Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT

Journal of Statistical Planning and Inference 115 (2003) 103–121

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals

K. Krishnamoorthy[a,] *, Thomas Mathew[b]

[a] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA
[b] Department of Mathematics and Statistics, University of Maryland, Baltimore, MD 21250, USA

Received 19 July 2001; accepted 17 February 2002

Abstract

The lognormal distribution is widely used to describe the distribution of positive random variables; in particular, it is used to model data relevant to occupational hygiene and to model biological data. A problem of interest in this context is statistical inference concerning the mean of the lognormal distribution. For obtaining confidence intervals and tests for a single lognormal mean, the available small sample procedures are based on a certain conditional distribution, and are computationally very involved. Occupational hygienists have in fact pointed out the difficulties in applying these procedures. In this article, we have first developed exact confidence intervals and tests for a single lognormal mean using the ideas of generalized p-values and generalized confidence intervals. The resulting procedures are easy to compute and are applicable to small samples. We have also developed similar procedures for obtaining confidence intervals and tests for the ratio (or the difference) of two lognormal means. Our work appears to be the first attempt to obtain small sample inference for the latter problem. We have also compared our test to a large sample test. The conclusion is that the large sample test is too conservative or too liberal, even for large samples, whereas the test based on the generalized p-value controls type I error quite satisfactorily. The large sample test can also be biased, i.e., its power can fall below type I error probability. Examples are given in order to illustrate our results. In particular, using an example, it is pointed out that simply comparing the means of the logged data in two samples can produce a different conclusion, as opposed to comparing the means of the original data.

Keywords: Coverage probability; Generalized confidence interval; Generalized p-value; Type I error probability

* Corresponding author. Tel.: +1-337-482-6702; fax: +1-337-482-5346.
E-mail address: krishna@louisiana.edu (K. Krishnamoorthy).

0378-3758/03/$ - see front matter
PII: S0378-3758(02)00153-2

1. Introduction

Random variables that are inherently positive occur in many real life applications. The suitability of the lognormal distribution has been validated for several such applications; in particular, for analyzing biological data (Koch, 1966), and for analyzing data on workplace exposure to contaminants (Oldham, 1953; Esmen and Hammad, 1977; Rappaport and Selvin, 1987; Selvin and Rappaport, 1989; Lyles and Kupper, 1996). Let X be a random variable having a lognormal distribution, and let μ and σ^2, respectively, denote the mean and variance of $\ln(X)$ so that $Y = \ln(X) \sim N(\mu, \sigma^2)$. Many of the parameters of interest concerning the lognormal distribution (for example the mean of X) turn out to be functions of both μ and σ^2 and it appears difficult to obtain exact and/or optimum tests and confidence intervals. In particular, the mean of the lognormal distribution is given by

$$E(X) = E(\exp(Y)) = \exp(\eta), \quad \text{where } \eta = \mu + \frac{\sigma^2}{2}. \tag{1.1}$$

Clearly, the computation of confidence intervals and test procedures concerning the mean of X is equivalent to the computation of the corresponding quantities for η. The problem of obtaining confidence intervals and tests concerning η has been addressed by Land (1971, 1972, 1973, 1975, 1988) in a series of articles. However, the tests and confidence intervals derived by Land are conditional (i.e., based on a certain conditional distribution) and this makes the procedure somewhat difficult to use in practice. In fact, concerning Land's procedures, Lyles and Kupper (1996, p. 9) comment that "... this method is apparently not used extensively by environmental scientists. This lack of use may be due to the fact that extensive tables required for the procedure are somewhat obscure..." A similar remark concerning the difficulties associated with Land's method is also pointed out in Zhou and Gao (1997, p. 784). It should however be noted that some of the tables required for implementing Land's procedure is reproduced in Gibbons and Coleman (2001), who have also summarized Land's procedure; see Gibbons and Coleman (2001, Chapter 19). The table values depend on the observed sample standard deviation of the logged sample data; the table cannot be used if this observed value is different from one of the tabulated values.

Some simple procedures for obtaining confidence intervals for a single lognormal mean are reviewed and compared in Zhou and Gao (1997). These include a large sample method due to Cox, reported in Land (1972), a conservative method due to Angus (1988), and a parametric bootstrap method, also due to Angus (1994). The numerical results in Zhou and Gao (1997) show that in terms of coverage probability, all of these procedures are too conservative or too liberal, unless the sample size is big, in which case, the procedure due to Cox is satisfactory. Thus, for obtaining confidence intervals or hypotheses tests for a single lognormal mean, satisfactory small sample procedures are unavailable except the results due to Land mentioned above. The problem of testing the equality of two lognormal means is investigated in Zhou et al. (1997), and an approach similar to the Cox procedure is recommended for large samples. For this problem, no small sample results are available so far. Yet, in many applications where lognormal data come up, small samples are quite common and small sample inference is called for. In fact, in the context of analyzing occupational exposure data using the lognormal distribution, Lyles et al. (1997, p. 69) mention that "personal exposure monitoring is relatively time consuming and costly, so typical samples will seldom be large in a statistical sense".

Perhaps an obvious question to ask is why it is of interest to do inference on the mean of a skewed distribution such as the lognormal. It may appear more meaningful to deal with the median rather than the mean. Furthermore, consideration of the lognormal median has the added advantage that the inference problem will reduce to that concerning a normal mean. However, there are applications that specifically require inference concerning the mean of a lognormal distribution—most notably in the context of analyzing data on occupational exposure to contaminants. The need to have statistical inference on the lognormal mean, and the lack of simple and easy to use procedures for the same, are clearly pointed out in the literature on occupational exposure; see, for example, Rappaport and Selvin (1987), Spear and Selvin (1989), and Lyles and Kupper (1996). In particular, regarding exposure data analysis using the lognormal distribution, Rappaport and Selvin (1987, p. 375) states: "Unfortunately, the median exposure has no physiological significance without additional information concerning the variance of the exposure distribution. Estimation procedures for evaluating the mean exposure per se have defined limit values not to be exceeded by either a single measured exposure or the estimated mean of a series of exposures". A major motivation for the present work is the need for easily computable tests and confidence regions for the lognormal mean, as required in the analysis of occupational exposure data. Other applications that require inferences on a single lognormal mean and the comparison of two lognormal means are discussed in Zhou and Gao (1997) and Zhou et al. (1997).

