Journal of Environmental Statistics
August 2013, Volume 5, Issue 1. http://www.jenvstat.org

Sample Size Determination and Hypothesis Testing for the Mean of a Lognormal Distribution

Dulal K. Bhaumik
Department of Psychiatry, UIC
CSPCC, Hines VA Hospital

Kush Kapur
Boston Children's Hospital
Harvard Medical School

Runa Bhaumik
Department of Psychiatry, UIC

Domenic J. Reda
CSPCC, Hines VA Hospital

Abstract

Lognormal distributions have applications in health sciences, occupational exposures, environmental monitoring and several other fields. In this article we compare power functions of four testing procedures based on Student-t, Edgeworth expansion, generalized p-value, and permutation test for testing the mean of a lognormal distribution. We recommend the generalized p-value and permutation test procedures for testing the mean of the distribution. Using these procedures we also determine the required sample size for obtaining a pre-specified power. The results are illustrated with an example from an environmental monitoring field.

Keywords: Lognormal distribution, Edgeworth expansion, Generalized p-value, Permutation test.

1. Introduction

The comparison of the average of a small number of on-site measurements to a larger collection of background (off-site) measurements is an extremely important problem in environmental statistics. Consider an example where a series of n on-site soil samples collected in an area of potential environmental concern are compared to the average of the background concentration distribution, characterized by m background measurements. In a similar problem, we compare n measurements obtained from a ground-water monitoring well located hydraulically downgradient of a waste disposal facility with a concentration limit computed from a series of

m measurements obtained from several ground-water monitoring wells located hydraulically upgradient of the facility. In the literature this kind of problem is generally resolved by (a) comparing the on-site sample mean with the off-site mean, (b) comparing the on-site sample mean with the upper confidence limit (UCL) of the background measurements, and (c) comparing the on-site sample mean with a prediction limit constructed by using the background data. The prediction limit procedure has been extensively discussed in Bhaumik and Gibbons (2004) and Gibbons, Bhaumik, and Aryal (2009). In this article we restrict to the testing and UCL procedures. In addition, we will focus on determining sample sizes while testing the mean and constructing the upper confidence limit (UCL) of a lognormal distribution.

Comparisons of means using t-tests are generally undesirable in this setting because the number of on-site/downgradient samples (n) is generally quite small (*e.g.*, $n = 1$ to 10) relative to the number of background samples ($m = 10$ to 1000). Hence it is critically important to develop an appropriate test statistic for comparing the mean of the on-site data when n is small.

Sample size determination is of great importance for planning studies in environmental and health research. The problem of sample size determination is well defined in the context of hypothesis testing for maintaining nominal Type I error rates and pre-specified power. When conducting a comparison study, there are several fundamental questions that must be answered: (a) what is the minimum number of on-site samples n that are required in order to attain a pre-specified power for testing significance of the parameter of interest?, and (b) how should n be adjusted if it cannot control the pre-specified Type I error rate (α)? Large sample based tests often fail to maintain the aforementioned constraints when in reality sample sizes are small. To test the mean of a right skewed distribution such as lognormal, investigators often use t-tests. Practitioners hardly check whether a t-test with a limited sample size can simultaneously control Type I error rates and produce the desired power. The problem of sample size determination for computing an upper confidence or prediction bound for on-site samples using a lognormally distributed random variable has not been theoretically studied.

The lognormal distribution is often used to model positively skewed contaminant concentration (see Ott 1995; Singh, Singh, and Engelhardt 1997; Gibbons *et al.* 2009). Singh *et al.* (1997) used log-normal distributions for computing the UCL of contaminant concentration data to account for the biased sampling, multiple populations and/or outliers. This distribution has been applied in medicine to model critical dose level (Gaddum 1933; Bliss 1934), in engineering to model particle size in naturally occurring aggregates (Johnson, Kotz, and Balakrishnan (1994)), in pharmacokinetic studies to model the rate and extent of absorption (Hauschke, Steinijans, Diletti, and Burke 1992), in ecological studies to find out the importance of eating habits and the number of fish sampled in the estimation of mercury environmental contamination (Rincon-Leon, Zurera-Cosano, Moreno-Rojas, and Amaro-Lopez 1993). A comprehensive review of applications of lognormal distributions in economics and finance is presented in Crow and Shimizu (1988). In biological fields, the log normal distribution is used to characterize the distribution of microbial diversity and abundance levels (Mandoiu and Zelikovsky 2007). Williams (1937) proposed the use of logarithm transformations to prevent swamping of the results for comparison of number of insects caught under varying conditions in reference to the ones captured in a light trap. In the context of agricultural research, Cochran (1938) discussed some of the difficulties in the analysis of data where the errors from the analysis of variance are dependent on the means. Further, several biological mechanisms such as inheritance of fruit and flower size have been known to fit

log-normal distributions (see Koch 1966, 1969). Pearce (1945), Chapter 10 contains several applications of lognormal distribution in the area of small particle statistics, economics and sociology, biology, anthropometry, household sizes and industrial processes. In addition we refer to Johnson, Kotz, and Balakrishnan (1994) for further details of lognormal distributions.

For lognormal distributions the geometric mean is attractive as a measure of comparison as it leads to considerable mathematical simplification. But the geometric mean is typically not appropriate, as it can dramatically underestimate the prediction limit for the mean. In practice the arithmetic mean is used for comparing the on-site samples with the background measurements. Note that the use of the exponential of a normal confidence or prediction limit for the mean of n on-site samples computed from the natural log transformed data, provides a $(1-\alpha)100\%$ confidence or prediction limit for the geometric mean or median of the N on-site samples and not the arithmetic mean. Hence the aforementioned transformation should be avoided when the interest is in the arithmetic mean. Confidence interval estimation involving means has drawn a significant attention in recent years. In the parametric set up Krishnamoorthy and Mathew (2003), Krishnamoorthy, Mathew, and Ramachandran (2006) have used generalized p-value approach and Zou (2009) has used separate confidence limits for mean and variance to construct confidence interval for the mean(s) of lognormal variable(s).

