s+1} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m_i}(u_{ij} - \hat{\mu}_{s+1})^2 + \sum_{i=1}^{k}\sum_{j=m_i+1}^{n_i}(u_{ij} - \hat{\mu}_{s+1})^2}{\sum_{i=1}^{k} m_i + \sum_{i=1}^{k}\sum_{j=m_i+1}^{n_i}\gamma(\frac{u_{ij} - \hat{\mu}_s}{\hat{\sigma}_s})}$$

where the function $\gamma(t)$ is defined as:

$$\gamma(t) = W(t)(W(t) + t) \quad \text{and} \quad W(t) = \frac{\phi(t)}{\Phi(t)}$$

More details about the EM algorithm procedure can be found in Wolynetz (1979). Convergence is achieved if both $|\hat{\mu}_s - \hat{\mu}_{s+1}| < 0.00001$ and $|\hat{\sigma}_s - \hat{\sigma}_{s+1}| < 0.00001$ occur. When these convergence criteria are met, the maximum likelihood estimates for μ and σ are then given by $\hat{\mu} = \hat{\mu}_s$ and $\hat{\sigma} = \hat{\sigma}_s$, respectively.

3.2 Maximum Likelihood Estimates under H_{A1N}

Under the hypothesis H_{A1N} x_{ij} are assumed to be normally distributed with mean μ_i and standard deviation σ_i, for $i = 1, 2, ..., k$ and $j = 1, 2, ..., n_i$. Thus the likelihood function under H_{A1N} is given by:

$$L_{H_{A1N}}(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k) = \prod_{i=1}^{k} \left((\frac{1}{\sqrt{2\pi}})^{m_i} (\sigma_i)^{-m_i} \; e^{-\frac{1}{2}\sum_{j=1}^{m_i} (\frac{x_{ij}-\mu_i}{\sigma_i})^2} \right) \cdot \prod_{i=1}^{k} \left(\prod_{j=m_i+1}^{n_i} \Phi\left(\frac{DL_{ij} - \mu_i}{\sigma_i} \right) \right) \tag{3.5}$$

Hence, the corresponding log-likelihood function $\ell_{H_{A1N}}(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k)$ of (3.5) which is defined as $\log(L_{H_{A1N}}(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k))$ is given by:

$$\ell_{H_{A1N}}(\mu_1, \mu_2, ..., \mu_k; \sigma_1, \sigma_2, ..., \sigma_k) = \sum_{i=1}^{k} -\frac{m_i}{2} \log(2\pi) + \sum_{i=1}^{k} -m_i \log \sigma_i - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{m_i} \left(\frac{x_{ij} - \mu_i}{\sigma_i} \right)^2 + \sum_{i=1}^{k} \sum_{j=m_i+1}^{n_i} \log \Phi \left(\frac{DL_{ij} - \mu_i}{\sigma_i} \right) \tag{3.6}$$

The maximum likelihood estimates $\hat{\mu}_i$ and $\hat{\sigma}_i$ of μ_i and σ_i are the solutions to equations (3.7) and (3.8) for $i = 1, 2, ..., k$.

$$\frac{\partial \ell_{H_{A1N}}(\mu_i, \sigma_i)}{\partial \mu_i} = \sum_{j=1}^{m_i} \left(\frac{x_{ij} - \mu_i}{\sigma_i} \right) - \sum_{j=m_i+1}^{n_i} W(\xi_{ij}) = 0 \tag{3.7}$$

$$\frac{\partial \ell_{H_{A1N}}(\mu_i, \sigma_i)}{\partial \sigma_i} = \sum_{j=1}^{m_i} \left(\frac{x_{ij} - \mu_i}{\sigma_i} \right)^2 - m_i - \sum_{j=m_i+1}^{n_i} W(\xi_{ij}) \, \xi_{ij} = 0 \tag{3.8}$$

where $W(\xi_{ij}) = \frac{\phi(DL_{ij})}{\Phi(DLi_{ij})}$ and $\xi_{ij} = \frac{DL_{ij} - \mu_i}{\sigma_i}$ for $i = 1, 2, ..., k$.

The single sample EM algorithm estimation method can be used to obtain the maximum likelihood estimates $\hat{\mu}_i$ and $\hat{\sigma}_i$ of μ_i and σ_i for $i = 1, 2, ..., k$ as follows. At step 0 of the EM algorithm all non-censored observations are used to calculate the initial estimates of μ_i and σ_i for $i = 1, 2, ..., k$ as follows:

$$\hat{\mu}_{i0} = \bar{x}_{m_i} = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij} \; , \quad and \quad \hat{\sigma}^2_{i0} = s^2_{m_i} = \frac{1}{m_i} \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{m_i})^2$$

For $i = 1, 2, ..., k$ let $\hat{\mu}_{is}$ and $\hat{\sigma}_{is}$ be the maximum likelihood estimates of μ_i and σ_i at step s of this procedure. At step $s + 1$, each censored observation x_{ij} (where $i =$

$1, 2, ..., k;$ and $j = m_i + 1, ..., n_i)$ is replaced by an estimate of $\hat{\mu}_{is} - \hat{\sigma}_{is} W(\frac{DL_{ij} - \hat{\mu}_{is}}{\hat{\sigma}_{is}})$.

Let the values t_{ij} be calculated at step $s+1$ as follows:

$$t_{ij} = \begin{cases} x_{ij} \ , & \text{for } \ i = 1, 2, ..., k \text{ and } j = 1, ..., m_i \\ \\ \hat{\mu}_{is} - \hat{\sigma}_{is} W(\frac{DL_{ij} - \hat{\mu}_{is}}{\hat{\sigma}_{is}}) & \text{for } i = 1, 2, ..., k \text{ and } j = m_i + 1, ..., n_i \end{cases}$$

So for $i = 1, 2, ..., k$ the updated estimates $\hat{\mu}_{is+1}$ and $\hat{\sigma}_{is+1}$ of μ_i and σ_i are given by

$$\hat{\mu}_{is+1} = \frac{\sum_{j=1}^{m_i} t_{ij} + \sum_{j=m_i+1}^{n_i} t_{ij}}{n_i}$$

and

$$\hat{\sigma}^2_{is+1} = \frac{\sum_{j=1}^{m_i} (t_{ij} - \hat{\mu}_{is+1})^2 + \sum_{j=m_i+1}^{n_i} (t_{ij} - \hat{\mu}_{is+1})^2}{m_i + \sum_{j=m_i+1}^{n_i} \gamma(\frac{t_{ij} - \hat{\mu}_{is}}{\hat{\sigma}_{is}})}$$

where the function $\gamma(v)$ is defined as:

$$\gamma(v) = W(v)(W(v) + v) \quad \text{and} \quad W(v) = \frac{\phi(v)}{\Phi(v)}$$

For $i = 1, 2, ..., k$, convergence is achieved if $|\hat{\mu}_{is} - \hat{\mu}_{is+1}| < 0.00001$, and $|\hat{\sigma}_{is} - \hat{\sigma}_{is+1}| < 0.00001$ occur. When these convergence criteria are met, the maximum likelihood estimates for μ_i and σ_i are then given by $\hat{\mu}_i = \hat{\mu}_{is}$ and $\hat{\sigma}_i = \hat{\sigma}_{is}$, respectively.