The first goal of this article is to come up with exact tests and confidence intervals for η in (1.1) using the novel concepts of generalized p-values and generalized confidence intervals. In particular, we obtain test procedures and confidence intervals applicable to small samples. The generalized p-value has been introduced by Tsui and Weerahandi (1989) and the generalized confidence interval by Weerahandi (1993); see the book by Weerahandi (1995a) for a detailed discussion along with numerous examples. Weerahandi (1995a, p. 109) has in fact mentioned the applicability of the generalized p-value for dealing with parameters of the type η in (1.1). The concepts of generalized p-values and generalized confidence intervals have turned out to be extremely fruitful for obtaining tests and confidence intervals involving "non-standard" parameters, such as η in (1.1). Several articles have appeared in the literature describing such applications; see Weerahandi and Johnson (1992), Zhou and Mathew (1994), Weerahandi (1995b), Weerahandi and Berger (1999), and Krishnamoorthy and Mathew (2002). In the next section we discuss the problem of testing hypotheses and computing confidence intervals for η in (1.1), using generalized p-values and generalized confidence intervals. For two independent lognormal distributions with means $\exp(\eta_1)$ and $\exp(\eta_2)$, we have also considered the problem of testing hypotheses and computing confidence intervals for $\eta_1 - \eta_2$. This is described in Section 3. Such a problem can come up when we want to compare occupational exposure data (or pollution measurements) at two different sites. The problem also comes up in the context of comparing the average medical costs for African American patients and White patients with type I diabetes, described in Zhou et al. (1997). These authors have noted that such data follow the lognormal distribution. Note that inference concerning $\eta_1 - \eta_2$ is equivalent to that for the ratio of two lognormal means, namely, $\exp(\eta_1)/\exp(\eta_2)$. Furthermore, we have also provided a procedure for constructing confidence limits for the difference between the lognormal means, namely, $\exp(\eta_1) - \exp(\eta_2)$.

For each of the problems that we have considered, we have also explained the required computational procedures. Furthermore, in Section 4, we have illustrated our procedures with two examples that involve the comparison of two lognormal means. One of the applications is on the comparison of the means of two sets of carbon monoxide emission measurements, and the problem is to check if an oil refinery is overestimating the carbon monoxide emissions. A second example is on the comparison of the amount of rainfall from clouds seeded with silver nitrate and from unseeded clouds. For this example, comparison of the means from the logged data (which are normally distributed) gave a conclusion that was different from the comparison of the lognormal means.

2. Tests and confidence intervals for a lognormal mean

Suppose the random variable X follows the lognormal distribution so that $Y = \ln(X) \sim N(\mu, \sigma^2)$. Then the mean of X is as defined in (1.1). Consider the problem of testing

$$\mathrm{H}_0\colon \eta \leqslant \eta_0 \quad \text{vs.} \quad \mathrm{H}_1\colon \eta > \eta_0, \tag{2.1}$$

where $\eta = \mu + \sigma^2/2$ and η_0 is a specified quantity. Let $X_1, X_2, \ldots, X_n$ be a random sample from the lognormal distribution, and let $Y_i = \ln(X_i)$, $i = 1, 2, \ldots, n$. We shall develop a test for the hypotheses in (2.1) and a confidence interval for η based on the sufficient statistics

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2. \tag{2.2}$$

We shall also denote by $\bar{y}$ and s^2, the observed values of $\bar{Y}$ and S^2, respectively. In order to define a p-value (referred to as the *generalized p-value*) for testing the hypotheses in (2.1), we shall first define a *generalized test variable* T_1 that is a function of the random variables $\bar{Y}$ and S^2, their observed values $\bar{y}$ and s^2, where T_1 could also depend on the unknown parameters. However, T_1 is required to satisfy the following conditions:

(a) The distribution of T_1 is stochastically monotone in η.

(b) The observed value of T_1 is free of any unknown parameters.

(c) At $\eta = \eta_0$, the distribution of T_1 is free of any unknown parameters. (2.3)

Let t_1 denote the observed value of a generalized test variable T_1 satisfying the three conditions in (2.3). If T_1 is stochastically increasing in η, the generalized p-value for testing the hypotheses in (2.1) is defined as $P(T_1 \geqslant t_1|\eta=\eta_0)$. On the other hand, if T_1 is stochastically decreasing in η, the generalized p-value is defined as $P(T_1 \leqslant t_1|\eta=\eta_0)$.

2.1. A test for (2.1) and a confidence interval for η

We shall now define a generalized test variable satisfying the conditions in (2.3). Let

$$T_1 = \bar{y} - \frac{\bar{Y}-\mu}{S/\sqrt{n}} s/\sqrt{n} + \frac{1}{2}\frac{\sigma^2}{S^2}s^2 - \eta$$

$$= \bar{y} - \frac{Z}{U/\sqrt{n-1}}\frac{s}{\sqrt{n}} + \frac{1}{2}\frac{s^2}{U^2/(n-1)} - \eta, \qquad (2.4)$$

where $Z = \sqrt{n}(\bar{Y}-\mu)/\sigma \sim N(0,1)$ independently of $U^2 = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Here χ^2_r denotes the central chisquare distribution with df $= r$. The observed value of T_1 is obtained by replacing $\bar{Y}$ and S^2 by $\bar{y}$ and s^2, respectively, in the first expression in (2.4), and this observed value is zero. It is clear that T_1 satisfies the conditions in (2.3) and the distribution of T_1 is stochastically decreasing in η. The generalized p-value for testing the hypotheses in (2.1) is thus given by $P(T_1 \leqslant 0|\eta=\eta_0)$. The test based on the generalized p-value rejects H_0 if the generalized p-value is less than some specified level α (say, $\alpha = 0.05$). It should however be noted that type I error probability and the power of such a test may depend on unknown parameters. Consequently, it is necessary to simulate type I error probability in order to see whether the test controls type I error.

In order to obtain an upper confidence interval for η, let

$$T_2 = \bar{y} - \frac{\bar{Y}-\mu}{S/\sqrt{n}} s/\sqrt{n} + \frac{1}{2}\frac{\sigma^2}{S^2}s^2$$

$$= \bar{y} - \frac{Z}{U/\sqrt{n-1}}\frac{s}{\sqrt{n}} + \frac{1}{2}\frac{s^2}{U^2/(n-1)}, \qquad (2.5)$$

where the various quantities in (2.5) are as defined before for T_1. Notice that T_2 reduces to η when $\bar{Y} = \bar{y}$ and $S^2 = s^2$, and the distribution of T_2 is free of any unknown parameters. If $T_2(1-\alpha)$ denotes the $100(1-\alpha)$th percentile of T_2, then $T_2(1-\alpha)$ is the $100(1-\alpha)\%$ generalized upper confidence interval for η. Once again, the actual coverage probability of this interval may not be $1-\alpha$; the coverage could depend on unknown parameters, and it is necessary to simulate the coverage probability to study the behavior of the confidence interval. Such simulation results are reported in Section 2.2. A $100(1-\alpha)\%$ generalized lower confidence interval for η can be similarly obtained as $T_2(\alpha)$. A two-sided $100(1-\alpha)\%$ generalized confidence interval for η is given by $(T_2(\alpha/2), T_2(1-\alpha/2))$. We recall that, unlike Land's results (see Land 1975, pp. 386–387), our pivot variable T_2 can be used for constructing both one-sided and equi-tailed two-sided confidence intervals.

Note that with T_1 and T_2 as defined in (2.4) and(2.5), respectively, $T_1 = T_2 - \eta$ and the generalized p-value for testing (2.1) is given by $P(T_2 \leqslant \eta_0)$. It is also easy to verify that for testing the hypotheses in (2.1), the test based on the generalized

p-value is equivalent to a test procedure based on the generalized lower confidence limit $T_2(\alpha)$. Rejecting H_0 when the generalized p-value is less than α is easily seen to be equivalent to rejecting H_0 when $T_2(\alpha) > \eta_0$.

