STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:2125–2135
Published online 11 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2329

Interval estimation of the mean response in a log-regression model

Jianrong Wu[1,*,†], A. C. M. Wong[2] and Wei Wei[3]

[1]*Department of Biostatistics, St Jude Children's Research Hospital, 332 North Lauderdale St., Memphis, TN 38105, U.S.A.*
[2]*SASIT, Atkinson Faculty of Liberal and Professional Studies, York University, 4700 Keele St., North York, Ontario, Canada M3J 1P3*
[3]*Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.*

SUMMARY

A standard approach to the analysis of skewed response data with concomitant information is to use a log-transformation to normalize the distribution of the response variable and then conduct a log-regression analysis. However, the mean response at original scale is often of interest. El-Shaarawi and Viveros developed an interval estimation of the mean response of a log-regression model based on large sample theory. There is however very little information available in the literature on constructing such estimates when the sample size is small. In this paper, we develop a small-sample corrected interval by using the likelihood-based inference method developed by Barndorff-Nielson and Fraser *et al*. Simulation results show that the proposed interval provides almost exact coverage probability, even for small samples. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: confidence interval; coverage probability; log-regression model; mean response; r^*-formula

1. INTRODUCTION

Skewed response data with concomitant information, are often obtained in diverse forms of medical, ecological, and econometrics research. For example, with patients having acute myelogenous leukaemia (AML), a short period of survival is associated with over-proliferation of white blood cells (WBC) [1]. Another example concerns contaminated concentration with supplemented measurements on environmental factors in ecology [2]. A standard approach to

*Correspondence to: Jianrong Wu, Department of Biostatistics, St Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, TN 38105, U.S.A.
†E-mail: jianrong.wu@stjude.org

Contract/grant sponsor: National Cancer Institute; contract/grant number: CA21765
Contract/grant sponsor: American Lebanese Syrian Associated Charities (ALSAC)
Contract/grant sponsor: Natural Science and Engineering Council

Received 6 December 2004
Accepted 6 June 2005

the analysis of skewed data is to use a log-transformation to normalize the distribution of the response variable and then conduct a log-regression analysis. Often however, the mean response on the original scale is of interest. For example, the U.S. Environmental Protection Agency requires that risks be characterized in terms of the mean concentration of the contaminant [3]. This requirement is partly responsible for the notable emphasis in environmental literature on inferences about the mean concentration of the contaminant.

Finney [4] developed a minimum variance unbiased estimation of parameters for the log-normal distribution, and Bradu and Mundlak [5] extended those results to the log-normal regression model. El-Shaarawi and Viveros [2] developed a large-sample interval estimation of the mean response of a log-regression model. There is however very little literature available on how to construct the interval estimate for the mean response under a log-regression model when such data arise from a small-sample; for example, a small number $(n=17)$ of patients with AML and high WBC count [1] or a small data set $(n=10)$ on the annual production and market prices of ground nuts and cotton in Israel from 1954 to 1963 [5].

The primary goal of this paper is to develop a small-sample corrected interval estimate of the mean response at a specific value of the concomitant variates. The paper is organized as follows: Section 2 describes the large-sample interval estimate developed by El-Shaarawi and Viveros [2]. A small-sample corrected interval estimate is also developed in Section 2 by using the likelihood-based inference method developed by Barndorff-Nielson [6, 7] and Fraser *et al.* [8]. Simulation results are reported in Section 3. In Section 4, the proposed intervals are applied using two small-sample examples and one moderate-sample example. Conclusions are given in Section 5.

2. INTERVAL ESTIMATIONS

Let $(T_i, \mathbf{z}_i')$ be the given experimental data, where T_i is the response variable measured at the ith set of p concomitant variates $\mathbf{z}_i = (z_{1i}, \ldots, z_{pi})'$. By taking a logarithmic transformation of the response variable T, a general regression model of $\log(T)$ on $\mathbf{z}$ is given of

$$y_i = \log(T_i) = \alpha + \mathbf{z}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where α and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ are the regression coefficients, and $\varepsilon_i, i = 1, \ldots, n$ are i.i.d. from a normal distribution with mean zero and variance σ^2. Note that (1) is generally referred to as the log-regression model. The mean response T at a specific value of the concomitant variate $\mathbf{z} = \mathbf{z}_0$ is given by

$$\eta = E(T \mid \mathbf{z} = \mathbf{z}_0) = \mathrm{e}^{\alpha + \mathbf{z}_0'\boldsymbol{\beta} + \sigma^2/2}$$

