Journal of the KSQC
Vol. 12, No, 2, 1984

Interval Estimation of Variance Components for the Random Twofold Nested Classification Model

Jung Il Kim*

ABSTRACT

Some statistics are presented for construction of confidence intervals of variance components for the unbalanced twofold nested classification model and they are shown to be approximated by chi-square distributions.

1. INTRODUCTION

Several methods have been proposed for setting confidence intervals for variance components in some particular linear models (see, for example, Welch (1956)). Previous methods have been developed for balanced designs. There are few published methods for constructing confidence intervals in the unbalanced case of the twofold nested random effects model.

In this paper statistics are presented which enable us to construct confidence intervals for variance components in the unbalanced random twofold nested classification model.

2. DEVELOPMENT OF NEW STATISTICS

The model under consideration is the unbalanced twofold nested random effects model

(2.1) $y_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk}$,

with $i = 1, \cdots, r$; $j = 1, \cdots, s$; and $k = 1, \cdots, n_{ij}$ Here μ is a general mean (constant), while α_i, β_{ij} and e_{ijk} are independent normal random variables with zero means and variances σ_1^2, σ_2^2 and σ_0^2 respectively.

Define $\bar{y}_{ij} = (1/n_{ij}) \sum_{k=1}^{n_{ij}} y_{ijk}$; $i=1, \cdots, r$; $j=1, \cdots, s$,

then (2.2) $\bar{y}_{ij} = \mu + \alpha_i + \beta_{ij} + \bar{e}_{ij}$

with $\bar{e}_{ij} = (1/n_{ij}) \sum_{k=1}^{n_{ij}} e_{ijk}$. The formular (2.2) can be written in a matrix form as

$$\bar{y} = \mu \underline{1}_{rs} + X\alpha + Z\beta + \bar{e}$$

with $\bar{y}' = (y_{11}, \cdots y_{rs})$, $\alpha' = (\alpha_1, \cdots, \alpha_r)$,
$\beta' = (\beta_{11}, \cdots \beta_{rs})$

* Department of Computer Science and Statistics, Seoul National University.

and $X = \text{diag}(\underline{1}_s, \cdots, \underline{1}_s)$, $Z = I_{rs}$ where $\underline{1}_s$ and I_{rs} are the column vectors of ones of dimensions and the identity matrix of order $rs \times rs$, respectively. The random vector $\bar{e}$ is normally distributed with a zero mean vector and a convariance matrix $V(\bar{e}) = \sigma_0^2 K$, where $K = \text{diag}(1/n_{11}, \cdots, 1/n_{rs})$. The convariance matrix for $\bar{y}$ turns out as

$$V(\bar{y}) = XX' \sigma_1^2 + I_{rs} \sigma_2^2 + K\sigma_0^2.$$

Consider the transformation $R' = (R_1', R_2')$, where $R_1 = \text{diag}(\underline{1}_s'/s^{\frac{1}{2}}, \cdots, \underline{1}_s'/s^{\frac{1}{2}})$, and R_2 consists of $r(s-1)$ orthonomal rows in the orthogonal complement of the row space of R_1 (see, Hultquist and Thomas (1978)). It follows that

$$E(R\bar{y}) = E\begin{pmatrix} R_1\bar{y} \\ R_2\bar{y} \end{pmatrix} = \begin{pmatrix} \mu s^{\frac{1}{2}} \underline{1}_r \\ \underline{0}_{r(s-1)} \end{pmatrix},$$

$$V(R\bar{y}) = \sigma_1^2 RXX'R' + \sigma_2^2 RR' + \sigma_0^2 RKR'.$$

Next form the $r \times r$ orthogonal matrix $H' = (h_1', H_2')$ with $h_1 = (1/r^{\frac{1}{2}}) \underline{1}_r'$ and H_2 consists of $r-1$ orthonormal rows in the orthogonal complement of h_1. Then

$$E(HR_1\bar{y}) = H\mu s^{\frac{1}{2}} \underline{1}_r = \begin{pmatrix} \mu(rs)^{\frac{1}{2}} \\ \underline{0}_{(r-1)} \end{pmatrix},$$

$$V(HR_1\bar{y}) = s\sigma_1^2 I + \sigma_2^2 I + \sigma_0^2 HR_1KR_1'H'.$$

Let P and Q be $(r-1)\times(r-1)$ and $r(s-1)\times r(s-1)$ matrices which diagonalize $H_2R_1KR_1'H_2'$ and R_2KR_2', respectively.
Let