2. An Example

In order to motivate our ideas we would like to consider an example from a contaminated site. A "brownfield" investigation in which a now closed plating facility is being investigated for future industrial use. At this facility, a portion of the area may have been used as a plating waste sludge disposal area between 1942 and 1944. The capacity of the pond was approximately 5 million gallons. The entire area had been filled and graded by 1955. The area was subsequently developed and is now covered by the eastern portion of an assembly building, concrete paved parking areas, a main electrical substation, and a switch gear house. Investigations were conducted between June 1995 and April 1996 to determine whether regulated constituents were present in soil or groundwater at levels that could pose an unacceptable risk to human health or the environment at this location. Based on the results of the initial investigation, additional investigation was implemented to characterize the vertical and horizontal extent of lead impacts to unsaturated soil. In April 1996, $n = 5$ soil borings were installed to delineate the extent of lead-impacted soil. To determine if the on-site mean lead concentration at this area of the facility exceeded background, $m = 15$ off-site soil samples were collected in areas that were not influenced by the activities at the facility. The data for both off-site and on-site locations are displayed in Table 1.

Table 1: Lead in mg/kg for soil boring samples in off-site and on-site locations.

off-site	26	63	3	70	16
	5	1	57	5	3
	24	2	1	48	3
on-site	50	82	95	103	88

We used Anderson-Darling, Kolmogorov-Smirnov, Cramer-von Mises, and Shapiro Wilk tests

of normality for the background data. The null hypothesis of normality was not rejected by Anderson-Darling, Kolmogorov-Smirnov, and Cramer-von Mises tests. However, the null hypothesis of normality of the background data using the Shapiro Wilk test was rejected at the 5% level ($W = .787$, 95% critical value = .881), but the null hypothesis of log-normality was not rejected at the 5% level ($W = 0.882$). This motivates us to look for a test based on a lognormal distribution. Two fundamental questions that arise for this problem are (i) how to compare the on-site measurements with the off-site measurements?, and (ii) is the on-site sample size 5 sufficient for any valid comparison?

3. Statistical Foundation

Let y be a lognormal random variable, i.e., $ln(y) \sim N(\mu, \sigma^2)$, where $\mu = E(ln(y))$, and σ^2=Var$(ln(y))$. Assume that a random sample $y_1, \cdots, y_m$ of size m is available from this distribution. Let

$$\hat{\mu} = \frac{\sum_{i=1}^{m}(ln(y_i))}{m} \text{ and } \hat{\sigma^2} = \frac{\sum_{i=1}^{m}(ln(y_i) - \hat{\mu})^2}{m-1}. \tag{1}$$

Note that $\hat{\mu}$ and $\hat{\sigma^2}$ are jointly sufficient statistics for μ and σ^2, and they are consistent estimators for large m. Let $x_1, \cdots, x_n$ be a random sample of size n from on-site location. Also, let $X = \sum_{k=1}^{n} x_k$ and $\bar{x} = X/n$.

Let the minimum variance unbiased estimator (MVUE) of $E(\bar{x})$ be denoted by $\hat{\theta}$. The MVUEs of Var$(\hat{\theta})$ and Var(x) are denoted respectively by $\hat{\phi}$ and v. The expressions of $\hat{\theta}$ and v are given in Finney (1941), and that of $\hat{\phi}$ is given in Hoyle (1968). Gilbert (1987) observed that for small sample sizes, the MVUEs can perform poorly. Gilbert (1987) gives less efficient but simpler estimators that perform well for small samples. His estimators are

$$\begin{aligned} \hat{\theta} &= e^{\hat{\mu}+\hat{\sigma}^2/2}, \\ v &= \hat{\theta}^2\left[e^{\hat{\sigma}^2} - 1\right]. \end{aligned} \tag{2}$$

Johnson, Kotz, and Balakrishnan (1994) give a simplified estimator $\hat{\phi}$ of the variance of $\hat{\theta}$.

$$\hat{\phi} = \widehat{Var}(\hat{\theta}) = e^{2\hat{\mu}+\hat{\sigma}^2}\left(\frac{\hat{\sigma}^2}{m} + \frac{\hat{\sigma}^4}{2m}\right). \tag{3}$$

The differences in the tail properties of a t-distribution and standard normal start showing up for samples less than 9. Hence, in practice sample sizes of less than 9 are being considered as small. For the off-site measurements of our example, the maximum likelihood estimates of μ and σ^2 are 2.1815 and 2.3463 respectively. Hence $\hat{\theta} = 28.6354$, $v = 774.62$ and $\hat{\phi} = 278.73$. For the rest of the article we will consider the values of μ and σ^2 of the off-site population as 2.1815 and 2.3463 respectively.

4. Hypothesis Testing

For the comparison of the on-site population with the off-site population we develop our methodology in the context of the example presented in Section 2. Our fundamental goal

in this article is to determine the required sample size for comparing the mean lead concentration of the on-site facility with that of the background facility. We make this comparison by two different ways as stated above, namely, (i) a testing procedure that assumes that the parameters estimated from the background samples are consistent; more precisely these estimates will be treated as the parameters of the background population, (ii) the UCL procedure that compares the on-site mean with the UCL of the background measurements. In the testing procedure we compare the on-site sample mean with the mean of the background distribution by four different methods: (a) Student-t distribution, (b) Edgeworth expansion, (c) Generalized p-value, and (d) Permutation test.

4.1. Student-t Test

Let $\bar{x}$ and s_x^2 be respectively the sample mean and variance based on n independent observations obtained from the on-site facility. Let ν and ν_0 be respectively the on-site and off-site mean concentrations on the original scale. To test $H_0 : \nu = \nu_0$, we use Student t-test denoted by T that has the following expression

$$T = \frac{\sqrt{n}(\bar{x} - \nu_0)}{s_x}, \tag{4}$$

where the assumption is that T follows a t distribution with degrees of freedom (df) $(n - 1)$. Investigators often assume that the sample mean of the background facility is a good representation of ν and variances of these populations do not change significantly. Even if the assumptions are correct, it is clear that any comparison based on T for this problem will be inappropriate as the data is non-normal and the on-site sample size is small.