2.1. Z-interval

Based on the large-sample normal approximation to the uniformly minimum variance unbiased estimator, El-Shaarawi and Viveros [2] derived an $100(1-\alpha)$ per cent confidence interval for $\psi = \log(\eta) = \alpha + \mathbf{z}_0'\boldsymbol{\beta} + \frac{\sigma^2}{2}$

$$\tilde{\alpha} + \mathbf{z}_0'\tilde{\boldsymbol{\beta}} + \frac{\tilde{\sigma}^2}{2} \pm z_{\alpha/2}\tilde{\sigma}\left\{v + \frac{\tilde{\sigma}^2}{2(n-p-1)}\right\}^{1/2}$$

In this interval, $(\tilde{\alpha}, \tilde{\boldsymbol{\beta}}')' = (X'X)^{-1}X'\mathbf{y}$ and $\tilde{\sigma}^2 = \mathbf{y}'[I - X(X'X)^{-1}X']\mathbf{y}/(n-p-1)$, where $\mathbf{y} = (y_1, \ldots, y_n)'$ is the $n \times 1$ vector of the log responses, $X = (\mathbf{1}, Z)$ is the $n \times (p+1)$ design matrix, and $v = (1, \mathbf{z}_0')(X'X)^{-1}(1, \mathbf{z}_0')'$. This interval is referred to as the Z-interval. Our simulation results (Section 3) show that this Z-interval is liberal in terms of low coverage probability, especially in the small-sample size setting.

2.2. *r-interval and r*-interval*

Suppose the joint log-likelihood function based on sample data is $\ell(\boldsymbol{\theta}) = \ell(\psi, \boldsymbol{\lambda})$, where ψ is a scalar parameter of interest, and $\boldsymbol{\lambda}$ is a nuisance vector parameter. One can construct an approximate confidence interval for ψ based on the signed log-likelihood ratio

$$r \equiv r(\psi) = \operatorname{sgn}(\hat{\psi} - \psi)\{2[\ell(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell(\psi, \hat{\boldsymbol{\lambda}}_\psi)]\}^{1/2} \tag{2}$$

where $(\hat{\psi}, \hat{\boldsymbol{\lambda}}')'$ denotes the maximum likelihood estimator of $(\psi, \boldsymbol{\lambda}')'$, and $\hat{\boldsymbol{\lambda}}_\psi$ denotes the maximum likelihood estimator of $\boldsymbol{\lambda}$ for a fixed ψ. The approximate $100(1-\alpha)$ per cent confidence interval for ψ can be obtained as

$$\{\psi;\ |r(\psi)| < z_{\alpha/2}\}$$

which is referred to as the r-interval. However, the signed log-likelihood ratio r is asymptotically distributed as a standard normal variate with first-order accuracy [9]. We will show that neither the Z-interval nor the r-interval have good coverage probabilities in the small-sample setting.

The asymptotic normality of r can be improved by certain adjustments. In this paper, we consider the modified signed log-likelihood ratio r^* introduced by Barndorff-Nielsen [6, 7] and Barndorff-Nielsen and Cox [10]. This ratio is generally known as the r^*-formula

$$r^* \equiv r^*(\psi) = r(\psi) + r(\psi)^{-1} \log\left\{\frac{u(\psi)}{r(\psi)}\right\} \tag{3}$$

The general form of $u(\psi)$ is given in Appendix. Barndorff-Nielsen [6, 7] showed that r^* is asymptotically distributed as a standard normal variate with third-order accuracy. Therefore, an approximate $100(1-\alpha)$ per cent confidence interval for ψ based on r^* is given by

$$\{\psi; |r^*(\psi)| < z_{\alpha/2}\}$$

which is referred to as the r^*-interval. This r^*-interval, unlike the Z-interval or the r-interval, calculates the confidence limits from the observed asymmetric likelihood-based function $r^*(\psi)$ and achieves a more accurate coverage probability and symmetric upper- and lower-error probabilities, in a small-sample size setting.

For the log-regression model (1), the log-likelihood function of $\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda}')' = (\psi, \boldsymbol{\beta}', \sigma)'$ is given by

$$\ell(\theta) = -n \log \sigma - \frac{1}{2\sigma^2} s_{p+1} + \frac{1}{\sigma^2}\left(\psi - \frac{\sigma^2}{2}\right) s_0 - \frac{1}{\sigma^2}\mathbf{s}'\boldsymbol{\beta} - \frac{1}{2\sigma^2}\sum\left(\psi - \frac{\sigma^2}{2} - \mathbf{d}_i'\boldsymbol{\beta}\right)^2$$

where $\psi = \alpha + \mathbf{z}_0{}'\boldsymbol{\beta} + \frac{\sigma^2}{2}$; $\mathbf{d}_i = \mathbf{z}_i - \mathbf{z}_0 = (z_{1i} - z_{10}, \ldots, z_{pi} - z_{p0})'$. The minimal sufficient statistic $\mathbf{t} = (s_0, \mathbf{s}', s_{p+1})'$ is given by the variables $s_0 = \sum y_i$, $s_{p+1} = \sum y_i^2$ and $\mathbf{s} = (s_1, \ldots, s_p)' = (\sum d_{1i} y_i, \ldots, \sum d_{pi} y_i)'$.

The maximum likelihood estimators are

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}')' = (X'X)^{-1}X'\mathbf{y}$$

$$\hat{\sigma}^2 = \frac{1}{n}\mathbf{y}'[I - X(X'X)^{-1}X']\mathbf{y} \quad \text{and}$$

$$\hat{\psi} = \hat{\alpha} + \mathbf{z}_0'\hat{\boldsymbol{\beta}} + \frac{\hat{\sigma}^2}{2}$$

The constrained maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 are the solutions of the following recursive equations:

$$\hat{\boldsymbol{\beta}}_{\psi} = \left(\sum \mathbf{d}_i\mathbf{d}_i'\right)^{-1}\left\{\mathbf{s} - \left(\psi - \frac{\hat{\sigma}_{\psi}^2}{2}\right)\sum \mathbf{d}_i\right\} \quad \text{and}$$

$$\hat{\sigma}_{\psi}^2 = 2\left\{1 + \frac{1}{n}(s_{p+1} + 2\psi \sum \mathbf{d}_i'\hat{\boldsymbol{\beta}}_{\psi} + \sum(\mathbf{d}_i'\hat{\boldsymbol{\beta}}_{\psi})^2 + n\psi^2 - 2\psi s_0 - 2\mathbf{s}'\hat{\boldsymbol{\beta}}_{\psi})\right\}^{1/2} - 2$$

The signed log-likelihood ratio $r(\psi)$ can be calculated from (2), and $r^*(\psi)$ can be calculated from (3). We can therefore construct a confidence interval for ψ based on $r(\psi)$ or $r^*(\psi)$. Let (ψ_L, ψ_U) be a $100(1-\alpha)$ per cent confidence interval for ψ, then (e^{ψ_L}, e^{ψ_U}) is the corresponding $100(1-\alpha)$ per cent confidence interval for the mean response η.

In general, there is no explicit analytic interval available based on r and r^*, but a simple numerical iterative procedure is developed to obtain the upper- and lower-bound limits [11].

3. SIMULATION STUDIES

In this section, we carry out simulation studies to compare the performance of the Z-interval, r-interval, and r^*-interval for constructing 90 and 95 per cent two-sided confidence intervals for ψ in small- or moderate-sample size settings. The performance of a method is judged using the criteria addressed in Reference [11], such as the coverage probability, coverage error, upper- and lower-error probabilities, average bias, and average length. The desired values of these criteria for the confidence intervals (90 per cent, 95 per cent) are as follows: coverage probability (0.9, 0.95), coverage error (0, 0), upper- and lower-error probabilities (0.05, 0.025), and average bias (0, 0). These values reflect the desired properties of the coverage probability, accuracy, and symmetry of the upper- and lower-error probabilities. It is recommended that the average length not be used as a major judgment criterion.