3.3 Maximum Likelihood Estimates under H_{A2N}

Under the hypothesis H_{A2N} x_{ij} are assumed to be normally distributed with mean μ_i and standard deviation σ, for $i = 1, 2, ..., k$ and $j = 1, 2, ..., n_i$. Thus the likelihood function under H_{A2N} is given by:

$$\begin{aligned} L_{H_{A2N}}(\mu_1, \mu_2, ..., \mu_k; \sigma) = & \prod_{i=1}^{k} \left((\frac{1}{\sqrt{2\pi}})^{m_i} (\sigma)^{-m_i} \ e^{-\frac{1}{2} \sum_{j=1}^{m_i} (\frac{x_{ij} - \mu_i}{\sigma})^2} \right) \\ & \prod_{i=1}^{k} \left(\prod_{j=m_i+1}^{n_i} \Phi \left(\frac{DL_{ij} - \mu_i}{\sigma} \right) \right) \end{aligned} \tag{3.9}$$

Hence, the corresponding log-likelihood function $\ell_{H_{A2N}}(\mu_1, \mu_2, ..., \mu_k; \sigma)$ of (3.9) which is defined as $\log(L_{H_{A2N}}(\mu_1, \mu_2, ..., \mu_k; \sigma))$ is given by:

$$\begin{aligned}\ell_{H_{A2N}}(\mu_1, \mu_2, ..., \mu_k; \sigma) = &\sum_{i=1}^{k} -\frac{m_i}{2}\log(2\pi) + \sum_{i=1}^{k} -m_i \log\sigma \\ &- \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(\frac{x_{ij}-\mu_i}{\sigma}\right)^2 + \sum_{i=1}^{k}\sum_{j=m_i+1}^{n_i}\log\Phi\left(\frac{DL_{ij}-\mu_i}{\sigma}\right)\end{aligned} \tag{3.10}$$

For $i = 1, 2, ..., k$ the maximum likelihood estimates $\hat{\mu}_i$ and $\hat{\sigma}$ of μ_i and σ are the solutions to equations (3.11)-(3.12).

$$\frac{\partial \ell_{H_{A2N}}(\mu_1, \mu_2, ..., \mu_k; \sigma)}{\partial \mu_i} = \sum_{j=1}^{m_i}\left(\frac{x_{ij}-\mu_i}{\sigma}\right) - \sum_{j=m_1+1}^{n_i} W(\xi_{ij}) = 0 \tag{3.11}$$

$$\frac{\partial \ell_{H_{A2N}}(\mu_1, \mu_2, ..., \mu_k; \sigma)}{\partial \sigma} = \sum_{1=1}^{k}\sum_{j=1}^{m_i}\left(\frac{x_{ij}-\mu_i}{\sigma}\right)^2 - m - \sum_{1=1}^{k}\sum_{j=m_i+1}^{n_i} W(\xi_{ij})\,\xi_{ij} = 0 \tag{3.12}$$

where $m = m_1 + m_2 + ... + m_k$, $W(\xi_{ij}) = \frac{\phi(DL_{ij})}{\Phi(DL_{ij})}$ and $\xi_{ij} = \frac{DL_{ij}-\mu_i}{\sigma}$ for $i = 1, 2, ..., k$ and $j = m_i + 1, 2, ..., n_i$.

The expectation maximization (EM) algorithm will be used iteratively to obtain the solutions $\hat{\mu}_i$, and $\hat{\sigma}$ to the maximum likelihood equations (3.11) and (3.12), for $i = 1, 2, ..., k$. At step 0 of the EM algorithm all non-censored observations are used to calculate the initial estimates of μ_i and σ as follows:

$$\hat{\mu}_{i0} = \bar{x}_{m_i} = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij} \qquad \text{for} i = 1, 2, ..., k,$$

and

$$\hat{\sigma}_0^2 = s_m^2 = \frac{1}{m}\sum_{i=1}^{k}\sum_{j=1}^{m_i}(x_{ij} - \hat{\mu}_0)^2$$

where

$$\hat{\mu}_0 = \frac{m_1\hat{\mu}_{10} + m_2\hat{\mu}_{20} + ... + m_k\hat{\mu}_{k0}}{m_1 + m_2 + ... + m_k}$$

Let $\hat{\mu}_{is}$ and $\hat{\sigma}_s$ be the maximum likelihood estimates of μ_i and σ at step s of this procedure for $i = 1, 2, ..., k$. At step $s + 1$, each censored observation x_{ij} (where $i = 1, 2, ..., k$; $j = m_i + 1, 2, ..., n_i$) is replaced by an estimate of $\hat{\mu}_{is} - \hat{\sigma}_s W(\frac{DL_{ij}-\hat{\mu}_{is}}{\hat{\sigma}_s})$.