4.2. Edgeworth Expansion for Student-t

In this subsection we propose a test based on the Edgeworth expansion of the distribution of T defined in equation 4. Results regarding the distribution of the sample (arithmetic) mean for small samples using Edgeworth or Gram-Charlier's expansions are available in the literature. Sugden, Smith, and Jones (2000) investigated Cochran's rule for the minimum sample size to ensure adequate coverage of nominal 95% confidence intervals for the population mean using the Edgeworth's expansion. Bhaumik and Gibbons (2004) used Gram-Charlier Type A series to construct the upper prediction limit for the mean of a sample. The main problem that we encounter when a distribution obtained by the Edgeworth expansion is implemented for hypothesis testing with small samples is its conservativeness.

Let U_n be the Student-t statistic (which is defined in 4) based on a sample of size n. Using the Edgeworth expansion, Sugden *et al.* (2000) provides the following approximate expression for the distribution function of U_n.

$$\begin{aligned} F(u) &= Pr(U_n \le u) = \Phi(u) + \frac{q_1(u)\phi(u)}{\sqrt{n}} + \frac{q_2(u)\phi(u)}{n} + O(n^{-3/2}), \qquad (5) \\ \text{where } q_1(u) &= \gamma_1\Big(\frac{1}{2} + \frac{(u^2-1)}{3}\Big), \\ \text{and } q_2(u) &= u\Big[\gamma_2\Big(\frac{(u^2-3)}{12}\Big) - \Big(1 + \frac{(u^2-3)}{4}\Big) \\ &\quad -\gamma_1^2\Big(1 + \frac{2(u^2-3)}{3} + \frac{(u^4-10u^2+15)}{18}\Big)\Big]. \end{aligned}$$

and γ_1 and γ_2 are respectively the skewness and kurtosis of the distribution. Here $\Phi(u)$ and $\phi(u)$ are the cumulative distribution and probability density functions of a standard normal distribution. Thus from the above expansion of the distribution function of U_n we see that it is adjusted with the skewness and kurtosis of U_n. For a symmetric parent distribution $\gamma_1 = \gamma_2 = 0$. Thus the influence of $q_1(u)$ completely disappears and the contribution of the third term in equation 5 diminishes for large values of n. Hence, the resultant distribution is more close to a normal distribution under this scenario. On the other hand, for a skewed distribution the impact of both second and third terms in equation 5 will start revealing and results will depart from a normal distribution.

We use the lead background data and estimate the relevant parameters for U_n. The estimates of γ_1 and γ_2 from the lead background data are 0.8705 and -0.8433 respectively. In Figure 1 we compare $F(U)$ with a standard normal distribution $\Phi(u)$ and see that $F(U)$ is right skewed for $n = 5$.

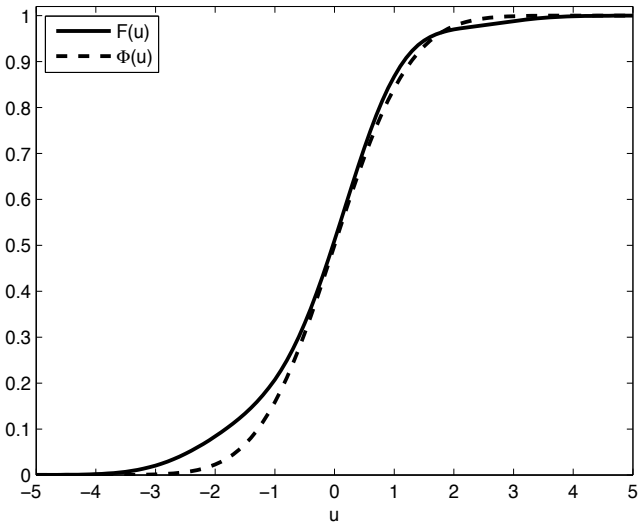


Figure 1: Cumulative Distribution Function by Edgeworth Expansion for $n = 5$.

Denote the upper $(1 - \alpha)$th percentile point of the distribution of U_n by $U_n(1 - \alpha)$. Using this expansion we construct a test for the on-site measurements as follows: reject H_0 if $U_n > U_n(1 - \alpha)$. In Section 4.4 we explore the performance of this test numerically and compare it with two other tests.

4.3. Generalized p-value

Using the concept of generalized p-value, Krishnamoorthy and Mathew (2003) developed a test for the mean of a lognormal distribution. Let $Z = ln(x) \sim N(\mu_z, \sigma_z^2)$. Also let $\bar{Z}$ and S_z^2 be respectively the sample mean and sample variance on the natural log scale. For a particular given sample, when we know the values of $\bar{Z}$ and S_z^2 , denote them by $\bar{z}$ and s_z^2 and call them as observed values. Note that $\nu = E(x) = e^{\mu_z + \frac{\sigma_z^2}{2}}$ and $\tau^2 = var(x) = (e^{2\mu_z + \sigma_z^2})(e^{\sigma_z^2} - 1)$. We set the hypothesis at $H_0 : \beta = \mu_z + \frac{\sigma_z^2}{2} \leq \beta_0$, against the alternative $H_1 : \beta > \beta_0$. Under the

null hypothesis β_0 is known. The test is based on a statistic known as the generalized test statistic and we denote it by $G_p(\beta)$. $G_p(\beta)$ is required to satisfy the following three conditions:

1. The distribution of $G_p(\beta)$ is stochastically decreasing in β, i.e., $P(G_p(\beta) \geq c)$ is a decreasing function of β.
2. For any given known β, the distribution of $G_p(\beta)$ does not depend on any unknown parameters.
3. At the observed value, i.e., $\bar{Z} = \bar{z}$ and $S_z^2 = s_z^2$, $G_p(\beta)$ does not depend on any unknown parameters.