Both the generalized p-value and the generalized confidence interval can be computed using the following algorithm.

Algorithm 1.
For a given data set $x_1,\ldots,x_n$, set $y_i = \ln(x_i)$, $i = 1,\ldots,n$.
Compute $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $s^2 = (1/(n-1))\sum_{i=1}^{n}(y_i - \bar{y})^2$
For $i = 1$ to m
Generate $Z \sim N(0,1)$ and $U^2 \sim \chi^2_{n-1}$
Set $T_{2i} = \bar{y} - (Z/(U/\sqrt{n-1}))s/\sqrt{n} + \frac{1}{2}s^2/U^2/(n-1)$
(end i loop)
Let $K_i = 1$ if $T_{2i} \leqslant \eta_0$, else $K_i = 0$
$(1/m)\sum_{i=1}^{m} K_i$ is a Monte Carlo estimate of the generalized p-value for testing (2.1)
The $100(1-\alpha)$th percentile of $T_{21},\ldots,T_{2m}$, denoted by $T_2(1-\alpha)$, is a Monte Carlo estimate of the $100(1-\alpha)\%$ generalized upper confidence limit for $\eta = \mu + \sigma^2/2$.

The $100(1-\alpha)$th percentile of T_2 can also be obtained using a numerical integration and a root finding method as shown below. Noting that Z is distributed as $-Z$, and rearranging the terms in (2.5), we write

$$T_2 = \bar{y} + s\sqrt{\frac{n-1}{n}}\left(\frac{Z}{U} + \frac{s\sqrt{n(n-1)}}{2U^2}\right).$$

Let $c_{1-\alpha}$ denote the $100(1-\alpha)$th percentile of $(Z/U + s\sqrt{n(n-1)}/2U^2)$. Using the fact that $Z \sim N(0,1)$ independently of $U^2 \sim \chi^2_{n-1}$, it can be easily shown that $c_{1-\alpha}$ is the root of the equation

$$\int_0^\infty \Phi\left(\left(\sqrt{x}c_{1-\alpha} - \frac{s\sqrt{n(n-1)}}{2\sqrt{x}}\right)\right) f(x; n-1)\,\mathrm{d}x = 1 - \alpha, \tag{2.6}$$

where $f(x; n-1)$ is the probability density function of the chi-square distribution with df $= n-1$. The $100(1-\alpha)\%$ upper limit on the basis of $c_{1-\alpha}$ is given by

$$\bar{y} + c_{1-\alpha}s\sqrt{\frac{n-1}{n}}. \tag{2.7}$$

Although the confidence limit (2.7) may be more accurate than the one based on Algorithm 1, evaluating $c_{1-\alpha}$ is numerically involved. Algorithm 1 is simple to use, and the confidence limit (2.7) and the one based on Algorithm 1 with 100,000 runs are very close. For instance, when $\bar{y} = 1.0$, $s = 0.5$ and $n = 3$, the 95% upper limit using (2.7) is 3.718, and Algorithm 1 yielded 3.729; for the same values of $\bar{y}$ and s, and $n = 5$ both yielded 1.948; when $n = 10$ both produced the same limit of 1.517. Therefore, for practical use we recommend Algorithm 1 to construct confidence limits for $\eta = \mu + \sigma^2/2$.

2.2. Numerical results on the coverage probability

In order to understand the performance of the generalized p-value and the generalized confidence limits, we estimated the coverage probabilities of the generalized confidence interval as follows:

Algorithm 2. For specified values of n, μ, σ and $0 < \alpha < 1$:
For $i = 1,\ m_1$
  Generate $\bar{y}$ from $N(\mu, \sigma^2/n)$
  Generate Q from χ^2_{n-1}, and set $s^2 = \sigma^2 Q/(n-1)$
For $j = 1,\ m_2$
  Generate $Z \sim N(0,1)$ and $U^2 \sim \chi^2_{n-1}$
  Set $T_{2j} = \bar{y} - (Z/(U/\sqrt{n-1}))s/\sqrt{n} + \frac{1}{2}s^2/U^2/(n-1)$
(end j loop)
  If the $100(1-\alpha)$th percentile $T_2(1-\alpha)$ of $\{T_{21}, \ldots, T_{2m_2}\}$ is greater than $\eta = \mu + \sigma^2/2$, set $K_i = 1$; else set $K_i = 0$
(end i loop)
  $(1/m_1)\sum_i^{m_1} K_i$ is an estimate of the coverage probability of the generalized upper confidence limit.

For an accurate interval estimation procedure, estimated coverage probabilities should be equal to the nominal level $1-\alpha$. We estimated the coverage probabilities of the one-sided upper limits of $\mu + \sigma^2/2$ using the above method with $m_1 = m_2 = 10{,}000$. We used the IMSL subroutine RNCHI to generate chi-square random numbers and the function subroutine RNNOF to generate normal random numbers. Following Algorithm 2, the coverage probabilities were computed for the parameter values $\mu = 1$, $\sigma = 0.1$, 0.5, 2, 5, 10, $n = 3$, 10, 20, and $1-\alpha = 0.90$, 0.95 and 0.99. The estimated coverage probabilities coincided with the nominal levels in all the cases considered for the simulation (for this reason table values are not reported here). Consequently, we also conclude that for testing the hypotheses in (2.1), the estimated type I error probabilities of the test based on the generalized p-value will coincide with the corresponding nominal significance level, at least for the above parameter combinations considered for the simulation.

2.3. Comparison with parametric bootstrap and Land's (1973) procedure

We compared our generalized confidence limits with those of Land (1973) for various values of n and s, and with the confidence limits obtained by the parametric bootstrap method, described in Angus (1994). Land's procedure can be described as follows. Define

$$T = \frac{\sqrt{n}(\bar{Y} - \eta)}{S} \quad \text{and} \quad V = [(n-1)S^2 + n(\bar{Y} - \eta)^2]^{1/2}.$$

Let $f(t|v)$ denote the conditional density of T given $V = v$, and let $t(1-\alpha; \eta, v)$ denote the $100(1-\alpha)$th percentile of $f(t|v)$. Land's $100(1-\alpha)\%$ upper limit for η is