$$PH_2R_1KR_1'H_2'P' = C = \text{diag}(c_1, \cdots, c_{r-1})$$

$$QR_2KR_2'Q' = D = \text{diag}(d_1, \cdots, d_{r(s-1)}),$$

where $c_1, \cdots, c_{r-1}$ and $d_1, \cdots, d_{r(s-1)}$ are the eigenvalues of $H_2R_1KR_1'H_2'$ and R_2KR_2', respectively. Then we have

$$PH_2R_1\bar{y} \sim N(\underline{0}, s\sigma_1^2 I + \sigma_2^2 I + \sigma_0^2 C),$$

$$QR_2\bar{y} \sim N(\underline{0}, \sigma_2^2 I + \sigma_0^2 D).$$

If P_i and Q_j denote the ith and jth rows of P and Q, respectively, then

$$u_i = P_iH_2R_1\bar{y} \sim N(\underline{0}, s\sigma_1^2 + \sigma_2^2 + c_i\sigma_0^2), \ r = 1, \cdots\cdots r-1;$$

$$v_j = Q_jR_2\bar{y} \sim N(\underline{0}, \sigma_2^2 + d_j\sigma_0^2), \ j = 1, \cdots, r(s-1);$$

and

$$u_i^2/(s\sigma_1^2 + \sigma_2^2 + c_i\sigma_0^2) \sim X^2(1), \ i = 1, \cdots, r-1;$$

$$v_j^2/(\sigma_2^2 + d_j\sigma_0^2) \sim X^2(1), \ j = 1, \cdots, r(s-).$$

Since P and Q are orthogonal,

$$\sum_{i=1}^{r-1} u_i^2 = \sum_{i=1}^{r-1} \bar{y}'R_1'H_2'P_i'P_iH_2R_1\bar{y} = \bar{y}'R_1'H_2'H_2R_1\bar{y} = s\sum_{i=1}^{r}(\bar{y}_{i.} - \bar{y}_{..})^2 = s(r-1)S_1^2,$$

$$\sum_{j=1}^{r(r-1)} v_j^2 = \sum_{j=1}^{r(s-1)} \bar{y}'R_2'Q_j'Q_jR_2\bar{y} = y'R_2'R_2y = \sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij} - \bar{y}_{i.})^2 = r(s-1)S_2^2.$$

where $\bar{y}_{i.} = (1/s)\sum_{j=1}^{s}\bar{y}_{ij}$,

$\bar{y}_{..} = (1/rs)\sum_{i=1}^{r}\sum_{j=1}^{s}\bar{y}_{ij}$ and

$$S_1^2 = \sum_{i=1}^{r}(\bar{y}_{i.} - \bar{y}_{..})^2/(r-1),$$

$$S_2^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij} - \bar{y}_{i.})^2/r(s-1).$$

Let $\tilde{n} = rs/\sum_{i=1}^{r}\sum_{j=1}^{s}(1/n_{ij})$, that is, the harmonic mean of the n_{ij}.

Then $\sum_{i=1}^{r-1} c_i = tr[H_2 R_1 K R_1' H_2'] = (r-1)/\tilde{n}$

and $\sum_{j=1}^{r(r-1)} d_j = tr[R_2 K R_2'] = r(s-1)/\tilde{n}$.

Now we define

$$W = \sum_{i=1}^{r-1} u_i^2 / (s\sigma_1^2 + \sigma_2^2 + \sigma_0^2/\tilde{n}),$$

$$V = \sum_{j=1}^{r(r-1)} v_j^2 / (\sigma_2^2 + \sigma_0^2/\tilde{n}),$$

or $$W = s(r-1) S_1^2 / (s\sigma_1^2 + \sigma_2^2 + \sigma_0^2/\tilde{n}),$$

$$V = r(s-1) S_2^2 / (\sigma_2^2 + \sigma_0^2/\tilde{n}).$$

Then it is proposed that the random variables W and V have approximately chi-square distributions with $r-1$ and $r(s-1)$ degrees of freedom, respectively. This fact can be proved asymptotically. Since.

$$W = \sum_{i=1}^{r-1} [\{u_i^2/(s\sigma_1^2 + \sigma_2^2 + c_i\sigma_0^2)\}$$
$$\{(s\sigma_1^2 + \sigma_2^2 + c_i\sigma_0^2)$$
$$/(s\sigma_1^2 + \sigma_2^2 + \sigma_0^2/\tilde{n})\},$$

$$V = \sum_{j=1}^{r(r-1)} [\{v_j^2/(\sigma_2^2 + d_j\sigma_0^2)\}$$
$$\{(\sigma_2^2 + d_j\sigma_0^2)/(\sigma_2^2 + \sigma_0^2/\tilde{n})\}],$$

applying the Slutsky's Theorem (see, for example, Bickel and Doksum (1977)) when $\tilde{n} \to \infty$ we obtain the results that the distributions of W and V are asymptotically same as those of $\sum_{i=1}^{r-1} u_i^2/(s\sigma_2^2 + \sigma_2^2 + c_i\sigma_2^2)$ and $\sum_{j=1}^{r(r-1)} v_j^2/(\sigma_2^2 + d_j\sigma_0^2)$, that is $X^2(r-1)$ and $X^2(r(s-1))$, respectively.