The first simulation involves a log-regression model with a single concomitant variate. The sample size considered is $n = 11$. The parameter configurations are $\alpha = 6$ and $\beta = -1$, and σ ranges from 0.1 to 2. The concomitant variate $\mathbf{z} = \log(3, 5, 10, 30, 40, 50, 60, 80, 100, 120, 160)'$, and $z_0 = \log(70)$. For each parameter configuration, we generated 10 000 random samples (as the log responses) from the normal distribution with the mean equal to $\alpha + \beta z$ and variance σ^2

Table I. Evaluation of performance criteria for $Z-$, $r-$, and r^*-intervals for constructing a two-sided 90 per cent confidence interval in a small-sample (n=11) setting with one concomitant variate.

		Performance criteria					
σ	Interval	Coverage probability	Coverage error	Upper error probability	Lower error probability	Average length	Average bias
0.1	Z	0.8643	0.0357	0.0723	0.0634	0.1108	0.0178
	r	0.8477	0.0523	0.0804	0.0719	0.1065	0.0262
	r^*	0.8966	0.0034	0.0515	0.0519	0.1223	0.0017
0.5	Z	0.8686	0.0314	0.0861	0.0453	0.5866	0.0204
	r	0.8506	0.0494	0.0923	0.0571	0.5615	0.0247
	r^*	0.8980	0.0020	0.0520	0.0500	0.6652	0.0010
1.0	Z	0.8656	0.0344	0.1083	0.0261	1.3556	0.0411
	r	0.8509	0.0491	0.1098	0.0393	1.2756	0.0352
	r^*	0.8998	0.0002	0.0562	0.0440	1.6185	0.0061
1.5	Z	0.8657	0.0343	0.1232	0.0111	2.4170	0.0560
	r	0.8406	0.0594	0.1251	0.0343	2.2494	0.0454
	r^*	0.8957	0.0043	0.0547	0.0496	3.0316	0.0026
2.0	Z	0.8599	0.0401	0.1351	0.0050	3.8213	0.0650
	r	0.8407	0.0593	0.1339	0.0254	3.5090	0.0542
	r^*	0.8985	0.0015	0.0550	0.0465	4.9245	0.0042

to construct the 90 per cent confidence intervals. The simulated coverage probabilities, upper- and lower-error probabilities, and average biases and lengths for each interval are given in Table I.

From Table I, we observe that the performance of the Z-interval is slightly better than that of the r-interval. However, the coverage probability is low, the coverage errors and average biases are large, and the error probabilities are quite asymmetric, particularly for cases in which the variance is large. In contrast, the coverage probability of the small-sample corrected r^*-interval is comparable in all cases; its coverage errors are near zero; and it has the smallest average bias for the three intervals in all the cases studied. In addition, the upper- and lower-error probabilities of the r^*-interval are much more symmetric and accurate than those of the Z- or r-intervals. Overall, the small-sample corrected r^*-interval performs much better than the large-sample based Z-interval or r-interval.

The second simulation involves a log-regression model with two concomitant variates. The sample size considered is $n=10$. The parameter configurations are $\alpha=3.5901, \beta_1=2.8405$, $\beta_2=-0.3553$, and $\sigma=0.311155$. The two concomitant variates z_1 and z_2 are defined as in Example 2, Section 4 and $\mathbf{z}_0=(-0.443, 10)'$. For each parameter configuration, we generated 10 000 random samples from the normal distribution with the mean equal to $\alpha+\beta_1 z_1+\beta_2 z_2$ and variance σ^2. The simulated coverage probabilities, upper- and lower-error probabilities, and average biases and lengths for each interval are given in Table II.

This simulation also showed that the r^*-interval performs much better than the Z-interval or r-interval, in terms of coverage probability and symmetry of the error probabilities.

Table II. Evaluation of performance criteria for $Z-$, $r-$, and r^*-intervals for constructing a two-sided 95 per cent confidence interval in a small-sample ($n=10$) setting with two concomitant variate.

	Performance criteria					
Interval	Coverage probability	Coverage error	Upper error probability	Lower error probability	Average length	Average bias
Z	0.9063	0.0437	0.0562	0.0375	0.7954	0.0218
r	0.8827	0.0673	0.0680	0.0493	0.7330	0.0336
r^*	0.9442	0.0058	0.0291	0.0267	0.9388	0.0029

Table III. Evaluation of performance criteria for $Z-$, $r-$, and r^*-intervals for constructing a two-sided 95 per cent confidence interval in a moderate-sample ($n=31$) setting with two concomitant variate.