Let the values v_{ij} be calculated at step $s+1$ as follows:

$$v_{ij} = \begin{cases} x_{ij}\ , & \text{for } \ i = 1, 2, ..., k \text{ and } j = 1, ..., m_i \\ \\ \hat{\mu}_{is} - \hat{\sigma}_s W(\frac{DL_{ij} - \hat{\mu}_{is}}{\hat{\sigma}_s}) & \text{for } i = 1, 2, ..., k \text{ and } j = m_i + 1, ..., n_i \end{cases}$$

So the updated estimates $\hat{\mu}_{is+1}$ and $\hat{\sigma}_{s+1}$ of μ_i and σ are given by

$$\hat{\mu}_{is+1} = \frac{\sum_{j=1}^{m_i} v_{ij}}{m_i} \qquad \text{for} i = 1, 2, ..., k,$$

and

$$\hat{\sigma}^2_{s+1} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m_i} (v_{ij} - \hat{\mu}_{is+1})^2}{\sum_{i=1}^{k} m_i + \sum_{i=1}^{k} \sum_{j=m_i+1}^{n_i} \gamma(\frac{v_{ij} - \hat{\mu}_{is}}{\hat{\sigma}_s})}$$

where the function $\gamma(t)$ is defined as:

$$\gamma(t) = W(t)(W(t) + t) \quad \text{and} \quad W(t) = \frac{\phi(t)}{\Phi(t)}$$

For $i = 1, 2, ..., k$ convergence is achieved if $|\hat{\mu}_{is} - \hat{\mu}_{is+1}| < 0.00001$ and $|\hat{\sigma}_s - \hat{\sigma}_{s+1}| < 0.00001$ occur. When these convergence criteria are met, the maximum likelihood estimates for μ_i and σ are then given by $\hat{\mu}_i = \hat{\mu}_{is}$ and $\hat{\sigma} = \hat{\sigma}_s$, respectively.

4 Asymptotic Chi-Square Test

The estimated log-likelihood functions $\hat{\ell}_{H_{0N}}$, $\hat{\ell}_{H_{A1N}}$ and $\hat{\ell}_{H_{A2N}}$ under the hypotheses H_{0N} (overall homogeneity), H_{A1N} (overall heterogeneity) and H_{A2N} (mean heterogeneity, variance homogeneity), respectively; are obtained by replacing population parameters by their maximum likelihood estimates. Therefore from (3.2), (3.6) and (3.10) we get:

$$\begin{aligned} \hat{\ell}_{H_{0N}}(\hat{\mu}, \hat{\sigma}) = & \sum_{i=1}^{k} \left(-m_i \log \hat{\sigma} + m_i \log(\frac{1}{\sqrt{2\pi}}) - \frac{1}{2} \sum_{j=1}^{m_i} (\frac{x_{ij} - \hat{\mu}}{\hat{\sigma}}) \right) \\ & + \sum_{i=1}^{k} \sum_{j=m_i+1}^{n_i} \log \left(\Phi \left(\frac{DL_{ij} - \hat{\mu}}{\hat{\sigma}} \right) \right) \ , \end{aligned} \tag{4.1}$$

$$\hat{\ell}_{H_{A1N}}(\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_k; \hat{\sigma}_1, \hat{\sigma}_2, ..., \hat{\sigma}_k) = \sum_{i=1}^{k} -\frac{m_i}{2}\log(2\pi) + \sum_{i=1}^{k} -m_i \log \hat{\sigma}_i - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(\frac{x_{ij} - \hat{\mu}_i}{\hat{\sigma}_i}\right)^2 + \sum_{i=1}^{k}\sum_{j=m_i+1}^{n_i} \log \Phi\left(\frac{DL_{ij} - \hat{\mu}_i}{\hat{\sigma}_i}\right), \tag{4.2}$$

and

$$\hat{\ell}_{H_{A2N}}(\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_k; \hat{\sigma}) = \sum_{i=1}^{k} -\frac{m_i}{2}\log(2\pi) + \sum_{i=1}^{k} -m_i \log \hat{\sigma} - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(\frac{x_{ij} - \hat{\mu}_i}{\hat{\sigma}}\right)^2 + \sum_{i=1}^{k}\sum_{j=m_i+1}^{n_i} \log \Phi\left(\frac{DL_{ij} - \hat{\mu}_i}{\hat{\sigma}}\right), \tag{4.3}$$

In general, the asymptotic $\alpha-$level chi-square test used to test the null hypothesis $H_0 : \theta = 0$ versus the alternative hypothesis $H_a : \theta \neq 0$ is defined by

$$\chi_0^2 = -2(\hat{\ell}_{H_0}(\hat{\theta}_0) - \hat{\ell}_{H_{A1N}}(\hat{\theta}_a)) > \chi^2_{(\alpha,df)},$$

where χ_0^2 has a chi-square distribution with degrees of freedom df, which is defined by the number of free parameters under the alternative hypothesis H_a minus the number of free parameters under the null hypothesis H_0, and $\chi^2_{(\alpha,df)}$ is the upper α-point value obtained from the chi-square table with degrees of freedom df.

The asymptotic $\alpha-$level chi-square tests used in both Test 1: H_{0N} versus H_{A1N}, overall homogeneity versus overall heterogeneity, and Test 2: H_{0N} versus H_{A2N}, overall homogeneity versus mean heterogeneity and variance homogeneity are now described.

Test 1: Overall Homogeneity versus Overall Heterogeneity

$$H_{0N}: \ \mu_1 = \mu_2 = ... = \mu_k = \mu \quad \textbf{and} \quad \sigma_1 = \sigma_2 = ... = \sigma_k = \sigma$$

versus

$$H_{A1N}: \ \mu_1 \neq \mu_2 \neq ... \neq \mu_k \quad \textbf{and} \quad \sigma_1 \neq \sigma_2 \neq ... \neq \sigma_k$$

The asymptotic $\alpha-$level chi-square test to test the null hypothesis H_{0N} versus the alternative hypotheses H_{A1N} is defined by:

$$\chi^2_{0A1} = -2(\hat{\ell}_{H_{0N}} - \hat{\ell}_{H_{A1N}}) > \chi^2_{[\alpha,2(k-1)]} \tag{4.4}$$

where $\chi^2_{(\alpha,2)}$ is the upper $\alpha-$point for a chi-square random variable with 2 degrees of

freedom. The p-value of this test statistic is defined by:

$$p - value = P(\chi^2_{[2(k-)]} > \chi^2_{0A1}). \tag{4.5}$$

Thus in this case the null hypothesis will be rejected if $\chi^2_{0A1} > \chi^2_{[\alpha,2(k-1)]}$ or equivalently if $p - value < \alpha$.