Now define $S_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2$, then S_0^2/σ_0^2 has a chi-square distribution with $(n-rs)$ degrees of freedom where $n = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij}$.

Furthermore S_0^2 is stochastically independent of W and V .

From the fact that the statistics W and V have approximately chi-square distributions, we can construct an approximate $100(1-\alpha)\%$ confidence interval on $s\sigma_1^2 + \sigma_2^2 + \sigma_0^2/\tilde{n}$ and $\sigma_2^2 + \sigma_0^2/\tilde{n}$ as follows (see, Burdick and Graybill (1984)):

$$(r-1)S_1^2/X^2_{\frac{\alpha}{2}}(r-1) < s\sigma_1^2 + \sigma_2^2 + \sigma_0^2/\tilde{n}$$
$$< (r-1)S_1^2/X^2_{1-\frac{\alpha}{2}}(r-1)$$

and $$r(s-1)S_2^2/X^2_{\frac{\alpha}{2}}(r(s-1)) < \sigma_2^2 + \sigma_0^2/\tilde{n}$$
$$< r(s-1)S_2^2/X^2_{1-\frac{\alpha}{2}}(r(s-1)).$$

The statistics W and V also can be used for constructing confidence intervals on linear combinations of variance components in the random twofold nested classification model.

3. NUMERICAL JUSTIFICATION

To examine the fact that the previously stated W and V have approximately chi-squared distributions with $r-1$ and $r(s-1)$ degrees of freedom, respectively, for the moderate sample sizes in each cell, six designs (given in Table 1) were selected for study. $P[W < X_\alpha^2]$ and $P[V < X_\alpha^2]$ were calculated for 4 values of α where $P[X^2 < X_\alpha^2] = \alpha$. These values are : 0.01, 0.05, 0.95, 0.99.

Although the accuracy of approximation depends to some extent on the values of α and n_{ij}, we have observed that the chi-square distributions are good approximations to the distributions of W and V . The example for the case $\sigma_1^2/\sigma_0^2 = \sigma_2^2/\sigma_0^2 = 0.25$ is given in Table 2, where the observations y_{ijk} are generated by the subroutine GGNML of IMSL.

TABLE 1

Design	Number of (r, s)	Values of n_{ij}
1	(3,2)	5,10,15,20,25,30
2	(3,2)	10,20,30,40,50,60
3	(4,3)	5,10,15,20,25,30,35,40,45,50,55,60
4	(4,3)	2,2,2,2,2,2,2,2,2,2,2,100
5	(3,5)	2,2,2,2,2,2,2,2,2,2,2,2,2,2,100
6	(5,2)	3,4,5,6,7,8,9,10,11,12

TABLE 2

Design		0.010	0.050	0.950	0.990
1	W	0.010	0.043	0.950	0.989
	V	0.010	0.041	0.940	0.989
2	W	0.018	0.050	0.952	0.992
	V	0.007	0.045	0.952	0.996
3	W	0.004	0.039	0.950	0.988
	V	0.014	0.050	0.939	0.986
4	W	0.013	0.052	0.954	0.994
	V	0.011	0.047	0.953	0.988
5	W	0.012	0.053	0.956	0.990
	V	0.012	0.053	0.959	0.994
6	W	0.009	0.052	0.948	0.992
	V	0.014	0.057	0.940	0.993

REFERENCE

(1) Bickel, P. J. and Doksum, K. A. (1977), Mathematical Statistics, Holden-Day, Inc., 460-462.

(2) Burdick, R. K. and Graybill, F. A. (1984), "Confidence intervals on linear combinations of variance components in the unbalanced one-way classification," Technometrics 26, 131-136.

(3) Hultquist, R. A. and Thomas, J. D. (1978), "Interval estimation for the unbalanced case of the one-way random effects model," Annals of Statistics, 6, 582-587.

(4) Welch, B. L. (1956), "On linear combinations of several variances," Journal of the American Statistical Association, 51, 132-148.