	Performance criteria					
Interval	Coverage probability	Coverage error	Upper error probability	Lower error probability	Average length	Average bias
Z	0.9424	0.0076	0.0305	0.0271	0.1281	0.0038
r	0.9375	0.0125	0.0331	0.0294	0.1256	0.0062
r^*	0.9523	0.0023	0.0246	0.0231	0.1337	0.0012

The third simulation is also a log-regression model with two concomitant variates but with a moderate sample size $n=31$. The parameter configurations are $\alpha=-6.620, \beta_1=1.976$, $\beta_2=1.119$, and $\sigma=0.077667$. The two concomitant variates z_1 and z_2 are defined as in Example 3, Section 4 and $\mathbf{z}_0=\log(20.6, 87)'$. For each parameter configuration, we generated 10 000 random samples from the normal distribution with the mean equal to $\alpha+\beta_1 z_1+\beta_2 z_2$ and variance σ^2. The simulated coverage probabilities, upper- and lower-error probabilities, average biases, and average lengths for each method are given in Table III.

With a moderate sample size, all three methods showed similar coverage probability and upper- and lower-error probabilities. However, the r^*-interval still outperformed the Z-interval or the r-interval.

4. EXAMPLES

The first real data that we will use as an example is the survival time (T, weeks) from diagnosis of 17 patients with AML [1]. Leukaemia is characterized by an over-proliferation of white blood cells; the higher the WBC count, the more severe the disease, and the lower the probability of survival.

Figure 1(a) is the scatter plot of the survival time and WBC count. The linear trend of the scatter plot between logarithm of survival time (y) and log WBC count (z) (Figure 1(b)) suggests a log-regression model between survival time and log(WBC count). The maximum likelihood estimators of α and β are $\hat{\alpha}=11.07456$ and $\hat{\beta}=-0.8178$, and these give the predicted model