Test 2: Overall Homogeneity versus Mean Heterogeneity and Variance Homogeneity

$$H_{0N}: \ \mu_1 = \mu_2 = ... = \mu_k = \mu \quad \textbf{and} \quad \sigma_1 = \sigma_2 = ... = \sigma_k = \sigma$$

versus

$$H_{A2N}: \ \mu_1 \neq \mu_2 \neq ... \neq \mu_k \quad \text{and} \quad \sigma_1 = \sigma_2 = ... = \sigma_k = \sigma$$

The asymptotic $\alpha-$level chi-square test to test the null hypothesis H_{0N} versus the alternative hypotheses H_{A12N} is defined by:

$$\chi^2_{0A2} = -2(\hat{\ell}_{H_{0N}} - \hat{\ell}_{H_{A2N}}) > \chi^2_{[\alpha,k-1]} \tag{4.6}$$

where $\chi^2_{[\alpha,k-1]}$ is the upper $\alpha-$point for a chi-square random variable with $k-1$ degrees of freedom. The p-value of this test statistic is defined by:

$$p - value = P(\chi^2_{[k-1]} > \chi^2_{0A2}). \tag{4.7}$$

Thus in this case the null hypothesis will be rejected if $\chi^2_{0A2} > \chi^2_{[\alpha,k-1]}$ or equivalently if $p - value < \alpha$.

Computer Programs: To facilitate the application of the test procedure and parameter estimation method described in this article, a computer program called "*K.Lognormal.Estimation*" is written in the R language to automate parameters estimation from multiply left-censored data sets that are normally or log-normally distributed and to obtain the estimated values of the log-likelihood functions under the null and the alternative hypotheses . In addition, this computer program will be used to obtain the asymptotic $\alpha-$level chi-square test statistic and its p-value. A Copy of the source code is given in the Appendix section and is available upon request.

For the sake of simplicity, in the remaining part of this article Test 1 (H_{0N} versus H_{A1N}) will be considered. Test 2 can be easily programmed and computed.

5 Example:

The following data sets are simulated from a lognormal distribution with mean $\mu = 3$ and standard deviation $\sigma = 1$. Data are given in Table 1. Each data set is

Table 1: Three Simulated Data Sets of size $n = 50$ from a Lognormal Distribution with Mean $\mu = 3$ and standard deviation $\sigma = 1$

Data 1	19	6	1	24	17	76	16	22	19	73	43	30	14	3	34	82	13
	10	13	11	1	36	3	73	3	69	26	29	13	42	16	31	46	19
	23	4	57	6	36	21	18	15	9	15	88	72	5	102	13	15	
Data 2	115	27	8	56	133	65	28	45	30	4	7	3	46	3	21	12	77
	18	15	10	5	57	10	7	56	114	21	24	9	25	36	140	12	7
	16	19	36	6	45	97	124	32	21	30	64	5	14	19	17	47	
Data 3	23	23	18	33	21	94	80	7	77	5	4	40	102	114	43	8	4
	14	25	42	11	9	59	10	122	25	24	58	8	7	59	20	6	8
	46	33	37	45	18	56	33	30	15	2	12	12	37	44	112	87	

artificially censored at the $10th$, $20th$ and $30th$ quantiles. Table 2 contains the censored data sets and censored indicators ($0 =$ noncensored, $1 =$ censored). The first data set contains three distinct detection limits 4, 10 and 13, and has censoring level of 24%. The second data set contains three distinct detection limits 6, 9 and 13, and has censoring level of 30%. The third data set contains three distinct detection limits 7, 9 and 13, and has censoring level of 30%. Accordingly, the pooled data contains six distinct detection limits 4, 6, 7, 9, 10 and 13, and has censoring level of 28%.

Table 3 contains estimates of the normal and log-normal population parameters. The p-value results associated with the application of the recommended asymptotic chi-square test to the simulated censored data presented are also included in Table 3. The p-value of testing the null hypothesis $H_{0N} : \mu_1 = \mu_2 = \mu_3$ and $\sigma_1 = \sigma_2 = \sigma_3$ versus the alternative hypothesis $H_{A1N} : \mu_1 \neq \mu_2 \neq \mu_3$ and $\sigma_1 \neq \sigma_2 \neq \sigma_3$ is 0.8834. Therefore the hypothesis of equal normal (lognormal) parameters is accepted at significance level of $\alpha = 0.05$.

6 Simulation Study

In this simulation study, type I error rates and power of the test procedure introduced in this article are investigated. A computer program was written in the R language for this purpose. For each combination of the population parameters μ_1 , μ_2, μ_3, σ_1, σ_2 and σ_3 described below, two sample size cases were considered: in case one, $n_1 = n_2 = n_3 = 25$ and in the second case, $n_1 = n_2 = n_3 = 75$. The first case will be referred to as the small sample size case and the second as the large sample size case. Censoring at three different detection limits was used in each simulated sample. The simulation study was performed with 10,000 repetitions ($N = 10,000$) of sample normal distributions for each combinations of n, μ_1 , μ_2, μ_3, σ_1, σ_2, σ_3, and censoring levels. Simulated data were artificially censored twice at the 10th, 20th, 30th,and at 30th, 40th, and 50th percentiles as shown in Tables 4 and 5. In order to check the Type I error, the population parameters were specified as $\mu_1 = \mu_2 = \mu_3 = 0$, and

Table 2: The Three Simulated Data Sets Artificially Censored at $10th$, $20th$ and $30th$ quantiles

Data Set 1				Data Set 2				Data Set 3			
y_1	cenc1	y_1	cenc1	y_2	cenc2	y_2	cenc2	y_3	cenc3	y_3	cenc3
19	0	10	1	115	0	27	0	23	0	23	0
4	1	24	0	9	1	56	0	18	0	33	0
17	0	76	0	133	0	65	0	21	0	94	0
16	0	22	0	28	0	45	0	80	0	9	1
19	0	73	0	30	0	6	1	77	0	7	1
43	0	30	0	9	1	6	10	7	1	40	0
14	0	4	1	46	0	6	1	102	0	114	0
34	0	82	0	21	0	13	1	43	0	9	1
13	0	13	1	77	0	18	0	7	1	14	0
13	0	13	1	15	0	13	1	25	0	42	0
4	1	36	0	6	1	57	0	13	1	13	1
4	1	73	0	13	1	9	1	59	0	13	1
4	1	69	0	56	0	114	0	122	0	25	0
26	0	29	0	21	0	24	0	24	0	58	01
13	0	42	0	13	1	25	0	9	1	9	1
16	0	31	0	36	0	140	0	59	0	20	0
46	0	19	0	13	1	9	1	7	1	9	1
23	0	10	1	16	0	19	0	46	0	33	0
57	0	10	1	36	0	9	1	37	0	45	0
36	0	21	0	45	0	97	0	18	0	56	0
18	0	15	0	124	0	32	0	33	0	30	0
10	1	15	0	21	0	30	0	15	0	7	1
88	0	72	0	64	0	6	1	13	1	13	1
10	1	102	0	14	0	19	0	37	0	44	0
13	0	15	0	17	0	47	0	112	0	87	0

$\sigma_1 = \sigma_2 = \sigma_3 = 1$ as shown in Table 4. In order to check the power, the population parameters were specified as $\mu_1 = -1.0(0.1) - 0.1$, $\mu_2 = 0$, $\mu_3 = 0.1(0.1)1.0$, $\sigma_1 = 1$, $\sigma_2 = 1.1(0.1)2.0$, and $\sigma_3 = 1.2(0.2)3.0$ as shown in Table 5.

