Journal of Statistical Planning and Inference 141 (2011) 3793–3807

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Confidence intervals for variance components in unbalanced one-way random effects model using non-normal distributions

Brent D. Burch *

Department of Mathematics and Statistics, Northern Arizona University Flagstaff, Arizona 86011-5717, USA

A R T I C L E I N F O

Article history:
Received 3 December 2010
Received in revised form
8 April 2011
Accepted 15 June 2011
Available online 25 June 2011

Keywords:
Fiducial inference
Pivotal quantity
Restricted maximum likelihood estimation

A B S T R A C T

In scenarios where the variance of a response variable can be attributed to two sources of variation, a confidence interval for a ratio of variance components gives information about the relative importance of the two sources. For example, if measurements taken from different laboratories are nine times more variable than the measurements taken from within the laboratories, then 90% of the variance in the responses is due to the variability amongst the laboratories and 10% of the variance in the responses is due to the variability within the laboratories. Assuming normally distributed sources of variation, confidence intervals for variance components are readily available. In this paper, however, simulation studies are conducted to evaluate the performance of confidence intervals under non-normal distribution assumptions. Confidence intervals based on the pivotal quantity method, fiducial inference, and the large-sample properties of the restricted maximum likelihood (REML) estimator are considered. Simulation results and an empirical example suggest that the REML-based confidence interval is favored over the other two procedures in unbalanced one-way random effects model.

1. Introduction

Understanding and quantifying the sources of variation that influence a response variable is an integral part of statistical science. Fisher (1970) considered the study of variation, which gives rise to the concept of frequency distribution, as one of the three main aspects of statistics. In particular, constructing confidence intervals for functions of variance components provide information regarding the impact of a particular source of variation. If σ_1^2 and σ_2^2 denote variances associated with two sources that effect a response, then $\theta=\sigma_1^2/\sigma_2^2$ or $\rho=\sigma_1^2/(\sigma_1^2+\sigma_2^2)$ serve as ways to measure the importance of one effect compared to the other effect.

In this paper confidence intervals for θ or ρ are under consideration in the framework of one-way random effects model. Furthermore, unbalanced models are employed in an effort to address scenarios in which class or group sizes are not all equal. Imbalance may arise when a planned balanced experiment goes awry and yields missing or faulty measurements. Alternatively, in some experiments it may be difficult to obtain the same number of experimental units for each class and thus an unbalanced model is realized.

The statistical literature on confidence intervals for variance components in unbalanced one-way random effects model relies heavily on normal distribution assumptions. See Wald (1940), Thomas and Hultquist (1978), Harville and Fenech (1985), and Burch and Iyer (1997) for authors who based inferences on a pivotal quantity approach. Burdick et al. (2006)

* Tel.: +1 9285236875; fax: +1 9285235847.
E-mail address: brent.burch@nau.edu

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.06.015

discussed the present status of confidence interval estimation for one-factor random models and El-Bassiouni and Zoubeidi (2008) investigated the validity of confidence intervals in one-way random effects model under normal theory. Hannig et al. (2006), Lidong et al. (2008), and Hannig (2009) developed fudicial intervals for variance components in unbalanced normal mixed linear models. Jiang (1996, 2005) described the large-sample properties of restricted maximum likelihood (REML) estimators of θ or ρ which can be used to build asymptotic confidence intervals for variance components.

Burch (2011) presented confidence interval results for θ or ρ for the balanced one-way random effects model without assuming normality. This work depended on Jiang's (2005) result in that, under regularity conditions, the REML estimator is asymptotically normally distributed and its asymptotic variance does not depend on normally distributed responses. Estimating the kurtosis of the associated distribution is necessary to implement this procedure. The work of Burch (2011) is extended in the current article to investigate confidence intervals for variance components in non-normal unbalanced one-way random effects model based on the pivotal quantity, fiducial, and REML methods.

Section 2 presents an overview of the unbalanced one-way random effects model and states the properties of quadratic forms used to estimate variance components without relying on normal distribution assumptions. Included are descriptions of the three different procedures used to construct confidence intervals for θ or ρ. Section 3 provides confidence interval simulation results for a variety of scenarios and distributions. An application involving arsenic concentrations in oyster tissue is given in Section 4. A discussion and summary comments are given in Section 5.

2. Unbalanced one-way random effects model

Consider the one-way random effects model given by

$$Y_{ij} = \mu + A_i + e_{ij}, \tag{1}$$

where $i=1,\ldots,a$, $j=1,\ldots,b_i$, and the total sample size is $n=\sum_{i=1}^{a} b_i$. If the b_i's are not all equal, then the model is unbalanced. The a classes of the random factor A represent a random sample of classes from a population of classes. Furthermore, from within the ith class of A, a random sample of size b_i is selected and thus Y_{ij} is the jth observation associated with the ith class of A. The e_{ij}'s are often referred to as random errors and may be interpreted as the deviations of the observations within the classes. μ is the overall mean of the observations and in this paper is a nuisance parameter since the subject under study is variation.