Table 1
Upper limits for $\eta = \mu + \sigma^2/2$ based on (a) Algorithm 1, (b) Land's formula (2.8), and (c) parametric bootstrap by Angus (1994); $\bar{y} = 1$

n	s	95% limits			99% limits		
		(a)	(b)	(c)	(a)	(b)	(c)
3	0.1	1.226	1.199	1.184	1.731	1.594	1.431
3	0.5	3.724	3.421	2.329	13.831	13.436	4.052
3	5	244.25	244.69	164.44	1242.41	1244.57	446.52
11	0.1	1.062	1.062	1.061	1.093	1.093	1.091
11	1	2.499	2.448	2.367	3.247	3.194	2.902
11	10	128.11	128.10	127.22	196.57	196.69	193.76
21	0.1	1.044	1.044	1.043	1.062	1.062	1.062
21	0.5	1.355	1.347	1.344	1.476	1.468	1.456
21	2	4.889	4.852	4.769	6.113	6.068	5.809
21	10	93.39	93.33	93.26	122.51	122.29	121.80
101	0.5	1.218	1.217	1.216	1.259	1.258	1.258
101	5	17.145	17.139	17.130	18.988	18.975	18.960
101	10	65.273	65.260	65.227	72.518	72.500	72.440
501	5	14.963	14.964	14.510	15.621	15.623	15.617
501	10	56.704	56.711	56.690	59.286	59.291	59.301
1001	5	14.510	14.510	14.510	14.951	14.954	14.953
1001	10	54.940	54.940	59.94	56.672	56.673	56.622

given by

$$\bar{y} - t(1-\alpha;\eta,v)\,\frac{s}{\sqrt{n}}. \tag{2.8}$$

Note that for computing this interval, one needs $t(1-\alpha;\eta,v)$. We used the table values given in Land (1975) for obtaining $t(1-\alpha;\eta,v)$. The resulting confidence interval (2.8), along with our generalized confidence interval, are given in Table 1.

From the numerical results in Table 1, it should be clear that our generalized confidence limit and the confidence limit obtained by Land's (1973) method practically coincide. The limits based on the parametric bootstrap (PB) method due to Angus (1994) are smaller than the other two limits whenever the sample sizes are small and s is large. Simulation studies due to Angus (1994, Table 1) also exhibit similar property. This observation indicates, contrary to the numerical studies by Zhou and Gao (1997), that the confidence limits based on the PB method could be liberal. To confirm this, we estimated the coverage probabilities of our generalized confidence limits and the PB confidence limits for various values of n, σ^2 and $\mu = -\sigma^2/2$. The coverage probabilities of the PB method are estimated using the algorithm given in Angus (1994). We used 5000 samples from $N(\mu,\sigma^2)$ distribution and 2000 bootstrap samples to estimate the coverage probabilities. The estimated coverage probabilities in Table 2 clearly indicate that the PB intervals are liberal when the sample sizes are small and the σ^2's are large, whereas the coverage probabilities of the generalized limits are always close to the nominal level. Thus, the computational simplicity of our approach along with the fact that its coverage probability coincides with the nominal

Table 2
Coverage probabilities of two-sided limits for $\eta=\mu+\sigma^2/2$ based on (a) Algorithm 1, and (c) PB by Angus (1994); $\mu=-\sigma^2/2$

Nominal level		90%		95%	
n	σ^2	(a)	(c)	(a)	(c)
5	0.5	0.895	0.889	0.947	0.946
	1	0.880	0.867	0.949	0.934
	5	0.893	0.857	0.948	0.896
	20	0.897	0.864	0.948	0.911
10	0.500	0.895	0.889	0.952	0.943
	1	0.897	0.886	0.950	0.942
	5	0.899	0.879	0.950	0.932
	20	0.901	0.888	0.949	0.933
15	0.5	0.899	0.891	0.948	0.942
	1	0.896	0.884	0.947	0.943
	5	0.892	0.880	0.951	0.928
	20	0.901	0.893	0.950	0.945
25	0.5	0.904	0.903	0.950	0.948
	1	0.895	0.897	0.946	0.942
	5	0.900	0.895	0.947	0.939
	20	0.909	0.902	0.949	0.947
	100	0.897	0.897	0.947	0.946

level (at least for the parameter combinations considered for simulation) make the generalized confidence interval and the generalized p-value attractive options for inference concerning the mean of a lognormal distribution.

3. Comparing the means of two lognormal distributions

Let X_1 and X_2 be two independent lognormal random variables, so that $Y_1=\ln(X_1)\sim N(\mu_1,\sigma_1^2)$ and $Y_2=\ln(X_2)\sim N(\mu_2,\sigma_2^2)$. Let

$$\eta_1=\mu_1+\frac{\sigma_1^2}{2} \quad \text{and} \quad \eta_2=\mu_2+\frac{\sigma_2^2}{2} \tag{3.1}$$

so that $E(X_1)=\exp(\eta_1)$ and $E(X_2)=\exp(\eta_2)$. Thus, for comparing the two lognormal means, the problem reduces to inference concerning the difference $\eta_1-\eta_2$. We shall now address the problem of testing hypotheses and deriving confidence intervals for the difference $\eta_1-\eta_2$.

3.1. Hypothesis testing

Let X_{1i}, $i=1,2,\ldots,n_1$, and X_{2i}, $i=1,2,\ldots,n_2$, denote random samples from the lognormal distributions of X_1 and X_2, respectively. Let $Y_{1i}=\ln(X_{1i})$, $i=1,2,\ldots,n_1$, and $Y_{2i} = \ln(X_{2i})$, $i = 1, 2, \ldots, n_2$. Define

$$\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \quad \text{and} \quad S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2, \ \ i = 1, 2. \tag{3.2}$$

Furthermore, let $\bar{y}_1$, $\bar{y}_2$, s_1^2 and s_2^2 denote the observed values of $\bar{Y}_1$, $\bar{Y}_2$, S_1^2 and S_2^2, respectively. Consider the problem of testing

$$\mathrm{H}_0\colon \eta_1 \leqslant \eta_2 \quad \text{vs.} \quad \mathrm{H}_1\colon \eta_1 > \eta_2. \tag{3.3}$$

Let

$$\begin{aligned} T_{3i} &= \bar{y}_i - \frac{\bar{Y}_i - \mu_i}{S_i/\sqrt{n_i}} s_i/\sqrt{n_i} + \frac{1}{2}\frac{\sigma_i^2}{S_i^2} s_i^2 \\ &= \bar{y}_i - \frac{Z_i}{U_i/\sqrt{n_i - 1}}\frac{s_i}{\sqrt{n_i}} + \frac{1}{2}\frac{s_i^2}{U_i^2/(n_i - 1)}, \quad i = 1, 2, \end{aligned} \tag{3.4}$$

where $Z_i = \sqrt{n_i}(\bar{Y}_i - \mu_i)/\sigma_i \sim N(0,1)$ and $U_i^2 = (n_i - 1)S_i^2/\sigma_i^2 \sim \chi^2_{n_i - 1}$ for $i = 1$, 2, and these random variables are also independent. Define the generalized test variable

$$T_3 = T_{31} - T_{32} - (\eta_1 - \eta_2) \tag{3.5}$$

and let

$$T_4 = T_{31} - T_{32} \tag{3.6}$$

so that $T_3 = T_4 - (\eta_1 - \eta_2)$. It is easily verified that T_3 satisfies the conditions in (2.3), with η replaced by $\eta_1 - \eta_2$ and η_0 replaced by 0. Furthermore, the distribution of T_3 is stochastically decreasing in $\eta_1 - \eta_2$. Thus the generalized p-value for testing the hypotheses in (3.3) is given by

$$P(T_3 \leqslant 0 | \eta_1 - \eta_2 = 0) = P(T_4 \leqslant 0). \tag{3.7}$$

3.2. Confidence intervals

One-sided (upper or lower) and two-sided confidence intervals for $\eta_1 - \eta_2$ can be obtained using T_4 defined above. Note that the observed value of T_4 is in fact $\eta_1 - \eta_2$. Thus the appropriate percentiles of T_4 can be used for obtaining confidence intervals for $\eta_1 - \eta_2$. The confidence limits for the difference between the lognormal means, that is, $\exp(\eta_1) - \exp(\eta_2)$, can be constructed using the percentiles of

$$T_5 = \exp(T_{31}) - \exp(T_{32}), \tag{3.8}$$

where T_{31} and T_{32} are given in (3.4).