$$\hat{y}=11.07456-0.8178z$$

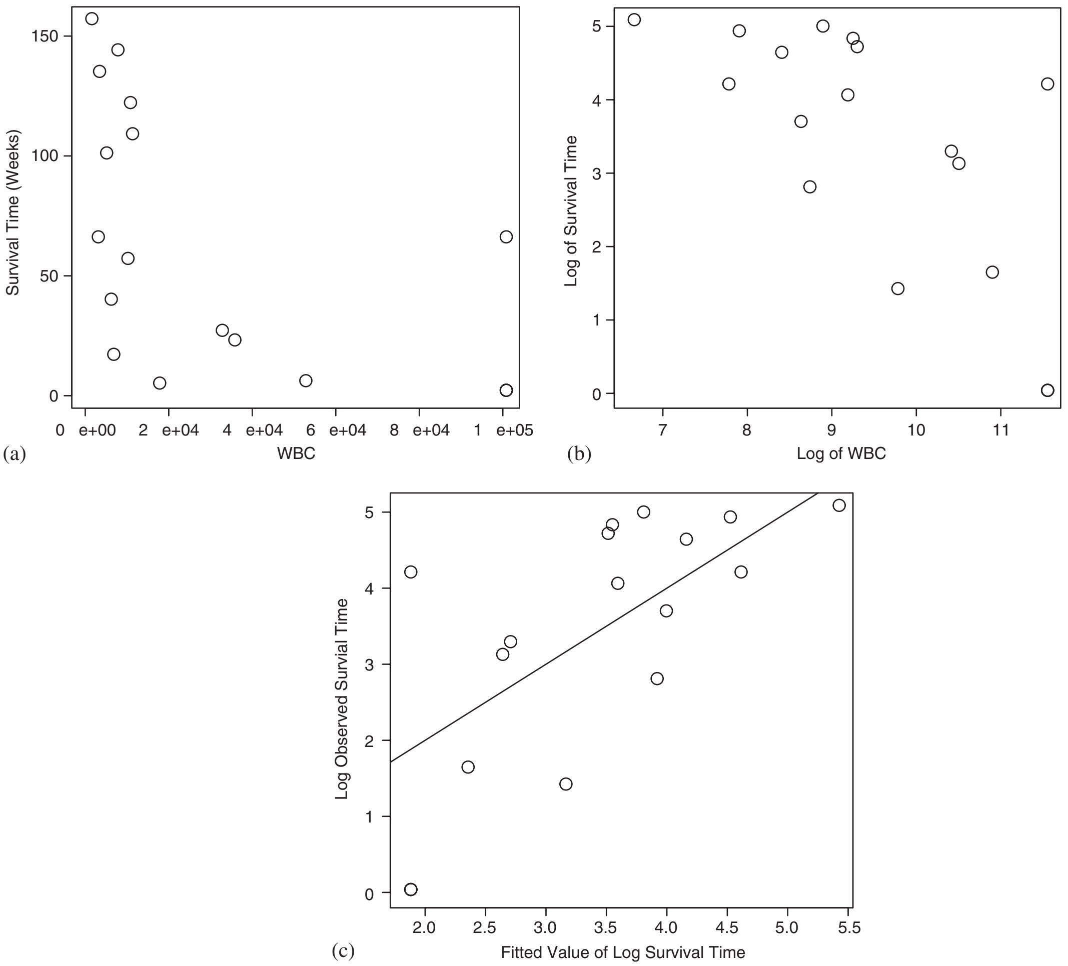

Figure 1. (a) Scatter plot of the survival time and WBC count; (b) scatter plot between logarithm of survival time and log WBC count; and (c) scatter plot of the observed values versus their fitted values.

As a partial check of the adequacy of the fitted model, a Shapiro–Wilk test of normality of the residuals ($y - \hat{y}$) gives a p-value of 0.52, which supports the log-normal distribution assumption. The scatter plot of the observed y values versus their fitted values ($\hat{y}$) (Figure 1(c)) lies along the identity line and suggests a reasonable model fit. The predicted value of y at the median value of z is 4.213, which gives a predicted mean survival time of 67.55 weeks at a median WBC count. The 95 per cent confidence intervals for ψ and e^{ψ} at the median WBC count are given in Table IV.

All three intervals yield similar lower bounds but quite different upper bounds. This is due to the fact that r^* corrects the skewness of the distribution, whereas Z and r do not.

The second example is a small data set on the annual production and market price of ground nuts and cotton in Israel from 1952 to 1963 [5]. Specifically, response variable (T)

Table IV. 95 per cent confidence intervals of Example 1.

Interval	ψ		$\eta = e^{\psi}$	
Z	3.497	5.107	33.03	165.17
r	3.611	5.206	36.99	182.29
r^*	3.642	5.539	38.17	254.54

Table V. 95 per cent confidence intervals for Example 2.

Year	z_{10}	z_{20}	Z-interval	r-interval	r^*-interval
1963	−0.443	10	(0.193, 0.518)	(0.201, 0.500)	(0.185, 0.595)
1964	−0.250	11	(0.175, 0.838)	(0.187, 0.790)	(0.160, 1.008)
1965	−0.308	12	(0.106, 0.492)	(0.112, 0.465)	(0.097, 0.591)
1966	0.003	13	(0.109, 1.368)	(0.121, 1.244)	(0.092, 1.788)

and concomitant variates $\mathbf{z} = (z_1, z_2)'$ are

$$T = \frac{\text{output of ground nuts}}{\text{output of cotton}} \text{ in year t}$$

$$z_1 = \log\left(\frac{\text{price of ground nuts}}{\text{price of cotton}}\right) \text{ in year t} - 2$$

$$z_2 = \text{time trend}, \quad 1954 = 1$$

The observed values of T can be readily calculated from the raw data, and following the author's methods, we fitted a log-regression model to their 10 observations for the period 1954–1963:

$$y_i = \alpha + \beta_1 z_1 + \beta_2 z_2 + \varepsilon_i, \quad i = 1, \ldots, 10$$

and developed the prediction model

$$\hat{y} = 0.35901 + 2.8405 z_1 - 0.3553 z_2$$

The confidence intervals for the average value of T for each year of the 4-year period 1963–1966 are reported in Table V.

Again, all three intervals yielded a similar lower bound, but a quite different upper bound, seemingly because the r^* corrects the skewness of the distribution. Furthermore, both the Z- and r-intervals are too short to provide 95 per cent coverage probability. This finding is confirmed by the simulation results in the previous section.