The following observations and conclusions are made from an examination of the simulation results reported in Tables 4 and 5.

From Table 4, one can see that the estimated simulated Type I error rates are slightly higher than 0.05 $(0.0534, 0.0516)$ for the small sample size case, and slightly less than 0.05 $(0.0482, 0.0469)$ for the large sample size case. The censoring levels do not seem to affect the value of Type I error rate, α.

From Table 5, one can see that the estimated simulated power is higher for large sample size case than the small sample size case, and slightly higher for the lower

Table 3: Estimates of normal and log-normal parameter values from the simulated data given in Table 2

H_{0N}	H_{A1N}		
Pooled Data	Data 1	Data 2	Data 3
Estimations of Normal Parameters			
$\hat{\mu}_0 = 2.9805$ $\hat{\sigma}_0 = 1.1516$	$\hat{\mu}_1 = 2.8857$ $\hat{\sigma}_1 = 1.0733$	$\hat{\mu}_2 = 2.9981$ $\hat{\sigma}_2 = 1.2131$	$\hat{\mu}_3 = 3.0585$ $\hat{\sigma}_3 = 1.1559$
Estimations of Lognormal Parameters			
$\hat{\mu}_{y1} = 31.8707$ $\hat{M}_{y1} = 17.9161$ $\hat{\sigma}_{y1} = 46.8881$ $\hat{S}_{y1} = 7.5979$	$\hat{\mu}_{y2} = 41.8422$ $\hat{M}_{y2} = 20.0474$ $\hat{\sigma}_{y2} = 76.6552$ $\hat{S}_{y2} = 11.6447$	$\hat{\mu}_{y3} = 41.5357$ $\hat{M}_{y3} = 21.2956$ $\hat{\sigma}_{y3} = 69.5545$ $\hat{S}_{y3} = 9.7196$	$\hat{\mu}_{y0} = 38.2289$ $\hat{M}_{y0} = 19.6977$ $\hat{\sigma}_{y0} = 63.5869$ $\hat{S}_{y0} = 9.5918$
The Asymptotic Chi-square Test: χ_0^2 (P-value)			
1.1678 (0.8834)			

level of censoring. Specifically, in the small sample size case with $h_1 = 10\%, h_2 = 20\%\text{and}h_3 = 30\%$; and $(h_1 = 30\%, h_2 = 40\%\text{and}h_3 = 50\%)$ censoring levels we reach a power of 0.9998 (0.9981) for the values of $\mu_1 = -1.0$, $\mu_2 = 0.0$ and $\mu_3 = 1.0$, and the values of $\sigma_1 = 1$, $\sigma_2 = 2$ and $\sigma_3 = 3.0$. Alternatively, in the large sample size case with censoring levels $h_1 = 10\%, h_2 = 20\%\text{and}h_3 = 30\%$ we reach a power above 0.99 for the values of $\mu_1 = -0.4$, $\mu_2 = 0.0$ and $\mu_3 = 0.4$, and the values of $\sigma_1 = 1.0$, $\sigma_2 = 1.4$ and $\sigma_3 = 1.8$ and a power of 1.0 for the values of $\mu_1 = -0.6$, $\mu_2 = 0.0$ and $\mu_3 = 0.6$, and the values of $\sigma_1 = 1.0$, $\sigma_2 = 1.6$ and $\sigma_3 = 2.2$; while with censoring levels $h_1 = 30\%, h_2 = 40\%\text{and}h_3 = 50\%$ we reach a power above 0.99 for the values of $\mu_1 = -0.5$, $\mu_2 = 0.0$ and $\mu_3 = 0.5$, and the values of $\sigma_1 = 1.0$, $\sigma_2 = 1.5$ and $\sigma_3 = 2.0$ and a power of 1.0 for the values of $\mu_1 = -0.7$, $\mu_2 = 0.0$ and $\mu_3 = 0.7$, and the values of $\sigma_1 = 1.0$, $\sigma_2 = 1.7$ and $\sigma_3 = 2.4$.

In summary, the test procedure introduced in this article maintains its stated significance level and has much power with larger sample size and a bit less power with greater censoring levels. In addition, the power decreases when the censoring levels moves from $0.10, 0.20, \text{and}0.30$ to $0.30, 0.40\text{and}0.50$. Also, the power increases greatly when the sample size moves from the order of 25 to the order of 75.

Table 4: The Estimated Simulated Type I Error Rates: $\mu_1 = \mu_2 = \mu_3 = 0$ and $\sigma_1 = \sigma_2 = \sigma_3 = 1$

Sample Size	Censoring Level	Estimated α
Small ($n = 25$)	10% , 20% , 30%	0.0534
Small ($n = 25$)	30% , 40% , 50%	0.0516
Large ($n = 75$)	10% , 20% , 30%	0.0482
Large ($n = 75$)	30% , 40% , 50%	0.0469

7 Conclusions and Remarks

The k-sample lognormal model provides an alternative to the nonparameteric models for testing the equality of the parameters of $k(k \geq 3)$ independent log-normal populations in environmental settings. The lognormal model provide additional information as to the overall homogeneity ($\sigma_1 = \sigma_2 = ... = \sigma_k$) or heterogeneity ($\sigma_1 \neq \sigma_2 \neq ... \neq \sigma_k$) of the k lognormal populations, which is important in the interpretation of the differences among medians. It is well known that the log-normal distribution is widely used in modelling environmental and biomedical censored data. This article has dealt with the problem of comparing the parameters of $k(k \geq 3)$ independent log-normal populations in the presence of left-censored data. The EM Algorithm is employed to obtain the maximum likelihood estimates of population parameters under different hypotheses. A parametric test procedure for testing the equality of $k(k \geq 3)$ independent log-normal parameters in the presence of censored data is presented and evaluated. The performance of the test procedure presented in this article is evaluated by means of simulation studies. We find analytically that the considered test procedure is doing well through comparing the size and power statistic test. To facilitate the application of the new test procedure a computer program is written in the R languages. I hope that my paper would be useful to the researchers who are considering log-normal distribution in their analysis of the left censored data.