3.3. Comparison with a large sample test: one-sided hypotheses

Zhou et al. (1997) have proposed a large sample test that can be used for testing the hypotheses in (3.3), and also for testing two-sided hypotheses. The test proposed by these authors is based on

$$Z = \frac{\bar{Y}_2 - \bar{Y}_1 + \frac{1}{2}(S_2^2 - S_1^2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2 + \frac{1}{2}(S_1^4/(n_1 - 1) + S_2^4/(n_2 - 1))}}. \tag{3.9}$$

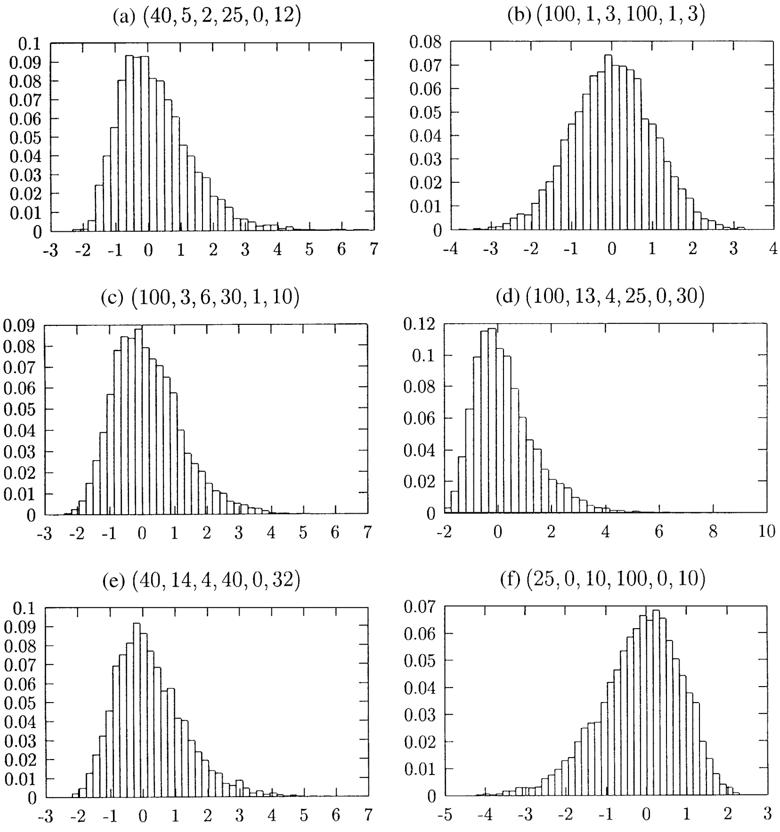

Fig. 1. Histograms of the Z-score statistic in (3.9); the numbers in parenthesis represent $(n_1, \mu_1, \sigma_1^2, n_2, \mu_2, \sigma_2^2)$.

Zhou et al. claimed that, for large samples, the distribution of the Z-score statistic in (3.9) is approximately normal under H_0 in (3.3). However, the histograms (based on 5000 simulated data) of the Z-score statistic in Fig. 1 indicate that such an approximation is valid only when both samples are large and (n_1, μ_1, σ_1^2) is approximately equal to (n_2, μ_2, σ_2^2). In other situations, the distributions of the Z-score statistic appear to be highly skewed, and hence a normal approximation is not valid.

We carried out a comparison of the Z-score test based on Z in (3.9), and the test based on the generalized p-value mentioned above using Monte Carlo simulation. The simulation study was carried out along the lines of Algorithm 2 given for the one-sample case. The numerical results in Zhou et al. (1997) are for the case of $n_1 \geqslant 25$ and $n_2 \geqslant 25$. For smaller values of n_1 and n_2, the Z-score test has type I probabilities which are either too large or too small, whereas the test based on the generalized p-value is extremely satisfactory for controlling type I error probability. Since the Z-score test is meant for large samples, we are mainly interested in studying its large sample properties. Numerical results on type I error probability and power are reported in Table 3 for various values of n_1 and n_2. For the simulations, we have taken $\mu_2=0$, without loss of generality. Various combinations of σ_1^2, σ_2^2, and μ_1 are considered in Table 3, and all the results correspond to a nominal level of 5%. Type I error probability corresponds to the parameter combinations satisfying $\mu_1 + \frac{1}{2}\sigma_1^2 = \mu_2 + \frac{1}{2}\sigma_2^2$. Rest of the numerical results in Table 3 give the power of the tests.

The numerical results show that the Z-score test is either too conservative, or too liberal, especially when the sample sizes are unequal. The pattern that emerges is the following. For H_1: $\eta_1 > \eta_2$, the Z-score test is too liberal when $n_1 > n_2$ and is too conservative when $n_1 < n_2$. In addition, our numerical results are in agreement with the histograms in Fig. 1. For instance, we observe from the histograms in Fig. 1(e) and (f), and their corresponding type I error probabilities in Table 3 (rows 24 and 30) that the Z-score test is too liberal when the histogram is right skewed and is too conservative when the histogram is left skewed. On the other hand, the test based on the generalized p-value controls type I error probability quite satisfactorily, regardless of n_1 and n_2. Also, there is a clear pattern concerning the power. In cases where the Z-score test is too liberal in terms of type I error probability, it has a larger power compared to the generalized p-value test. On the other hand, in cases where the Z-score test is too conservative, it has a smaller power compared to the generalized p-value test. We also observe from Table 3 that there are parameter configurations for which the sizes of the Z-score test exceed the nominal level of 0.05 even when the sample sizes are very large and equal. Furthermore, there are also situations where the power of the Z-score test can be smaller than the nominal level (see the last row in Table 3), indicating that the Z-score test is biased.

The overall picture that emerges from the numerical results is that the test based on the generalized p-value is extremely satisfactory, and it is applicable regardless of the sample size. The computations required to obtain the generalized p-value (or the generalized confidence interval) are simple and straightforward.

3.4. Comparison with a large sample test: two-sided hypotheses

Since the Z-score test is skewed, it is not appropriate for testing one-sided hypotheses; our numerical results in the previous section confirmed this. The problem addressed in Zhou et al. (1997) is that of testing the two sided hypotheses

$$H_0\colon \eta_1 = \eta_2 \quad \text{vs.} \quad H_1\colon \eta_1 \neq \eta_2. \tag{3.10}$$

For testing (3.10), Zhou et al. (1997) recommend the large sample test based on (3.9).