The expression of $G_p(\beta)$ is as follows.

$$G_p(\beta) = \bar{z} - (\frac{\bar{Z} - \mu_z}{S_z/\sqrt{n}})(s_z/\sqrt{n}) + \frac{\sigma_z^2 s_z^2}{2S_z^2} - \beta. \tag{6}$$

The observed value of $G_p(\beta)$ is zero and hence it does not depend on any parameters. The generalized p-value for this test is $p = P(G_p(\beta) \leq 0 | \beta = \beta_0)$. The test procedure is to reject H_0 if p is less than the pre-specified nominal level α. This is equivalent to reject H_0 if $G_p(\alpha, \beta_0) > 0$, where $G_p(\alpha, \beta_0)$ is the $\alpha 100\% th$ percentile point of the distribution of $G_p(\beta_0)$ under the null hypothesis. This is because of the monotonically decreasing distributional property of G_p and $P(G_p \leq 0) \leq P(G_p \leq G_p(\alpha, \beta_0 | \beta_0)) = \alpha$. We will use this concept in Algorithm 2 to compute simulated Type I rates and power of the test.

The cut-off value of $G_p(\alpha, \beta_0)$ is determined via simulation. A careful look at the expression of $G_p(\beta)$ tells that it basically depends on a standard normal distribution and a χ^2 distribution with df $n-1$. In its expression $\bar{z}$ and s_z^2 are fixed and they do not have any distributions. To compute the generalized p-value we follow Algorithm 1.

Algorithm 1

1. For a given data set $x_1, x_2, \cdots, x_n$, compute $\bar{z}$ and s_z^2.
2. For $i = 1, 2, \ldots, N$ generate $W_i \sim N(0,1)$ and $U_i^2 \sim \chi_{n-1}^2$.
3. For each i, compute $G_{pi}(\theta_0) = \bar{z} - (\frac{W_i\sqrt{n-1}}{U_i})(\frac{s_z}{\sqrt{n}}) + \frac{s_z^2(n-1)}{2U_i^2} - \beta_0$ and if $G_{pi}(\beta_0) \leq 0$, set $R_i = 1$ otherwise 0.
4. Compute $\bar{R} = \frac{\sum_{i=1}^N R_i}{N}$.

$\bar{R}$ is the Monte Carlo simulated p-value and the lower $\alpha 100\% th$ ordered $G_{pi}(\beta_0)$ is the Monte Carlo estimate of the $\alpha 100\% th$ percentile point of the distribution of $G_p(\beta_0)$. The test based on the generalized p-value is to reject H_0 if the generalized p-value is less than α.

In order to see the performance of this procedure we simulate the Type I error rates and power for values of μ_z and σ_z^2 estimated from the data discussed in the Example Section. We provide the simulation strategy in Algorithm 2.

Algorithm 2

1. Specify n, μ_z, and σ_z^2.
2. For $i = 1, 2, \ldots, N_1$, generate $\bar{z}_i$ from $N(\mu_z, \frac{\sigma_z^2}{n})$ and Q_i from χ^2_{n-1} and compute $s^2_{zi} = \frac{\sigma_z^2 Q_i}{n-1}$.
3. For $j = 1, 2, \ldots, N_2$, generate $W_j \sim N(0, 1)$ and $U_j^2 \sim \chi^2_{n-1}$.
4. Compute $G_p(j, \beta, i)$ using W_j, U_j^2, $\bar{z}_i$ and s^2_{zi}.
5. Get the lower $\alpha 100\% th$ ordered $G_p(j, \beta, i)$ for each i varying over j only. Denote it by $G_p(\alpha, \beta, i)$.
6. If $G_p(\alpha, \beta, i) > 0$, set $R_i = 1$, 0 otherwise.
7. Compute $\bar{R} = \frac{\sum_{i=1}^{N_1} R_i}{N_1}$.

$\bar{R}$ provides the simulated Type I error rate when each $G_p(j, \beta, i)$ is computed using the given null value of β_0. On the other hand, when each $G_p(j, \beta, i)$ is computed under the alternative hypothesis, $\bar{R}$ provides the simulated power of the test. Note that under the null hypothesis (i.e., $\beta = \beta_0$), μ_z and σ_z^2 must satisfy the condition $\beta_0 = \mu_z + \frac{\sigma_z^2}{2}$.

4.4. Permutation Test

Permutation tests are non-parametric that have applications in biology, medicine, engineering and many other areas. These tests have tremendous flexibility that cannot be achieved by many parametric tests. In real world examples, parametric methods require the introduction of some strong assumptions which are not easily justified. In contrast permutation based methods are intrinsically robust and provide inferences without losing much of efficiency. For most of the situations the Type I error rates of permutation tests are close to the nominal level for the entire range of baseline values with a compromise of power relative to asymptotic tests. In most cases, permutation based methods are conditional on the sufficient statistics under the null hypothesis. For more details regarding permutation tests we refer to Good (2005); Chihara and Hesterberg (2011). We outline an algorithm to implement the permutation test below:

1. Compute the difference in sample means between the two observed samples.
2. Pool the observations of two observed samples together.
3. Permute the observation labels in this pooled sample.
4. Split the permuted observations into two groups maintaining the observed sample ratio and compute the difference of the means of these groups.
5. Repeat Steps 3 and 4 for $i = 1, 2, \ldots, I$ (For example, $I = 5,000$).
6. Order all the I mean differences and obtain the 95^{th} percentile value of this ordered mean differences.

7. Reject the null if observed sample mean difference is greater than the 95^{th} percentile value of the ordered mean differences.

We denote this test by PT. Note that test of null hypothesis is rejected by all of these four procedures for our example.

4.5. Simulation Study

In this section we evaluate the performance of these fours tests in terms of Type I error rates for various combinations of $\mu = 0, (1), 4$, $\sigma^2 = 0.25, (0.50), 2.25$, and $n = 5, 15, 25$ and 50. The results are presented in Tables 2-5. The data were simulated from lognormal distributions for both background and on-sites with $10,000$ replications (*i.e.*, $N = N_1 = N_2 = 10,000$). For $n = 5$ and 15 the background samples were taken to be $2 \times n$ (i.e., the background sample size is twice of the on-site sample size). Inspecting these Tables we find that (i) for all the combinations of μ and σ^2 and for all the values of n the permutation test maintains the Type I error rate to its nominal value of 5%, (ii) for all combination μ and σ^2, both student-t test and U_n are very conservative and their performances improve for large values of n, (iii) even for $n = 50$ the Type I error rates for both T and U_n remain in the neighorhood of 0.03, and (iv) G_p has a very slight inflated Type I error rate for $n = 5$ and 15. For $n = 25$ and $n = 50$ the Type I error rates G_p are in the neighborhood of 5%. Hence, in this simulation study we establish that both G_p and PT perform quite well in terms of controlling Type I error rates. In addition, we study the power curve of G_p for (i) $\mu = 2$ and $\sigma^2 = 2.25$, and (ii) $\mu = 2$ and $\sigma^2 = 0.25$ for different values of $n = 5, 15, 25$ and 50. The power plots are presented in Figures 2a-b. As expected these power curves are highly depended on the sample size n. Comparing Figures 2a and 2b for the same value of μ and n, we get more power when the σ^2 is small. For $\mu = 2.00$ and $n = 15$, the power is 0.98 when $\sigma^2 = 0.25$ whereas it is close to 0.31 when $\sigma^2 = 2.25$. In order to study, the robustness of these procedures we simulated data from gamma distributions with shape parameters $\kappa = 0.25, (0.25), 3.00$ and scale parameter 1. Both T and U_n provided us conservative results $\kappa = 0.25, (0.25), 1.5$, however for $\kappa = 2.00, (0.25), 3.00$ Type I error rates are in the neighborhood of 4%. The PT test preserves the same behavior as mentioned before. Type I error rates of G_p varied from $4.9\% - 7.8\%$.

Table 2: Simulated Type I error rates of T, U_n, G_p, PT for various combinations of σ^2 and μ when sample size is fixed at 5.

σ^2	Tests	$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
0.25	T	0.016	0.016	0.017	0.017	0.018
	U_n	0.021	0.022	0.022	0.021	0.022
	G_p	0.066	0.066	0.065	0.065	0.064
	PT	0.050	0.050	0.050	0.050	0.050
0.75	T	0.005	0.005	0.006	0.005	0.006
	U_n	0.009	0.009	0.009	0.009	0.009
	G_p	0.067	0.066	0.066	0.065	0.065
	PT	0.050	0.050	0.050	0.050	0.050
1.25	T	0.003	0.003	0.003	0.003	0.003
	U_n	0.007	0.007	0.007	0.007	0.007
	G_p	0.065	0.066	0.066	0.067	0.065
	PT	0.050	0.050	0.050	0.050	0.050
1.75	T	0.001	0.001	0.001	0.001	0.001
	U_n	0.005	0.005	0.005	0.004	0.005
	G_p	0.066	0.067	0.066	0.066	0.067
	PT	0.050	0.050	0.050	0.050	0.050
2.25	T	0.001	0.001	0.001	0.001	0.001
	U_n	0.005	0.005	0.005	0.005	0.005
	G_p	0.067	0.067	0.067	0.066	0.066
	PT	0.050	0.050	0.050	0.050	0.050

Table 3: Simulated Type I error rates of T, U_n, G_p, PT for various combinations of σ^2 and μ when sample size is fixed at 15.

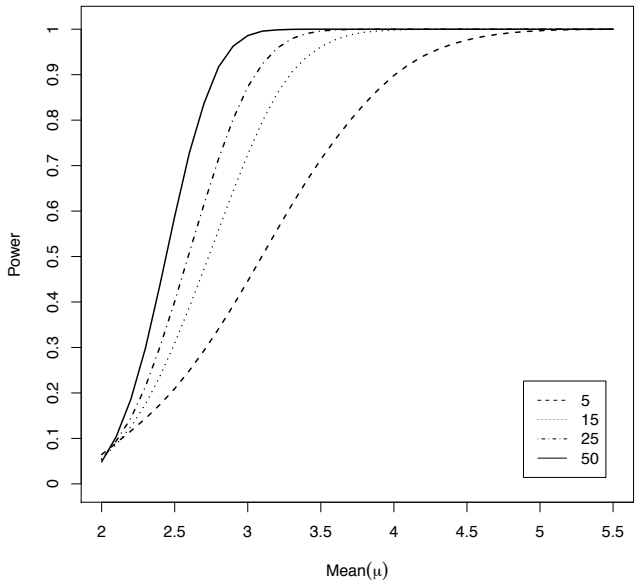
σ^2	Tests	$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
0.25	T	0.019	0.019	0.019	0.019	0.019
	U_n	0.023	0.025	0.024	0.023	0.024
	G_p	0.057	0.057	0.056	0.055	0.056
	PT	0.050	0.050	0.050	0.050	0.050
0.75	T	0.009	0.009	0.009	0.009	0.009
	U_n	0.012	0.012	0.012	0.013	0.013
	G_p	0.056	0.056	0.057	0.056	0.056
	PT	0.050	0.050	0.050	0.050	0.050
1.25	T	0.005	0.005	0.005	0.005	0.005
	U_n	0.007	0.009	0.008	0.008	0.009
	G_p	0.056	0.057	0.057	0.054	0.055
	PT	0.050	0.050	0.050	0.050	0.050
1.75	T	0.004	0.003	0.003	0.003	0.003
	U_n	0.008	0.008	0.006	0.008	0.008
	G_p	0.056	0.055	0.056	0.056	0.056
	PT	0.050	0.050	0.050	0.050	0.050
2.25	T	0.001	0.001	0.002	0.001	0.001
	U_n	0.004	0.004	0.004	0.004	0.004
	G_p	0.058	0.057	0.058	0.058	0.058
	PT	0.050	0.050	0.050	0.050	0.050

Table 4: Simulated Type I error rates of T, U_n, G_p, PT for various combinations of σ^2 and μ when sample size is fixed at 25.

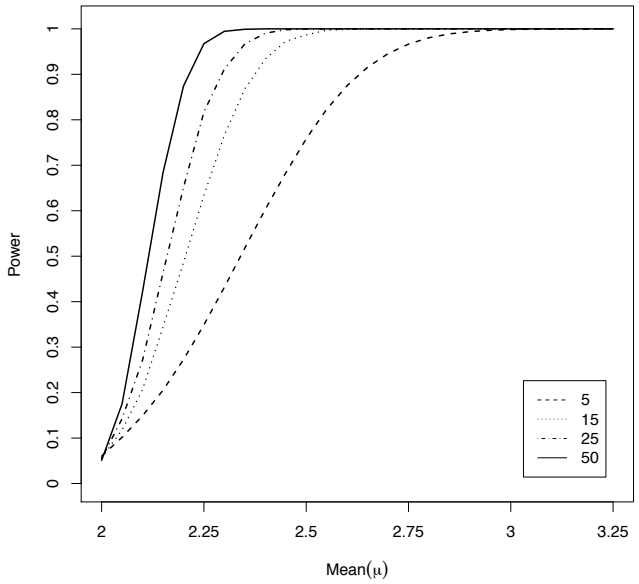
σ^2	Tests	$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
0.25	T	0.022	0.022	0.022	0.022	0.022
	U_n	0.025	0.024	0.024	0.024	0.024
	G_p	0.054	0.054	0.055	0.055	0.055
	PT	0.050	0.050	0.050	0.050	0.050
0.75	T	0.012	0.012	0.012	0.011	0.011
	U_n	0.015	0.015	0.015	0.015	0.016
	G_p	0.055	0.054	0.054	0.054	0.055
	PT	0.050	0.050	0.050	0.050	0.050
1.25	T	0.006	0.006	0.006	0.006	0.006
	U_n	0.009	0.009	0.009	0.010	0.009
	G_p	0.056	0.055	0.055	0.054	0.055
	PT	0.050	0.050	0.050	0.050	0.050
1.75	T	0.004	0.004	0.004	0.004	0.004
	U_n	0.006	0.006	0.006	0.006	0.006
	G_p	0.055	0.055	0.054	0.055	0.055
	PT	0.050	0.050	0.050	0.050	0.050
2.25	T	0.003	0.003	0.002	0.003	0.002
	U_n	0.005	0.004	0.004	0.004	0.004
	G_p	0.056	0.055	0.056	0.055	0.056
	PT	0.050	0.050	0.050	0.050	0.050

Table 5: Simulated Type I error rates of T, U_n, G_p, PT for various combinations of σ^2 and μ when sample size is fixed at 50.

σ^2	Tests	$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
0.25	T	0.029	0.028	0.030	0.029	0.029
	U_n	0.030	0.029	0.030	0.030	0.032
	G_p	0.051	0.052	0.051	0.050	0.052
	PT	0.050	0.050	0.050	0.050	0.050
0.75	T	0.017	0.017	0.016	0.015	0.015
	U_n	0.019	0.019	0.019	0.019	0.019
	G_p	0.052	0.052	0.051	0.052	0.052
	PT	0.050	0.050	0.050	0.050	0.050
1.25	T	0.009	0.009	0.009	0.009	0.009
	U_n	0.011	0.011	0.012	0.011	0.011
	G_p	0.052	0.051	0.052	0.050	0.051
	PT	0.050	0.050	0.050	0.050	0.050
1.75	T	0.007	0.007	0.006	0.006	0.006
	U_n	0.009	0.009	0.009	0.009	0.009
	G_p	0.052	0.052	0.052	0.053	0.053
	PT	0.050	0.050	0.050	0.050	0.050
2.25	T	0.003	0.004	0.004	0.004	0.004
	U_n	0.006	0.004	0.005	0.004	0.006
	G_p	0.052	0.052	0.053	0.052	0.053
	PT	0.050	0.050	0.050	0.050	0.050



(a) $\mu = 2, \sigma^2 = 2.25$



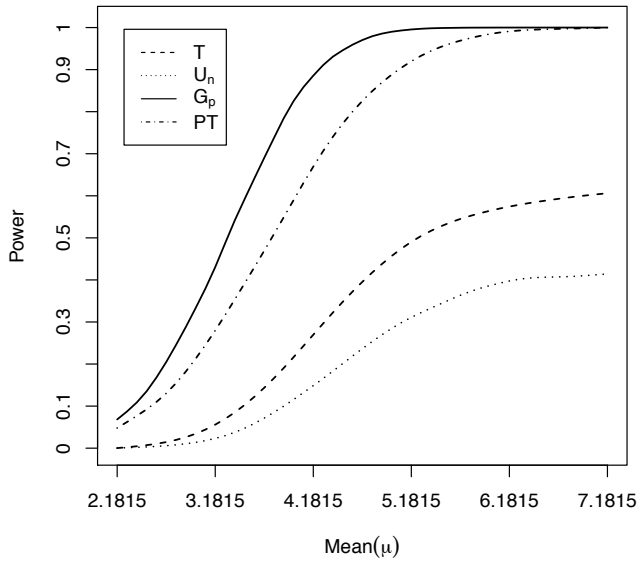
(b) $\mu = 2, \sigma^2 = 0.25$

Figure 2: Comparison of Power Curves of G_p for $n = 5, 15, 25$, and 50.

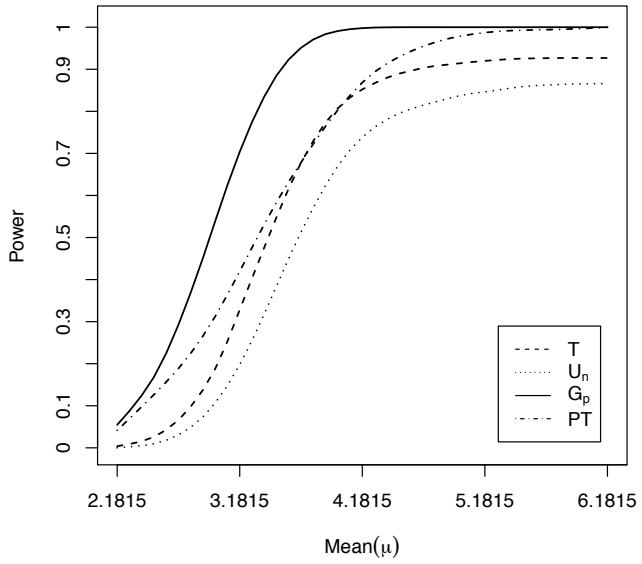
4.6. Results for the example

We have performed a simulation study to visualize the impact of skewness on testing and sample size determination for $n = 5$ and $n = 15$ in the context of the example presented in Section 2. To compare the on-site measurements with the off-site data by the hypothesis testing procedure we set the null, $H_0 : \mu = 2.1815$. We assume that the variance is $\sigma^2 = 2.3463$. Our simulation study is based on $N = N_1 = N_2 = 10,000$ replications from normal and chi-square distributions for the generalized p-value approach. We compute both Type I error rates and power for $\mu = 2.1815, 2.2815, \cdots, 9.1815$. The power curves for all four tests are presented in Figure 3. Inspecting Figure 3(a) we see that both Student-t (T) and Edgeworth expansion based test (U_n) are very conservative for $n = 5$. Contrary to that, the Generalized p-value test (G_p) has the Type I error rate close to 7% and that of Permutation test (PT) is close to 5%. Further, the power curves of G_p and PT rise up sharply whereas those of T and U_n cannot attain even 70% for $\mu = 9.1815$. The behavior of these power curves remain the same for $n = 15$ (see Figure 3(b)) except that the simulated Type I error rate of G_p reduces to 5.8%. The G_p needs a sample size close to 50 to achieve the Type I error rate to its nominal value of 5% in this example.

Remark: We have computed γ_1 and γ_2 for the Edgeworth expansion using the background sample. We have also computed these two measures using the third and fourth ordered moments of a lognormal distribution with $\mu = 2.1815$, and $\sigma^2 = 2.3463$. The behavior of the power function of the Edgeworth expansion with the latter approach was still unsatisfactory.



(a) $n = 5$



(b) $n = 15$

Figure 3: Comparison of Power Curves for $n = 5$ and $n = 15$.

5. Sample Size Determination

In the previous section we have established the superiority of the Generalized p-value and Permutation test compared to Student-t and Edgeworth expansion based tests via simulation study. We also noticed that both U_n and T are very conservative in terms of controlling Type I error rates. In the following we investigate the impact of this factor in determining sample sizes. Based on these four tests G_p, PT, U_n and T, we determine sample sizes using an iterative approach for testing $H_0 : \mu = 2.1815$ against alternatives $\mu_a = 2.2815, 2.3815, \cdots, 2.8815$ to achieve 80% power. For this iterative approach we start with a small value of n and then keep on increasing it until the desired 80% power is achieved. In order to nullify the over power of one group in the permutation test we take equal sample sizes (n) in both groups. We determine the Type I error rate for that particular value of n. Results are presented in Table 6. Inspecting this table we see that the sample sizes determined by G_p are significantly smaller compared to PT, U_n and T. In fact for different alternatives the required sample size by G_p is almost 50% less than those determined by U_n and T. In addition, we provide in this table the simulated Type I error rates and corresponding power. Table 6 reveals that all the above tests attain the target power of 80% and only G_p and PT maintain the Type I error rates close to the nominal level of 5%.

Table 6: Sample Sizes required by GP, Student t, and Edgeworth Expansion with their corresponding simulated Type I error rate (STR) and Power (SP).

	G_P			Student t			Edgeworth			Permutation		
μ_a	n	STR	SP	n	STR	SP	n	STR	SP	n	STR	SP
2.2815	3040	0.049	0.798	5195	0.027	0.799	5194	0.028	0.803	11284	0.050	0.803
2.3815	704	0.051	0.803	1240	0.016	0.800	1221	0.019	0.803	2616	0.051	0.803
2.4815	303	0.052	0.809	534	0.015	0.802	523	0.016	0.801	1096	0.050	0.810
2.5815	159	0.051	0.799	293	0.011	0.799	282	0.013	0.801	553	0.051	0.799
2.6815	101	0.049	0.801	190	0.008	0.801	178	0.012	0.805	342	0.049	0.805
2.7815	68	0.051	0.800	131	0.006	0.801	119	0.010	0.801	223	0.051	0.811
2.8815	48	0.053	0.811	97	0.004	0.804	88	0.007	0.805	155	0.051	0.805

6. Upper Confidence Limit and the Coverage Probability

Comparing the on-site sample with the background data by UCL is very common in the literature (see Singh *et al.* 1997). In this procedure we construct the UCL using the background data and compare the sample mean of the on-site measurement with the UCL. The concept of constructing the UCL of β by the generalized p-value approach (we restrict only to the Generalized p-value approach) is to obtain the $(1-\alpha)100$th percentile point of $G_p = \bar{z} - (\frac{\bar{Z}-\mu_z}{S_z/\sqrt{n}})(s_z/\sqrt{n}) + \frac{\sigma_z^2 s_z^2}{2S_z^2}$. Denote this percentile point by $G_p(1-\alpha)$. Then $G_p(1-\alpha)$ is the $(1-\alpha)100\%$ generalized UCL for β. In the context of the example discussed before, we set $\alpha = .05$ and determine $G_p(1-\alpha)$ for $\mu = 2.1815, 2.2815, \cdots, 2.8815$. In addition we numerically compute the corresponding coverage probability for each UCL via simulation based on $100,000$ replications for $n = 10, 15, \cdots, 30$. In Table 7 we provide simulated coverage

probability for each upper confidence limit determined by $G_p(1-\alpha)$. Table 7 reveals that for all values of n, the simulated coverage probability is close to 95%. For $n = 5$ the coverage probability is in the neighborhood of 93%.

Table 7: Coverage probability of 95% upper confidence limit based on G_p.