Table 5: The Estimated Simulated Power Rates

	Small Sample Size ($n = 25$)		Large Sample Size ($n = 75$)	
$(\mu_1, \mu_2, \mu_3; \sigma_1, \sigma_2, \sigma_3)$	Censoring Levels (10% , 20% , 30%)	Censoring Levels (30% , 40% , 50%)	Censoring Levels (10% , 20% , 30%)	Censoring Levels (30% , 40% , 50%)
(-0.1, 0, 0.1 ; 1, 1.1, 1.2)	0.0165	0.0247	0.1213	0.0647
(-0.2, 0, 0.2 ; 1, 1.2, 1.4)	0.1183	0.0935	0.6074	0.4395
(-0.3, 0, 0.3 ; 1, 1.3, 1.6)	0.3168	0.2665	0.9448	0.8634
(-0.4, 0, 0.4 ; 1, 1.4, 1.8)	0.5437	0.4367	0.9975	0.9897
(-0.5, 0, 0.5 ; 1, 1.5, 2.0)	0.7731	0.6528	0.9995	0.9986
(-0.6, 0, 0.6 ; 1, 1.6, 2.2)	0.8926	0.7846	1.0000	0.9993
(-0.7, 0, 0.7 ; 1, 1.7, 2.4)	0.9573	0.9085	1.0000	1.0000
(-0.8, 0, 0.8 ; 1, 1.8, 2.6)	0.9804	0.9541	1.0000	1.0000
(-0.9, 0, 0.9 ; 1, 1.9, 2.8)	0.9989	0.9872	1.0000	1.0000
(-1.0, 0, 1.0 ; 1, 2.0, 3.0)	0.9998	0.9981	1.0000	1.0000

References

Abdollahnezhad K., Babanezhad M. and Jafari A.A. (2012). *Inference on Difference of Means of two Log-Normal Distributions; A Generalized Approach*, Journal of Statistical and Econometric Methods 1 (2): 125-131.

Aboueissa A. A. (2015). *Comparison of Two Means of Two Log-Normal Distributions When Data is Singly Censored*, International Journal of Statistics and Probability 4 (2): 73-86.

Cohen A. C. R. (1959). *Simplified Estimators For The Normal Distribution When Samples Are Singly Censored Or Truncated*, Technometrics 3: 217-237.

Dempster A. P., N. Laird M. and Rubin D. B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*, The Journal Of Royal Statistical Society B 39: 1-38.

El-Shaarawi, A. H. (1989). *Inferences about the Mean from Censored Water Quality Data*, Water Resources Research 25: 685-690.

El-Shaarawi A. H. and Dolan D. M. (1989). *Maximum Likelihood Estimation Of Water Concentrations From Censored Data*, Canadian Journal Of Fisheries And Aquatic Sciences 46: 1033-1039.

El-Shaarawi A. H. and Esterby S. R. (1992). *Replacement Of Censored Observations By A Constant: An Evaluation*, Water Research 26(6): 835-844.

Gibbons RD. (1994). *Statistical Methods For Groundwater Monitoring*, John Wiley&Sons, New York.

Gilbert Richard O. (1987). *Statistical Methods For Environmental Pollution Monitoring*, Van Nostrand Reinhold: New York.

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

Gupta R. C. and Li X. (2006). *Statistical Inference for the Common Mean of two Log-normal Distributions and some Applications in Reliability*, Computational Statistics and Data Analysis 50: 3141-3164.

Harris G. A. (1991). *Two-samples Comparisons in the Presence of Less-than-detectable data*, Proceeding of the Section on Statistics and the Environment: American Statistical Association: 197-201.

Jianrong W., Jiang G., Wong A., and Xiang S. (2002). *Likelihood Analysis for the Ratio of Means of Two Independent Log-Normal Distributions*, Biometrics 58: 463-469.

Krishnamoorthy K., Mathew T. and Xu Z. (2014). *Comparison of Means of Two Lognormal Distributions Based on Samples with Multiple Detection Limits*, Journal of Occupational and Environmental Hygiene 11 (8): 538-546.

Krishnamoorthy K., Mathew T. and Xu Z. (2014). *Standardized Likelihood Inference for the Mean and Percentiles of a Lognormal Distribution Based on Samples with Multiple Detection Limits*, Journal of Environmental Statistics 6 (5): 1-18.

Krishnamoorthy K. and Mathew T. (2003). *Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals*, Journal of Statistical Planning and Inference 115: 103-121.

Krishnamoorthy K., Mathew T. and Ramachandran, G. (2006). *Generalized P-Values and Confidence Intervals: A Novel Approach for Analyzing Lognormally Distributed Exposure Data*, Journal of Occupational and Environmental Hygiene 3: 642-650.

Krishnamoorthy K., Avishek M. and Mathew T. (2011). *Inference for the Lognormal Mean and Quantiles Based on Samples with Left and Right Type I Censoring*, Technometrics 53 (1): 72-83.

Marco Bee (2005). *On Maximum Likelihood Estimation of Operational Loss Distributions*: Universita Degli, Studi Di Trento, Discussion paper No. 3

Millard S. P. and S. J. Deverel (1998). *Nonparametric Statistical Methods for Comparing two Sites Based on Data with multiple Nondetect Limits*, Water Resources Research 24: 2087 - 2098.

Prentice R. L. (1978). *Linear rank Tests with Right Censored Data*, Biometrika 65: 167-179.

Paul H. and Gary H. G.(2007). *A Comparison of Several Methods for Analyzing Censored Data*, Oxford University Press on behalf of the British Occupational Hygiene Society 51 (7): 611-632.

Schneider H. (1986). *Truncated and Censored Samples from Normal Population*, Marcel Dekker: New York.

Shumway R. H., Azari A. S. and Johnson P. (1989). *Estimating mean concentrations under transformation for environmental data with detection limits*, Technometrics 31: 347356.

Stoline Michael R. (1993). *Comparison of Two Medians Using a Two-Sample Log-normal Model in Environmental Contexts*, Environmetrics 4(3): 323-339.

Stavros Pouloukas (2004). *Estimation and comparison of Lognormal Parameters in the Presence of Censored Data*, Journal of Statistical Computation & Simulation 74(3): 157-169.

Wolynetz M. S. (1979). *Maximum Likelihood Estimations from Confined and Censored Normal Data*, Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (2): 185 - 195.