The two-sided generalized confidence interval mentioned above can also be used to test the hypotheses in (3.10), and the test procedure consists of rejecting H_0 when the generalized confidence interval for $\eta_1 - \eta_2$ does not contain zero. A generalized p-value for testing (3.10) can be developed as follows. Note that the null hypothesis in (3.10) will be rejected if a small p-value is obtained for testing against at least one of the alternatives: H_1: $\eta_1 > \eta_2$ or H_1: $\eta_1 < \eta_2$. In other words, we reject the null hypothesis in (3.10) if at least one of the generalized p-values, $P(T_4 \leqslant 0)$ or $P(T_4 \geqslant 0)$ is small, where T_4 is given in (3.6). Hence, for testing against the two-sided alternative

Table 3
Sizes and powers of the generalized p-value test (GP test) and the Z-score test at 5% significance level when $\mu_2 = 0$ and H_1: $\eta_1 > \eta_2$

n_1	n_2	μ_1	σ_1^2	σ_2^2	Size		n_1	n_2	μ_1	σ_1^2	σ_2^2	Power	
					GP test	Z-score						GP test	Z-score
4	4	1	2	4	0.0471	0.0963	4	4	0	12	4	0.1492	0.0376
		0	3	3	0.0412	0.0433			0	20	4	0.2910	0.0387
		5	2	12	0.0401	0.2067			3	2	4	0.1050	0.3726
		0	12	12	0.0400	0.0171			4	1	1	0.6220	0.9734
10	10	1	2	4	0.0442	0.0882	10	10	0	12	4	0.3880	0.2386
		0	3	3	0.0489	0.0397			0	20	4	0.6583	0.4370
		5	2	12	0.0510	0.1428			3	2	4	0.2743	0.5119
		0	12	12	0.0433	0.0286			4	1	1	0.9981	0.9992
25	25	0	1	1	0.0491	0.0530	25	25	1	1	1	0.8521	0.8880
		0	5	5	0.0479	0.0438			1	5	5	0.1751	0.2123
		0	10	10	0.0500	0.0420			1	10	10	0.1264	0.1150
		0	100	100	0.0495	0.0363			0	4	2	0.3754	0.3207
		2	4	8	0.0510	0.0789			0	9	7	0.1555	0.1079
		4	8	16	0.0511	0.0821			0	4	1	0.7641	0.6768
40	25	0	1	1	0.0490	0.0591	40	25	1	1	1	0.8590	0.9238
		0	5	5	0.0460	0.0606			1	5	5	0.2053	0.2650
		0	10	10	0.0510	0.0610			1	10	10	0.1071	0.1565
25	40	0	1	1	0.0461	0.0446	25	40	1	1	1	0.9466	0.9444
		0	5	5	0.0490	0.0302			1	5	5	0.2303	0.2177
		0	10	10	0.0511	0.0280			1	10	10	0.1190	0.0996
40	25	5	2	12	0.0487	0.1084	40	25	1	5	4	0.3954	0.4789
25	40	5	2	12	0.0534	0.0963			1	10	9	0.1482	0.2201
40	40	8	4	20	0.0496	0.0959	25	40	1	5	4	0.4829	0.4181
		14	4	32	0.0512	0.1002			1	10	9	0.2131	0.1580
100	25	0	1	1	0.0480	0.0711	100	25	1	1	1	0.9133	0.9531
		0	5	5	0.0470	0.0770			1	5	5	0.2438	0.3353
		0	10	10	0.0530	0.0860			1	10	10	0.1374	0.2045
25	100	0	1	1	0.0511	0.0341	25	100	1	1	1	0.9912	0.9851
		0	5	5	0.0472	0.0197			1	5	5	0.3401	0.2221
		0	10	10	0.0520	0.0130			1	10	10	0.1638	0.0788
100	25	13	4	30	0.0501	0.1221	100	25	0	2	1	0.3936	0.4744
25	100	13	4	30	0.0541	0.0780			0	3	1	0.8259	0.8535
100	100	13	4	30	0.0491	0.0810	25	100	0	2	1	0.3925	0.2906
									0	3	1	0.6933	0.5525
							100	25	1	5	4	0.4592	0.5785
									1	10	9	0.1765	0.2857
							25	100	1	5	4	0.5684	0.4930
									1	10	9	0.2627	0.1404

(continued on next page)

Table 3 (*continued*)

n_1	n_2	μ_1	σ_1^2	σ_2^2	Size		n_1	n_2	μ_1	σ_1^2	σ_2^2	Power	
					GP test	Z-score						GP test	Z-score
200	50	0	1	1	0.0481	0.0633	50	200	1	1	1	1.0000	1.0000
200	50	0	5	5	0.0512	0.0712	50	200	1	5	5	0.4925	0.4305
200	50	0	10	10	0.0530	0.0698	50	200	1	10	10	0.2125	0.1576
50	200	0	1	1	0.0517	0.0399	200	50	1	1	1	0.9982	0.9986
50	200	0	5	5	0.0487	0.0293	200	50	1	5	5	0.3751	0.4646
50	200	0	10	10	0.0482	0.0250	200	50	1	10	10	0.1823	0.2409
200	200	24	12	60	0.0471	0.0691	200	50	0	2	1	0.6814	0.7164
200	200	40	2	82	0.0498	0.0750	200	50	0	3	1	0.9805	0.9848
							50	200	0	2	1	0.6010	0.5086
							50	200	0	3	1	0.8940	0.8589
							25	300	0.8	12	12	0.1251	0.0357

in (3.10), the p-value can be computed as

$$2 \times \min\{P(T_4 \leqslant 0), P(T_4 \geqslant 0)\}. \tag{3.11}$$

Note that (3.11) is essentially in the spirit of the computation of the usual p-value for testing against two-sided alternatives; see Gibbons and Pratt (1975), or Pratt and Gibbons (1981, Section 4.5), for a discussion of the computation of such p-values. In fact, the computation in (3.11) follows one of the recommendation by these authors.

Table 4 gives numerical results on type I error probability and power of the generalized p-value test and the Z-score test for testing the hypotheses in (3.10), where the generalized p-value is computed following (3.11). In terms of controlling type I error probability, the large sample Z-score test now performs much more satisfactorily, compared to the testing of the one-sided hypotheses. However, there are still situations where the Z-score test is unsatisfactory. The pattern seems to be that the Z-score test performs well in terms type I probabilities when $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$. In fact the parameter configurations in the numerical results reported in Zhou et al. (1997) have the μ's close to each other and the σ^2's close to each other, and good performance was noted in terms type I error probabilities. Note that the generalized p-value test continues to provide satisfactory performance in terms of type I error. Regarding the power, the pattern that we noticed in Table 3 continues to hold in Table 4 as well.

4. Illustrative examples

Example 4.1. (Source: http://lib.stat.cmu.edu/DASL/) An oil refinery located at the northeast of San Francisco conducted a series of 31 daily measurements of the carbon monoxide levels arising from one of their stacks between April 16 and May 16, 1993. The measurements were submitted as evidence for establishing a baseline to the Bay Area Air Quality Management District (BAAQMD). BAAQMD personnel also made 9 independent measurements of the carbon monoxide concentration from the same stack over the period from September 11, 1990–March 30, 1993. As mentioned in the “data

Table 4
Sizes and Powers of the generalized p-value test (GP test) and the Z-score test at 5% significance level when $\mu_2 = 0$ and H_1: $\eta_1 \neq \eta_2$

n_1	n_2	μ_1	σ_1^2	σ_2^2	Size		n_1	n_2	μ_1	σ_1^2	σ_2^2	Power	
					GP test	Z-score						GP test	Z-score
4	4	1	2	4	0.0293	0.0720	4	4	0	12	4	0.0820	0.0301
		0	3	3	0.0387	0.0454			0	20	4	0.1580	0.0298
		5	2	12	0.0379	0.1773			3	2	4	0.0750	0.2930
		0	12	12	0.0299	0.0175			4	1	1	0.4060	0.9466
10	10	1	2	4	0.0415	0.0581	10	10	0	12	4	0.2810	0.0861
		0	3	3	0.0417	0.0409			0	20	4	0.5560	0.1601
		5	2	12	0.0473	0.1110			3	2	4	0.1690	0.4209
		0	12	12	0.0419	0.0148			4	1	1	0.9890	0.9986
25	25	0	1	1	0.0415	0.0478	25	25	1	1	1	0.7290	0.8117
		0	5	5	0.0414	0.0386			1	5	5	0.1280	0.1340
		0	10	10	0.0474	0.0333			1	10	10	0.0740	0.0620
		0	100	100	0.0487	0.0257			0	4	2	0.2730	0.1968
		2	4	8	0.0503	0.0532			0	9	7	0.0800	0.0602
		4	8	16	0.0441	0.0518			0	4	1	0.6710	0.5263
40	25	0	1	1	0.0410	0.0520	40	25	1	1	1	0.7960	0.8642
		0	5	5	0.0413	0.0422			1	5	5	0.1090	0.1907
		0	10	10	0.0511	0.0401			1	10	10	0.0600	0.0899
25	40	0	1	1	0.0423	0.0512	25	40	1	1	1	0.8570	0.8993
		0	5	5	0.0496	0.0459			1	5	5	0.1540	0.1332
		0	10	10	0.0476	0.0402			1	10	10	0.0750	0.0526
40	25	5	2	12	0.0467	0.0746	40	25	1	5	4	0.2620	0.3637
25	40	5	2	12	0.0480	0.0658			1	10	9	0.1010	0.1375
40	40	8	4	20	0.0474	0.0664	25	40	1	5	4	0.3470	0.2875
		14	4	32	0.0483	0.0696			1	10	9	0.1170	0.0838
100	25	0	1	1	0.0471	0.0563	100	25	1	1	1	0.7980	0.9207
		0	5	5	0.0467	0.0564			1	5	5	0.1270	0.2518
		0	10	10	0.0430	0.0575			1	10	10	0.0670	0.1426
25	100	0	1	1	0.0510	0.0566	25	100	1	1	1	0.9840	0.9669
		0	5	5	0.0410	0.0580			1	5	5	0.2050	0.1228
		0	10	10	0.0492	0.0600			1	10	10	0.0970	0.0404
100	25	13	4	30	0.0470	0.0823	100	25	0	2	1	0.2720	0.3619
25	100	13	4	30	0.0480	0.0557			0	3	1	0.6760	0.7678
100	100	13	4	30	0.0460	0.0592	25	100	0	2	1	0.3040	0.1580
									0	3	1	0.5950	0.3869
							100	25	1	5	4	0.3200	0.4631
									1	10	9	0.1060	0.2198
							25	100	1	5	4	0.4990	0.3230
									1	10	9	0.1760	0.0708

(*continued on next page*)

Table 4 (*continued*)

n_1	n_2	μ_1	σ_1^2	σ_2^2	Size		n_1	n_2	μ_1	σ_1^2	σ_2^2	Power	
					GP test	Z-score						GP test	Z-score
200	50	0	1	1	0.0532	0.0550	50	200	1	1	1	100	0.9999
200	50	0	5	5	0.0458	0.0544	50	200	1	5	5	0.4110	0.2913
200	50	0	10	10	0.0491	0.0556	50	200	1	10	10	0.1390	0.0847
50	200	0	1	1	0.0446	0.0514	200	50	1	1	1	0.9810	0.9967
50	200	0	5	5	0.0501	0.0518	200	50	1	5	5	0.2190	0.3654
50	200	0	10	10	0.0521	0.0547	200	50	1	10	10	0.1140	0.1725
200	200	24	12	60	0.0419	0.0493	200	50	0	2	1	0.5290	0.6074
200	200	40	2	82	0.0478	0.0532	200	50	0	3	1	0.9520	0.9645
							50	200	0	2	1	0.4730	0.3720
							50	200	0	3	1	0.8340	0.7638
							25	300	0.80	12	12	0.0880	0.0363

& story library" at the above web address, in this case, the refinery had an incentive to overestimate carbon monoxide emissions (to setup a baseline at a higher level), and the purpose of our analysis is to test this. The data are given below:

Carbon monoxide measurements by the refinery (in ppm): 45, 30, 38, 42, 63, 43, 102, 86, 99, 63, 58, 34, 37, 55, 58, 153, 75, 58, 36, 59, 43, 102, 52, 30, 21, 40, 141, 85, 161, 86, 161, 86, 71.

Carbon monoxide measurements by the BAAQMD (in ppm): 12.5, 20, 4, 20, 25, 170, 15, 20, 15.

We checked the assumption of lognormality using MINITAB, and found that a lognormal model adequately describes both sets of measurements but normal model does not fit the data at any practical levels of significance. The hypotheses to be tested are

$$\mathrm{H}_0\colon \eta_1 \leqslant \eta_2 \quad \text{vs.} \quad \mathrm{H}_1\colon \eta_1 > \eta_2,$$

where $\exp(\eta_1) = \exp(\mu_1 + \sigma_1^2/2)$ and $\exp(\eta_2) = \exp(\mu_2 + \sigma_2^2/2)$ denote the population mean of the refinery measurements and that of the BAAQMD measurements, respectively.

For logged measurements taken by the refinery, we have: $n_1 = 31$, sample mean$= \bar{y}_1 =$ 4.0743 and the sample SD $= s_1 = 0.5021$; for logged measurements collected by the BAAQMD, $n_2 = 9$, sample mean$= \bar{y}_2 = 2.963$ and sample SD $= s_2 = 0.974$. The 95% generalized lower confidence limit for $(\mu_1 + \sigma_1^2/2) - (\mu_2 + \sigma_2^2/2)$ is -0.40. Hence the 95% generalized lower confidence limit for the ratio of the lognormal population means (that is, for $\exp(\eta_1)/\exp(\eta_2)$) is $\exp(-0.40) = 0.67$, which is < 1. Using (3.7), we also computed the 95% lower limit for $\exp(\eta_1) - \exp(\eta_2)$ as -32.91. Finally, we computed the generalized p-value using (3.7) as 0.112. All these results lead to the conclusion that the data do not provide sufficient evidence to indicate that the mean measurement by the refinery is greater than that of BAAQMD, which is contrary to the speculation mentioned in the Data & Story Library.

Example 4.2. (Source: http://lib.stat.cmu.edu/DASL/) The data on the amount of rainfall (in acre-feet) from 52 clouds, 26 of which were chosen at random and seeded with silver nitrate are given in the aforementioned website. Probability plots indicate that normal models do not fit the data whereas lognormal models fit the data sets very well. The summary statistics for the logged data are as follows: For seeded clouds: $n_1 = 26$, $\bar{y}_1 = 5.134$, and $s_1 = 1.600$; for unseeded clouds: $n_2 = 26$, $\bar{y}_2 = 3.990$, and $s_2 = 1.642$. In order to understand the effect of silver nitrate seeding, we like to test H_0: $\eta_1 = \eta_2$ vs. H_1: $\eta_1 > \eta_2$. We computed the 95% generalized lower confidence limit for $\eta_1 - \eta_2$ as -0.20. This lower limit indicates that the data do not provide sufficient evidence to indicate that the mean rainfall from seeded clouds is higher than the mean rainfall from unseeded clouds. The generalized p-value turned out to be 0.078. On the other hand, application of the two-sample t test for the logged data (note that two-sample t test merely compares the means of the logged data, not of the original data) yields a p-value of 0.007, indicating $\mu_1 > \mu_2$, where μ_i's denote the means of logged data.

We also applied the Z-score test for the same problem. The 95% lower limit for $\eta_1 - \eta_2$ based on the Z-score test is -0.06203. The corresponding p-value is 0.060. The conclusion based on the Z-score test is consistent with the conclusion based on the generalized p-value test.

5. Concluding remarks

In this article, we have derived exact inference procedures (hypotheses tests and confidence intervals) concerning the mean of a single lognormal distribution, and for the ratio of the means of two independent lognormal distributions. The procedures are applicable to small samples, and are easy to compute and implement. The available procedures due to Land (1973) for deriving tests and confidence intervals concerning a single lognormal mean, being based on a conditional distribution, appear to be computationally more involved than the procedure given in this paper. The procedures described in Zhou and Gao (1997) and Zhou et al. (1997) are mostly applicable to large samples. The methods we have developed based on the concepts of the generalized p-value and generalized confidence interval are relatively easy to implement, and numerical results show that they give equivalent results compared to Land's (1973) procedure. Regarding the problem of hypotheses tests and confidence intervals concerning the ratio of two lognormal means and the difference between two lognormal means, this article appears to be the first to provide satisfactory solutions, applicable to small samples. The solutions are once again based on the concepts of the generalized p-value and generalized confidence interval. Numerical results show that our procedures based on the generalized p-value and generalized confidence interval are much more satisfactory compared to large sample procedures. Furthermore, one of our numerical examples show that if we compare the means of the logged data, as opposed to the means of the original data, the conclusions can be drastically different. Our procedures and conclusions should be of considerable interest to occupational hygienists and biologists, since lognormal data are commonly encountered in their applications.

Acknowledgements

The authors are grateful to two referees for their valuable comments and suggestions. This research has been supported by grant R01-0H03628-01A1 from the National Institute of Occupational Safety and Health (NIOSH).

References

Angus, J.E., 1988. Inferences on the lognormal mean for complete samples. Comm. Statist. Simulation Comput. 17, 1307–1331.

Angus, J.E., 1994. Bootstrap one-sided confidence intervals for the lognormal mean. Statistician 43, 395–401.

Esmen, N., Hammad, Y., 1977. Lognormality of environmental sampling data. Environ. Sci. Health A 12, 29–41.

Gibbons, R.D., Coleman, D.E., 2001. Statistical Methods for Detection and Quantification of Environmental Contamination. Wiley, New York.

Gibbons, J.D., Pratt, J.W., 1975. P-values: interpretation and methodology. Am. Statist. 29, 20–25.

Koch, A.L., 1966. The logarithm in biology. J. Theoret. Biol. 12, 276–290.

Krishnamoorthy, K., Mathew, T., 2002. Assessing occupational exposure via the one-way random effects model with balanced data. J. Agri. Biol. Environ. Statist., to appear.

Land, C.E., 1971. Confidence intervals for the linear function of the normal mean and variance. Ann. Math. Statist. 42, 1187–1205.

Land, C.E., 1972. An evaluation of approximate confidence interval estimation methods for lognormal means. Technometrics 14, 145–158.

Land, C.E., 1973. Standard confidence limits for linear functions of the normal mean and variance. J. Am. Statist. Assoc. 68, 960–963.

Land, C.E., 1975. Tables of confidence limits for linear functions of the normal mean and variance. In: Harter, H.L., Owen, D.B. (Eds.), Selected Tables in Math. Statist., Vol. III, American Mathematical Society, Providence, R.I., USA, pp. 385–419.

Land, C.E., 1988. Hypotheses tests and interval estimates. In: Crow, E.L., Shimizu, K. (Eds.), Lognormal Distribution. Marcel Dekker, New York, pp. 87–112.

Lyles, R.H., Kupper, L.L., 1996. On strategies for comparing occupational exposure data to limits. Am. Indust. Hyg. Assoc. J. 57, 6–15.

Lyles, R.H., Kupper, L.L., Rappaport, S.M., 1997. Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA model. J. Agric. Biol. Environ. Statist. 2, 64–86.

Oldham, P., 1953. The nature of the variability of dust concentrations at the coal face. British J. Indust. Med. 10, 227–234.

Pratt, J.W., Gibbons, J.D., 1981. Concepts of Nonparametric Theory. Springer, New York.

Rappaport, S.M., Selvin, S., 1987. A method for evaluating the mean exposure from a lognormal distribution. Am. Indust. Hyg. J. 48, 374–379.

Selvin, S., Rappaport, S.M., 1989. Note on the estimation of the mean value from a lognormal distribution. Am. Indust. Hyg. Assoc. J. 50, 627–630.

Tsui, K.W., Weerahandi, S., 1989. Generalized p-values in siginificance testing of hypotheses in the presence of nuisance parameters. J. Am. Statist. Assoc. 84, 602–607.

Weerahandi, S., 1993. Generalized confidence intervals. J. Am. Statist. Assoc. 88, 899–905.

Weerahandi, S., 1995a. Exact Statistical Methods for Data Analysis. Springer, New York.

Weerahandi, S., 1995b. ANOVA under unequal error variances. Biometrics 51, 589–599.

Weerahandi, S., Berger, V.W., 1999. Exact inference for growth curves with intraclass correlation structure. Biometrics 55, 921–924.

Weerahandi, S., Johnson, R.A., 1992. Testing reliability in a stress-strength model when x and y are normally distributed. Technometrics 34, 83–91.

Zhou, X.H., Gao, S., 1997. Confidence intervals for the lognormal mean. Statist. Med. 16, 783–790.

Zhou, L., Mathew, T., 1994. Some tests for variance components using generalized *p*-values. Technometrics 36, 394–402.

Zhou, X.H., Gao, S., Hui, S.L., 1997. Methods for comparing the means of two independent lognormal samples. Biometrics 53, 1129–1135.