The final example of real data is a set of measurements taken on 31 black cherry trees [12, p. 287]. For each sample unit, three measurements are given:

D is the diameter (inches) of the tree measured at a given height from the ground;

Table VI. 95 per cent confidence intervals for Example 3.

Interval	ψ		$\eta = \mathrm{e}^{\psi}$	
Z	4.2885	4.4244	72.859	83.464
r	4.2899	4.4232	72.961	83.360
r^*	4.2856	4.4278	72.649	83.750

H is the height (feet) of the tree; and
V is the volume (cubic feet) of timber.

The scatter plots of the (D, V) and (H, V) pairs suggest that a plausible relationship among the variables is

$$V = \beta_0 D^{\beta_1} H^{\beta_2}$$

After taking logarithms of all variables, we formulated the linear-regression model

$$y = \alpha + \beta_1 z_1 + \beta_2 z_2 + \varepsilon$$

where $y = \log(V), \alpha = \log(\beta_0), z_1 = \log(D)$, and $z_2 = \log(H)$. The following fitted model was derived from the data:

$$\hat{y} = -6.620 + 1.976 z_1 + 1.119 z_2$$

The 95 per cent confidence intervals of the mean volume at Diameter = 20.6 and Height = 87 are shown in Table VI.

Because of the moderate sample size, all three intervals are almost identical, which is consistent with the simulation in the previous section.

5. CONCLUSIONS

In this paper, we have proposed a small-sample corrected r^*-interval for the mean response of a log-regression model. The simulation studies showed that the proposed r^*-interval is uniformly better than that proposed by El-Shaarawi and Viveros [2] and it displays the almost exact coverage probability, even for small samples.

APPENDIX

Suppose that the log-likelihood function $\ell(\boldsymbol{\theta}; \mathbf{y})$ can be rewritten as $\ell(\boldsymbol{\theta}; \mathbf{t})$, where $\mathbf{t}$ is a minimum sufficient statistic with the same dimension as $\boldsymbol{\theta}$. Then the $u(\psi)$ in the r^*-formula has the following form [8]:

$$u(\psi) = \frac{|\ell_{;\mathbf{t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell_{;\mathbf{t}}(\psi, \hat{\boldsymbol{\lambda}}_{\psi})\ \ell_{\boldsymbol{\lambda};\mathbf{t}}(\psi, \hat{\boldsymbol{\lambda}}_{\psi})|}{|\ell_{\boldsymbol{\theta};\mathbf{t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})|} \left\{ \frac{|j(\hat{\psi}, \hat{\boldsymbol{\lambda}})|}{|j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_{\psi})|} \right\}^{1/2}$$

where the sample space derivatives are defined as

$$\ell_{;\mathbf{t}}(\boldsymbol{\theta}) = \frac{\partial}{\partial \mathbf{t}}\ell(\boldsymbol{\theta};\mathbf{t}) \quad \text{and} \quad \ell_{\boldsymbol{\theta};\mathbf{t}}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\ell_{;\mathbf{t}}(\boldsymbol{\theta};\mathbf{t})$$

$j(\boldsymbol{\theta})$ denotes the observed information matrices for the full parameter $\boldsymbol{\theta}$, and $j_{\boldsymbol{\lambda\lambda}}(\boldsymbol{\theta})$ denotes that of the nuisance parameter $\boldsymbol{\lambda}$. For the log-regression model (1), we have

$$\ell_{;\mathbf{t}}(\boldsymbol{\theta}) = \left\{ \frac{\left(\psi - \dfrac{\sigma^2}{2}\right)}{\sigma^2}, \frac{1}{\sigma^2}\boldsymbol{\beta}', -\frac{1}{(2\sigma^2)} \right\}'$$

$$\ell_{\boldsymbol{\theta};\mathbf{t}}(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{1}{\sigma^2} & \mathbf{0}'_{p\times 1} & -\dfrac{2\psi}{\sigma^3} \\ \mathbf{0}_{p\times 1} & \dfrac{1}{\sigma^2} I_p & -\dfrac{2}{\sigma^3}\boldsymbol{\beta} \\ 0 & \mathbf{0}'_{p\times 1} & \dfrac{1}{\sigma^3} \end{pmatrix}$$

$$j(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{n}{\sigma^2} & \dfrac{1}{\sigma^2}\sum \mathbf{d}'_i & \dfrac{2}{\sigma^3}s_0 - \dfrac{2}{\sigma^3}\sum(\psi + \mathbf{d}'_i\boldsymbol{\beta}) \\ \dfrac{1}{\sigma^2}\sum \mathbf{d}_i & \dfrac{1}{\sigma^2}\sum \mathbf{d}_i\mathbf{d}'_i & \dfrac{2}{\sigma^3}\mathbf{s} + \dfrac{2}{\sigma^3}\sum(\psi\mathbf{d}_i + \mathbf{d}_i\mathbf{d}'_i\boldsymbol{\beta}) \\ \dfrac{2}{\sigma^3}s_0 - \dfrac{2}{\sigma^3}\sum(\psi + \mathbf{d}'_i\boldsymbol{\beta}) & \dfrac{2}{\sigma^3}\mathbf{s}' + \dfrac{2}{\sigma^3}\sum(\psi\mathbf{d}'_i + \boldsymbol{\beta}'\mathbf{d}_i\mathbf{d}'_i) & j_{\sigma\sigma}(\boldsymbol{\theta}) \end{pmatrix}$$

where

$$\begin{aligned} j_{\sigma\sigma}(\boldsymbol{\theta}) = & -\frac{n}{\sigma^2} + \frac{3}{\sigma^4}s_{p+1} - \frac{6\psi}{\sigma^4}s_0 - \frac{6}{\sigma^4}\mathbf{s}'\boldsymbol{\beta} + \frac{3}{\sigma^4}\sum\left(\psi - \frac{\sigma^2}{2} - \mathbf{d}'_i\boldsymbol{\beta}\right)^2 \\ & + \frac{3}{\sigma^2}\sum\left(\psi - \frac{\sigma^2}{2} - \mathbf{d}'_i\boldsymbol{\beta}\right) + n \end{aligned}$$

Accordingly, the $u(\psi)$ can be calculated.

ACKNOWLEDGEMENTS

We gratefully acknowledge the anonymous referees whose careful reading and insightful comments improved the earlier version of this article. The first author's research was supported in part by the National Cancer Institute support Grant CA21765 and the American Lebanese Syrian Associated Charities (ALSAC). The second author's research was supported in part by the Natural Science and Engineering Council.

REFERENCES

1. Feigl P, Zelen M. Estimation of exponential survival probabilities with concomitant information. *Biometrics* 1965; **21**:826–838.
2. El-Shaarawi AH, Viveros R. Inference about the mean in log-regression with environmental applications. *Environmetrics* 1997; **8**:569–582.
3. Schmoyer RL, Beauchamp JJ, Brandt CC, Hoffman Jr FO. Difficulties with the lognormal model in mean estimation and testing. *Environmental and Ecological Statistics* 1996; **3**:81–97.
4. Finney DJ. On the distribution of variate whose logarithm is normally distributed. *Journal of the Royal Statistical Society, Series B* 1941; **7**:155–161.
5. Bradu D, Mundlak T. Estimation in lognormal linear models. *Journal of the American Statistical Association* 1970; **65**:198–211.
6. Barndorff-Nielsen OE. Inference on full and partial parameters, based on the standardized signed log-likelihood ratio. *Biometrika* 1986; **73**:307–322.
7. Barndorff-Nielsen OE. Modified signed log-likelihood ratio. *Biometrika* 1991; **78**:557–563.
8. Fraser DAS, Reid N, Wu J. A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* 1999; **86**:249–264.
9. Cox DR, Hinkley DV. *Theoretical Statistics*. Chapman & Hall: London, 1974.
10. Barndorff-Nielsen OE, Cox DR. *Inference and Asymptotics*. Chapman & Hall: London, 1994.
11. Wu J, Wong ACM, Jiang G. Likelihood-based confidence intervals for a log-normal mean. *Statistics in Medicine* 2003; **22**:1849–1860.
12. Ryan BF, Joiner BL, Edgar RS. *Minitab Handbook* (2nd edn). PWS-Kent Publishing Company: Boston, MA, 1985.