Yan Jin, Misty J. Hein, James A. Deddens and Cynthia J. Hines (2011). *Analysis of Lognormally Distributed Exposure Data with Repeated Measures and Values below the Limit of Detection Using SAS*, Oxford University Press, British Occupational Hygiene Society 55 (1): 97-112.

Zhou X., Sujuan G. and Hui S. L. (1997). *Methods for Comparing the Means of Two Independent Log-normal Samples*, Biometrics 53: 1129-1135.

Appendix

Computer Programs

The following computer program, "$K.Lognormal.Estimation$", is written in the R language to automate parameters estimation from left-censored data sets that are normally or log-normally distributed and to obtain the estimated values of the log-likelihood functions under the hypotheses H_{0N} and H_{A1N}. In addition, this computer program will be used to obtain the asymptotic $\alpha-$level chi-square test statistic and its p-value.

footnotesize

```
K.Lognormal.Estimation<-function(data1, data2, data3, kk, NumI,
LogN1, LogN2, LogN3) {
     #
     # NumI is the number of iterations suggested by users.
     # data1 and data2 are matrices containing of two columns each
     # the first column is the data set and the second column
     # is indicator 0 for uncensored and 1 for censored observations.
     # LogN = T if the data are log-normally distributed
     # kk is the number of lognormal populations.
    n1<-length(data1[,1])
```

```
    n2<-length(data2[,1])
    n3<-length(data3[,1])
    table1 <- table(data1[data1[, 2]==1, 1])
    DLV1<-as.numeric(dimnames(table1)[[1]])
    mcV1<-as.vector(table1)
    table2 <- table(data2[data2[, 2]==1, 1])
    DLV2<-as.numeric(dimnames(table2)[[1]])
    mcV2<-as.vector(table2)
    table3 <- table(data3[data3[, 2]==1, 1])
    DLV3<-as.numeric(dimnames(table3)[[1]])
    mcV3<-as.vector(table3)

    if(LogN1==T) data1[,1]<-log(data1[,1]) else data1[,1]<-data1[,1]
    if(LogN2==T) data2[,1]<-log(data2[,1]) else data2[,1]<-data2[,1]
    if(LogN3==T) data3[,1]<-log(data3[,1]) else data3[,1]<-data3[,1]

    datacomb<-rbind(data1,data2,data3)
    n<-length(datacomb[,1])
    table <- table(datacomb[datacomb[, 2]==1, 1])
    DLV<-as.numeric(dimnames(table)[[1]])
    mcV<-as.vector(table)
    k<-length(mcV)

####################   EM Algorithm    ####################

EMAmultvect<-function(data, NumI) {
     #
     # N is the number of iterations suggested by users.
     # data is a matrix containing of two columns
     # the first column is the data set and the second column
     # is indicator 0 for uncensored and 1 for censored obs.
     #
    n<-length(data[,1])
    table <- table(data[data[, 2]==1, 1])
    DLV<-as.numeric(dimnames(table)[[1]])

    mcV<-as.vector(table)

    Xmbar<-tapply(data[,1],list(data[,2]),mean)["0"]
    Smsquare<-tapply(data[,1],list(data[,2]),var)["0"]
    g<-Smsquare/(Xmbar-(sum(DLV)/2))^2
    n<-length(data[,1])
    m<-sum(data[,2]==0)
     #  print(m)
    k<-length(DLV)
    mc<-numeric(k) # vector of number of censored obs. at each DL.
    mc<-numeric(k)
    u<-numeric(n)
        for(r in 1:k) {
    for(i in 1:n) {
    if(data[i,1]==DLV[r] && data[i,2]==1)
    u[i]<-1
    else
    u[i]<-0
    }
```

```
mc[r]<-sum(u)
}
 #
 #   End of this part
 #
mu0.hat<-Xmbar
sig0.hat<-Smsquare

muhat<-numeric(NumI)
sighat<-numeric(NumI)
#
    w<-matrix(0,n,2)
ww<-matrix(0,n,2)
w[,2]<-data[,2]
ww[,2]<-data[,2]
#
    for(i in 1:n) {   #b11
    if(data[i,2]==1) {   #a22
    z0<-(data[i,1]-mu0.hat)/sqrt(sig0.hat)
    d0<-dnorm(z0)
    p0<-pnorm(z0)
      wdp0<-d0/p0
      w[i,1]<-mu0.hat-(sqrt((sig0.hat))*wdp0)
      ww[i,1]<-(wdp0)*(wdp0+z0)
            }   #a22
       else {  #e1
    w[i,1]<-data[i,1]
    ww[i,1]<-data[i,1]
    }   #e1
    muhat[1]<-mean(w[,1])
    num0<-sum((w[,1]-muhat[1])^2)
    dnum1<-tapply(ww[,1],list(ww[,2]),sum)["1"]
    dnum0<-m+dnum1
    sighat[1]<-num0/dnum0
    }    #b11

    for(j in 2:NumI) {     #a
    for(i in 1:n) {   #b1
    if(data[i,2]==1) {   #a2
    ze<-(data[i,1]-muhat[j-1])/sqrt(sighat[j-1])
    de<-dnorm(ze)
    pe<-pnorm(ze)
      wdpe<-de/pe
      w[i,1]<-muhat[j-1]-(sqrt((sighat[j-1]))*wdpe)
      ww[i,1]<-(wdpe)*(wdpe+ze)
    }   #a2
    else {    #e2
    w[i,1]<-data[i,1]
    ww[i,1]<-data[i,1]
        }   #e2
    muhat[j]<-mean(w[,1])
    nume<-sum((w[,1]-muhat[j])^2)
    dnum2<-tapply(ww[,1],list(ww[,2]),sum)["1"]
    dnume<-m+dnum2
```

```
        sighat[j]<-nume/dnume

        }  #b1

    if(abs(muhat[j]-muhat[(j-1)])<1e-007 && abs(sighat[j]-sighat[(j-1)])
    <1e-007) break
    muhatf<-muhat[j]
        sigsqhatf<-sighat[j]
    sighatf<-sqrt(sighat[j])

                }  #a
    musighat<-c(muhatf,sighatf)
    musighat
    }

