



Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Confidence interval estimation for lognormal data with application to health economics

Guang Yong Zou^{a,b,*}, Julia Taleban^a, Cindy Y. Huo^c^a Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada N6A 5C1^b Robarts Clinical Trials, Robarts Research Institute, University of Western Ontario, London, Ontario, Canada N6A 5K8^c Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada M4N 3M5

ARTICLE INFO

Article history:

Received 27 September 2008

Received in revised form 27 March 2009

Accepted 27 March 2009

Available online 2 April 2009

ABSTRACT

There has accumulated a large amount of literature on confidence interval construction involving lognormal data owing to the fact that many data in scientific inquiries may be approximated by this distribution. Procedures have usually been developed in a piecemeal fashion for a single mean, a single mean with excessive zeros, a difference between two means, and a difference between two differences (net health benefit). As an alternative, we present a general approach for all these cases that requires only confidence limits available in introductory texts. Simulation results confirm the validity of this approach. Examples arising from health economics are used to exemplify the methodology.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The lognormal distribution may be used to approximate right skewed data arising in a wide range of scientific inquiries (Limpert et al., 2001). Traditional statistical analysis of such data has usually been focused on the means of log-transformed data, resulting in inferences expressed in terms of geometric means rather than the arithmetic means. However, there are many situations, including in environmental science (Parkhurst, 1998) and in occupational health research (Rappaport and Selvin, 1987), in which arithmetic means may provide more meaningful information. Consequently, there has accumulated a relatively large amount of literature regarding statistical methods for this type of data, including Aitchison and Brown (1957) and Crow and Shimizu (1988), with more articles being added rapidly to the literature (Chen, 1994; Taylor et al., 2002; Wu et al., 2002, 2003, 2006; Gill, 2004; Tian and Wu, 2006; Shen et al., 2006; Krishnamoorthy et al., 2006; Bebu and Mathew, 2008; Fletcher, 2008).

Since many health cost data may be positively skewed (Thompson and Barber, 2000; Briggs et al., 2002), the literature dealing with the analysis of lognormal data in this context has also increased substantially. This includes procedures for a one sample mean, a difference between two independent sample means, a difference between two dependent sample means, and additional zero values for each of these cases (Zhou, 2002). Recent advances include a method based on the Edgeworth expansion (Zhou and Dinh, 2005; Dinh and Zhou, 2006). It is worthwhile to note that this approach not only fails to provide adequate coverage rates but also lacks invariance in the sense that a confidence interval for $-\theta$ differs from $(-u, -l)$ when confidence interval for θ is given by (l, u) . As a consequence, one may reach different conclusions depending on the labeling of groups in a comparative study. One could naturally suggest the bootstrap, but simulation results (Diciccio and Efron, 1996; Zhou and Dinh, 2005; Dinh and Zhou, 2006; Zou and Donner, 2008) suggest that it can fail in the case of

* Corresponding address: Robarts Clinical Trials, Robarts Research Institute, P. O. Box 5015, 100 Perth Drive, London, Ontario, Canada N6A 5K8. Tel.: +1 519 663 3400x34092; fax: +1 519 663 3807.

E-mail address: gzou@robarts.ca (G.Y. Zou).

lognormal data. A possible explanation is that the lognormal mean is a function of a normal variance and some bootstrap intervals have been shown to fail in confidence interval construction for a normal variance (Schenker, 1985).

Recently, a procedure relying on the simulation of pivotal statistics, commonly referred to as a generalized confidence interval, has generated a series of articles on lognormal data (see, e.g., Krishnamoorthy and Mathew, 2003; Tian, 2005; Chen and Zhou, 2006; Krishnamoorthy et al., 2006; Tian and Wu, 2007a,b; Bebu and Mathew, 2008).

Instead of adopting a simulation approach to each of the situations summarized above (Zhou, 2002; Chen and Zhou, 2006), we extend a simple confidence interval procedure proposed by Zou and Donner (2008) to each of these scenarios. One advantage of our procedure is that it relies only on techniques readily available in introductory texts. We also discuss confidence interval estimation for net health benefit (NHB), an alternative to incremental cost-effectiveness ratio (Stinnett and Mullahy, 1998; Willan, 2001). By assuming a value for the willingness-to-pay for a unit of effectiveness, a positive NHB indicates the treatment is cost-effective. Detailed principles for cost-effectiveness analysis in health care can be found in textbooks (e.g. Drummond et al., 2005; Willan and Briggs, 2006).