It is assumed that $A_i \overset{iid}{\sim} (0,\sigma_1^2)$ with some underlying distribution, $e_{ij} \overset{iid}{\sim} (0,\sigma_2^2)$, with some underlying distribution, and the A_i's and e_{ij}'s are mutually independent. Observations within the same class are correlated since $Cov(Y_{ij},Y_{ij'})=\sigma_1^2$ for $j \neq j'$ and observations from different classes are uncorrelated. It follows that $E(Y_{ij})=\mu$, $Var(Y_{ij})=\sigma_1^2+\sigma_2^2$, and the third and fourth central moments of Y_{ij} are $E[(Y_{ij}-\mu)^3]$ and $E[(Y_{ij}-\mu)^4]$, respectively. Furthermore, the kurtosis of A_i and e_{ij} are given by $\kappa_1 = E(A_i^4)/\sigma_1^4-3$ and $\kappa_2=E(e_{ij}^4)/\sigma_2^4-3$, respectively. If, for example, the A_i's and e_{ij}'s are normally distributed, then the Y_{ij}'s are jointly normally distributed with $E[(Y_{ij}-\mu)^3]=0$, $E[(Y_{ij}-\mu)^4]=3(\sigma_1^2+\sigma_2^2)^2$, and $\kappa_1=\kappa_2=0$.

The parameter $\rho=\sigma_1^2/(\sigma_1^2+\sigma_2^2)$ is commonly referred to as the intraclass correlation coefficient since it is the correlation between two observations within the same class. ρ is also the proportion of the total variance in the Y_{ij}'s attributed to the random factor A since $\rho = Var(A_i)/Var(Y_{ij})$. By definition, $0 \le \rho < 1$. The ratio of variance components, denoted by $\theta=\sigma_1^2/\sigma_2^2$, and ρ are one-to-one functions since $\rho=\theta/(\theta+1)$ where $0 \le \theta < \infty$.

In an effort to construct confidence intervals for θ or ρ, we first examine quadratic forms of the Y_{ij}'s that provide the same information about σ_1^2 and σ_2^2 (and thus θ and ρ) as does the n-dimensional sample $Y_1,\ldots,Y_n$. The formation of these quadratic forms was provided by Burch and Iyer (1997) and Burch and Harris (2001) in linear mixed models having two sources of variation. For the one-way random effects model, let the $n \times 1$ vector $\mathbf{Y}$ represent the sample so that (1) can viewed in matrix notation as $\mathbf{Y}=\mathbf{1}\mu+\mathbf{ZA}+\mathbf{e}$, where $\mathbf{1}$ is an $n \times 1$ vector of ones, $\mathbf{Z}$ is an $n \times a$ matrix whose elements in the ith column are ones for the b_i observations in class A_i, and $\mathbf{e}$ is the $n \times 1$ error vector. Let $\mathbf{H}$ be an $n \times (n-1)$ matrix whose columns span the space orthogonal to the space spanned by the column vector of ones, and satisfies $\mathbf{H}'\mathbf{H}=\mathbf{I}$ where $\mathbf{I}$ is an $(n-1)\times(n-1)$ identity matrix. A Helmert-type matrix can be used to form $\mathbf{H}$. See Burch (2011) for additional details.

The $(n-1)\times 1$ vector $\mathbf{H}'\mathbf{Y}$ has mean vector zero and variance–covariance matrix $\sigma_2^2\mathbf{I}+\sigma_1^2\mathbf{H}'\mathbf{ZZ}'\mathbf{H}$ where $\mathbf{I}$ is an $(n-1)\times(n-1)$ identity matrix. Let $0=\Delta_1<\cdots<\Delta_d$ be the distinct eigenvalues of $\mathbf{H}'\mathbf{ZZ}'\mathbf{H}$ having multiplicities $r_1,\ldots,r_d$, respectively. There exists an $(n-1)\times(n-1)$ orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}'(\mathbf{H}'\mathbf{ZZ}'\mathbf{H})\mathbf{P}$ is a diagonal matrix with entries $\Delta_1,\ldots,\Delta_1,\ldots,\Delta_d,\ldots,\Delta_d$, where each Δ_i is repeated r_i times, $i=1,\ldots,d$. Note that $\mathbf{P}=[\mathbf{P}_1,\ldots,\mathbf{P}_d]$ where $\mathbf{P}_i$ corresponding to Δ_i is of size $(n-1)\times r_i$, and consider the quadratic forms $Q_i=\mathbf{Y}'(\mathbf{HP_iP_i}'\mathbf{H}')\mathbf{Y}$, $i=1,\ldots,d$.

The vector of quadratic forms, denoted by $\mathbf{Q}=(Q_1,\ldots,Q_d)$, where $d<n$, form a set of minimal sufficient statistics associated with the model devoid of the parameter μ. These quadratic forms can be used to estimate functions of variance components. The number of quadratic forms and their corresponding distributions depend on the underlying model structure. LaMotte (1976) provided a complete description of the properties of the quadratic forms in a one-way random effects model under normal distribution theory.

In general, it is known that $\Delta_1=0$, $r_1=n-a$ (if at least one $b_i>1$) and the zero eigenvalue signifies that there is replication in the experiment (multiple observations per class). An equation that relates the eigenvalues and their