$\mu \backslash n$	10	15	20	25	30
2.181	0.952	0.948	0.956	0.955	0.947
2.281	0.949	0.955	0.948	0.951	0.955
2.381	0.947	0.956	0.950	0.952	0.953
2.481	0.954	0.955	0.950	0.949	0.951
2.581	0.951	0.955	0.953	0.952	0.950
2.681	0.953	0.951	0.948	0.952	0.949
2.781	0.947	0.949	0.953	0.952	0.944
2.881	0.955	0.951	0.946	0.953	0.949

The 95% UCL for β for the background data is 4.872 and the estimated β for the on-site data is 5.569. It is now clear that a sample of size 5 for the background is inadequate both for testing and constructing confidence interval for the mean.

7. Discussion

In this article we have focused on sample size determination and four testing procedures for the mean of a lognormal distribution. Results are illustrated with simulation studies and with an example from environmental field. Previous work in this area has focused on prediction limits of the sample mean of a lognormal distribution using the exact distribution and various approximate distributions. We have demonstrated that the generalized p-value approach is the most effective method for sample size determination. If the lognormal distribution is questionable then the Permutation test is a strong competitor to generalized p-value approach. The results have applications in many other fields like biological sciences, workplace exposure to contaminants, health monitoring, daily rainfall predictions etc. In addition, we have explored the performance of our procedures under certain types of violations of distributions. At this stage, we do not know how good the performance of our procedures will be under different types of violations of distributions. For example, it is not known how robust our results will be when data are generated from a mixture of gamma and lognormal distributions. A comparison of cost analysis can be made using each of the procedures developed in this article, when the real cost for collecting a sample unit is known. Hence from that point of view, the generalized p-value approach will be even more appealing.

Acknowledgments

We would like to thank the editor and two referees for their helpful comments and suggestions in order to improve the quality of this manuscript.

References

Bhaumik DK, Gibbons RD (2004). "An upper prediction limit for the arithmetic mean of a lognormal random variable." *Technometrics*, **46**(2), 239–248. ISSN 00401706.

Bliss C (1934). "The method of probits." *Science*, **79**, 38–39.

Chihara L, Hesterberg T (2011). *Mathematical statistics with resampling and R*. Wiley.

Cochran W (1938). "Some difficulties in the statistical analysis of replicated experiments." *Empire Journal of Experimental Agriculture*, **6**, 157–163.

Crow E, Shimizu K (1988). *Lognormal distributions:Theory and applications*. Marcel Dekker, New York.

Finney D (1941). "On the distribution of a variate whose logarithm is normally distributed." *Journal of Royal Statistical Society Series B*, **7**, 155–161.

Gaddum J (1933). *Reports on biological standards III. Methods of biological assay depending on a quantal response*, volume 183. Special Report Series, Medical Research Council, London.

Gibbons R, Bhaumik D, Aryal S (2009). *Statistical methods for groundwater monitoring*. John Wiley & Sons, Hoboken: New Jersey.

Gilbert R (1987). *Statistical methods for environmental pollution monitoring*. Van Nostrand Reinhold, New York.

Good P (2005). *Permutation, parametric, and bootstrap tests of hypotheses*. Springer.

Hauschke D, Steinijans V, Diletti E, Burke M (1992). "Sample size determination for bioequivalence assessment using a multiplicative model." *The Annals of Mathematical Statistics*, **20**, 557–561.

Hoyle M (1968). "The estimation of variances after using a Gaussianating transformation." *The Annals of Mathematical Statistics*, **39**, 1125–1143.

Johnson N, Kotz S, Balakrishnan N (1994). *Continuous univariate distributions*. John Wiley & Sons, New York.

Koch A (1966). "The logarithm in biology I. Mechanisms generating the log-normal distribution exactly." *Journal of Theoretical Biology*, **23**, 276–290.

Koch A (1969). "The logarithm in biology II. Distributions simulating the lognormal." *Journal of Theoretical Biology*, **23**, 251–268.

Krishnamoorthy K, 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, 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.

Mandoiu I, Zelikovsky A (2007). *Bioinformatics research and applications: third international symposium, ISBRA*. Springer.

Ott W (1995). *Environmental statistics and data analysis*. CRC press, Boca Raton: Florida.

Pearce S (1945). "Lognormal distribution." *Nature*, **156**, 747.

Rincon-Leon F, Zurera-Cosano G, Moreno-Rojas R, Amaro-Lopez M (1993). "Importance of eating habits and sample size in the estimation of environmental mercury contamination using biological indicators." *Environmental monitoring and assessment*, **27**, 193–200.

Singh A, Singh A, Engelhardt M (1997). "The lognormal distribution in environmental applications." *Technical Report EPA/600/S-97/006*, EPA Technology Support Center Issue, National Exposure Research Laboratory, Environmental Sciences Division.

Sugden R, Smith T, Jones R (2000). "Cochran's rule for simple random sampling." *Journal of Royal Statistical Society Series B*, **62**, 787–793.

Williams C (1937). "The use of logarithms in the interpretation of certain entomological problems." *Annals of Applied Biology*, **24**, 404–414.

Zou G (2009). "Simple confidence intervals for lognormal means and their differences with environmental applications." *Environmetrics*, **20**, 172–180.

Affiliation:

Dulal K. Bhaumik
Department of Psychiatry and
Division of Epidemiology and Biostatistics
University of Illinois at Chicago
1601 W. Taylor St.
Chicago, IL 60612
E-mail: dbhaumik@psych.uic.edu

Kush Kapur
Clinical Research Center
Boston Children's Hospital
Department of Neurology
Harvard Medical School
21 Autumn Street
Boston, MA 02215

Runa Bhaumik
Department of Psychiatry and
University of Illinois at Chicago
1601 W. Taylor St.
Chicago, IL 60612

Domenic J. Reda
Cooperative Studies Program Coordinating Center
Hines VA Hospital
5000 South 5th Avenue, Building 1
Hines, IL 60141

Journal of Environmental Statistics http://www.jenvstat.org
Volume 5, Issue 1 Submitted: 2012-07-15
August 2013 Accepted: 2013-01-07