MLE.EstimatesPooled<-EMAmultvect(datacomb,20)   #### MLEs under
H0N from combined sample (data)

MLE.Estimates1<-EMAmultvect(data1,20)     #### MLEs under HA1N
from sample1 (data1)

MLE.Estimates2<-EMAmultvect(data2,20)     #### MLEs under HA1N
from sample2 (data2)

MLE.Estimates3<-EMAmultvect(data3,20)     #### MLEs under HA1N
from sample2 (data3)



MLE.Estimates<-rbind(MLE.EstimatesPooled,MLE.Estimates1,MLE.Estimates2,
MLE.Estimates3)

datacombest<-numeric(n)

    for(i in 1:n){
    if(datacomb[i,2]==1) datacombest[i]<-log(pnorm((datacomb[i,1]-
    MLE.EstimatesPooled[1])/MLE.EstimatesPooled[2]))
        else datacombest[i]<-log((1/MLE.EstimatesPooled[2])*
        dnorm((datacomb[i,1]-MLE.EstimatesPooled[1])/
        MLE.EstimatesPooled[2]))
    }

    Loglikelihood.H0<-sum(datacombest)

data1est1<-numeric(n1)

    for(i in 1:n1){
    if(data1[i,2]==1) data1est1[i]<-log(pnorm((data1[i,1]-
    MLE.Estimates1[1])/MLE.Estimates1[2]))
        else data1est1[i]<-log((1/MLE.Estimates1[2])*dnorm((data1[i,1]-
        MLE.Estimates1[1])/MLE.Estimates1[2]))
    }

    Loglikelihood.HAdata1<-sum(data1est1)
```

```
data1est2<-numeric(n2)

    for(i in 1:n2){
    if(data2[i,2]==1) data1est2[i]<-log(pnorm((data2[i,1]-
    MLE.Estimates2[1])/MLE.Estimates2[2]))
        else data1est2[i]<-log((1/MLE.Estimates2[2])*dnorm((data2[i,1]-
        MLE.Estimates2[1])/MLE.Estimates2[2]))
    }

    Loglikelihood.HAdata2<-sum(data1est2)

data1est3<-numeric(n3)
    for(i in 1:n3){
    if(data3[i,2]==1) data1est3[i]<-log(pnorm((data3[i,1]-
    MLE.Estimates3[1])/MLE.Estimates3[2]))
        else data1est3[i]<-log((1/MLE.Estimates3[2])*dnorm((data3[i,1]-
        MLE.Estimates3[1])/MLE.Estimates3[2]))
    }
    Loglikelihood.HAdata3<-sum(data1est3)

    Loglikelihood.HA<-Loglikelihood.HAdata1 + Loglikelihood.HAdata2 +
    Loglikelihood.HAdata3

    dfs<-2*(kk-1)
    chisquare0<- -2*(Loglikelihood.H0 - Loglikelihood.HA)
    p.value<- 1 - pchisq(chisquare0 , dfs)

    Test.Result <- c(chisquare0,p.value)

    Test.Output<- rbind(MLE.EstimatesPooled,MLE.Estimates1,
    MLE.Estimates2,MLE.Estimates3,Test.Result)
    Test.Output
 As<-matrix(0,5,6)
    As[1,1]<-"------"
    As[1,2]<-"----------"
    As[1,3]<-"----------------"
    As[1,4]<-"----------------"
    As[1,5]<-"----------"
    As[1,6]<-"-------"
    As[2,1]<-round(MLE.EstimatesPooled[1], 4)
    As[2,2]<-round(MLE.EstimatesPooled[2], 4)
    As[2,3]<-round(Loglikelihood.H0, 4)
    As[2,4]<-round(Loglikelihood.HA ,4)
    As[2,5]<-round(chisquare0, 4)
    As[2,6]<-round(p.value, 4)
    As[3,1]<-round(MLE.Estimates1[1], 4)
    As[3,2]<-round(MLE.Estimates1[2], 4)
    As[3,3]<-"      "
    As[3,4]<-"      "
    As[3,5]<-"      "
    As[3,6]<-"      "
    As[4,1]<-round(MLE.Estimates2[1], 4)
    As[4,2]<-round(MLE.Estimates2[2], 4)
    As[4,3]<-"      "
    As[4,4]<-"      "
```

```
    As[4,5]<-"      "
    As[4,6]<-"      "
    As[5,1]<-round(MLE.Estimates3[1], 4)
    As[5,2]<-round(MLE.Estimates3[2], 4)
    As[5,3]<-"      "
    As[5,4]<-"      "
    As[5,5]<-"      "
    As[5,6]<-"      "

    dimnames(As)<-list(c( "    ", " Poold.Data:   ", "     Data 1:
    ", "     Data 2:   ", "     Data 3:   "), c("mu.hat","sigma.hate",
    "loglikelihood.H0","loglikelihood.HA", "Chisquare0", "P Value"))
    print(As,quote=F)
       invisible()

    }


### DATA: ### =====

data1<-matrix(c(19,10,4,24,17,76,16,22,19,73,43,30,14,4,34,82,13,13,13,
13,4,36,4,73,4,69,26,29,13,42,16,31,46,19,23,10,57,10,36,21,18,15,10,
15,88,72,10,102,13,15,0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,1,1,0,1,0,
1,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,1,0,0,0,1,0,0,0),50,2)

data2<-matrix(c(115,27,9,56,133,65,28,45,30,6,9,6,46,6,21,13,77,18,15,
13,6,57,13,9,56,114,21,24,13,25,36,140,13,9,16,19,36,9,45,97,124,32,21,
30,64,6,14,19,17,47,0,0,1,0,0,0,0,0,0,1,1,1,0,1,0,1,0,0,0,1,1,0,1,1,0,0,
0,0,1,0,0,0,1,1,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0),50,2)

data3<-matrix(c(23,23,18,33,21,94,80,9,77,7,7,40,102,114,43,9,7,14,25,
42,13,13,59,13,122,25,24,58,9,9,59,20,7,9,46,33,37,45,18,56,33,30,15,7,
13,13,37,44,112,87,0,0,0,0,0,0,0,1,0,1,1,0,0,0,0,1,1,0,0,0,1,1,0,1,0,0,
0,0,1,1,0,0,1,1,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0),50,2)

> K.Lognormal.Estimation(data1, data2, data3, 3, 20, T, T, T)

             mu.hat sigma.hate loglikelihood.H0 loglikelihood.HA Chisquare0 P Value
             ------ ---------- ---------------- ---------------- ---------- -------
Poold.Data:  2.9805 1.1516     -209.4725        -208.8886        1.1678     0.8834
    Data 1:  2.8857 1.0733
    Data 2:  2.9981 1.2131
    Data 3:  3.0585 1.1559
```