The rest of the article is structured as follows. In Section 2 we present a general approach applicable to confidence interval estimation, which will be referred to as the MOVER, standing for the method of variance estimates recovery. We then apply the MOVER to obtain confidence intervals for a single lognormal mean, a single lognormal mean with excessive zeros, a difference between two lognormal means, and the net health benefit in Section 3. In Section 4, we compare the performance of our approach to some existing methods, particularly to generalized confidence intervals, using simulation studies. We provide examples using data from previously published studies in Section 5. The article concludes with some final remarks in Section 6.

2. Confidence interval estimation by the method of variance estimates recovery

The complication in constructing a confidence interval for the lognormal mean appears to have been due to the fact that it involves two parameters, as reflected by the remark that ‘obtaining the confidence interval for the lognormal estimator is a non-trivial problem since it is a function of two transformed sample estimates’ (Briggs et al., 2005, p. 422). However, the confidence limits for each individual parameter (the normal mean and variance) are simple to obtain. Our strategy is to ‘recover’ variance estimates from these limits and then to form approximate confidence intervals for functions of the parameters, using similar arguments to those of Zou and Donner (2008) and Zou et al. (2009a).

Suppose we wish to construct a $100(1 - \alpha)\%$ two-sided confidence interval (L, U) for $\theta_1 + \theta_2$, where the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent. Using the central limit theorem, a lower limit (L) is given by

$$L = \hat{\theta}_1 + \hat{\theta}_2 - z_{\alpha/2} \sqrt{\text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2)},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. The limit L is not readily applicable because $\text{var}(\hat{\theta}_i)$ ($i = 1, 2$) is unknown.

Now, suppose that a $100(1 - \alpha)\%$ two-sided confidence interval for θ_i is given by (l_i, u_i) . Among all the plausible parameter values of θ_1 provided by (l_1, u_1) and that of θ_2 by (l_2, u_2) , we know L is in the neighborhood of $l_1 + l_2$. Inspired by the score interval approach (Bartlett, 1953), we proceed to estimate the variances needed for L at $\theta_1 + \theta_2 = l_1 + l_2$, i.e., when $\theta_1 = l_1$ and $\theta_2 = l_2$.

We have, by the central limit theorem,

$$l_i = \hat{\theta}_i - z_{\alpha/2} \sqrt{\widehat{\text{var}}(\hat{\theta}_i)},$$

which gives a variance estimate for $\hat{\theta}_i$ at $\theta_i = l_i$ of

$$\widehat{\text{var}}(\hat{\theta}_i) = (\hat{\theta}_i - l_i)^2 / z_{\alpha/2}^2.$$

Therefore, the lower limit L for $\theta_1 + \theta_2$ is given by

$$\begin{aligned} L &= \hat{\theta}_1 + \hat{\theta}_2 - z_{\alpha/2} \sqrt{\widehat{\text{var}}(\hat{\theta}_1) + \widehat{\text{var}}(\hat{\theta}_2)} \\ &= \hat{\theta}_1 + \hat{\theta}_2 - z_{\alpha/2} \sqrt{(\hat{\theta}_1 - l_1)^2 / z_{\alpha/2}^2 + (\hat{\theta}_2 - l_2)^2 / z_{\alpha/2}^2} \\ &= \hat{\theta}_1 + \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2}. \end{aligned} \quad (1)$$

Analogous steps with the notion that $u_1 + u_2$ is close to U , and the variance estimate at $\theta_i = u_i$ is

$$\widehat{\text{var}}(\hat{\theta}_i) = (u_i - \hat{\theta}_i)^2 / z_{\alpha/2}^2,$$

we obtain an upper limit U as

$$U = \hat{\theta}_1 + \hat{\theta}_2 + \sqrt{(u_1 - \hat{\theta}_1)^2 + (u_2 - \hat{\theta}_2)^2}. \quad (2)$$

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات