*Jour. Ind. Soc. Ag. Statistics*
*Vol. XXXVIII No. 2 (1986), pp.* 180-186

# ON THE CONFIDENCE INTERVAL FOR GROUP VARIANCE COMPONENT IN ONE WAY UNBALANCED RANDOM EFFECT MODEL

B. SINGH*
*C. I. R. G., Makhdoom, Mathura*

(Received : February, 1984)

SUMMARY

An approach for obtaining the confidence interval of the group variance component in one-way unbalanced random model is given by using two moments approximation for the distribution of a linear combination of independent chi-squares. Approximate confidence limits for the group variance component are obtained in terms of the population mean $\mu$ and the ratio of variance components $P$.

*Keywords* : Random model; Variance components ratio; Quadratic form; Confidence interval.

## Introduction

In balanced situations (equal group size), various methods for obtaining confidence interval of group variance component in one-way random effect model have been presented by several workers such as Bross [4], Bartlett [1], Green [6], Bulmer [5], William [13] and others. However, no such procedure in unbalanced situations (unequal group sizes) has yet been reported.

This paper attempts to find out a procedure for obtaining the confidence interval and hence the confidence limits for the group variance component for some known values of the population mean $\mu$ and the ratio of variance components $P$ in one way unbalanced random effect model.

*Present address : Livestock Economics and Statistics Division, IVRI, Izatnagar.

**Distribution and Confidence Interval**

In analysis of variance in one-way random classification $y_{ij}$, the $j$th observation belonging to $i$th group, is represented by :

$$y_{ij} = \mu + a_i + e_{ij}, \tag{2.1}$$

$$(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n_i;\ \sum_{i=1}^{k} n_i = N)$$

where

$\mu$ is the grand mean (fixed); $a_i$, the random effects due to groups, are *iid* normal with mean zero and variance $\sigma_a^2$; $e_{ij}$, the error variables independent of $a_i$, are *iid* normal with mean zero and variance $\sigma_e^2$, and $N$ is the total number of observations.

Here, $\sigma_a^2$, the group variance and $\sigma_e^2$, the error variance, are known as variance components of the model (2.1).

The between groups sum of squares SSB is defined by

$$\text{SSB} = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2, \tag{2.2}$$

where, $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{y}_{ij}$ and $\bar{y} = \frac{1}{N} \sum_{i=1}^{k} n_i \bar{y}_i$ are means.

Now define,

$$Z_i = a_i + \bar{e}_i$$

where, $\bar{e}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}$

SSB can be written as

$$\text{SSB} = \sum_{i=1}^{k} n_i (Z_i - \bar{Z})^2$$

where, $\bar{Z} = \frac{1}{N} \sum_{i=1}^{k} n_i Z_i$

Writing SSB in a matrix form as

$$\text{SSB} = Z^1 MZ, \tag{2.3}$$

where,

$$Z' = (Z_1, \ldots, Z_k) \text{ and } M = (m_{ij})_{k \times k}, \tag{2.4}$$

with,

$$m_{ij} = \begin{cases} n_i\left(1 - \frac{n_i}{N}\right), & i = j \\ -\frac{n_i n_j}{N}, & i \neq j \end{cases}$$

It can be shown that, under the assumption laid down in the model (2.1), the vector $Z$ is a multivariate normal with mean vector zero and variance-covariance matrix $V$, a diagonal matrix, given by

$$V = \text{diag}\left(\frac{\sigma_1^2}{n_1}, \ldots \frac{\sigma_k^2}{n_k}\right), \tag{2.5}$$

where,

$$\sigma_i^2 = \sigma_e^2 + n_i\sigma_a^2, \; i = 1, 2, \ldots k$$

By using the above relation for $\sigma_i^2$, the matrix $V$ can be written as

$$V = V_* + \sigma_a^2 I, \tag{2.6}$$

Where,

$$V_* = \text{diag}\left(\frac{\sigma_e^2}{n_1}, \ldots . \frac{\sigma_k^2}{n_k}\right).$$

With this, the expression (2.3) for SSB can be expressed as

$$\begin{aligned} \text{SSB} &= Z'MZ \\ &= Z'V^{-1}(V_* + \sigma_a^2 Z)\,MZ \\ &= Z_z V^{-1} V_* MZ + \sigma_a^2 Z'V^{-1}MZ, \end{aligned}$$

or,

$$\begin{aligned} \sigma_a^2 Z'V^{-1}MZ &= Z'(I - V^{-1}V_*)\,MZ \\ &= Z'\,\text{diag}\left(\frac{n_i P}{1 + n_i P}, \ldots, \frac{n_k P}{1 + n_k P}\right) MZ \\ &= \sum_{i=1}^{k} \frac{n_i^2 P Z_i (Z_i - \overline{Z})}{1 + n_i P}, \end{aligned} \tag{2.7}$$

Where, $P = \sigma_a^2/\sigma_e^2$, the ratio of variance components.

By using (2.1) and the definition of $Z_i$ given above, we can write (2.7) as

$$\sum_{i=1}^{k} n_i^2 P \frac{(\bar{y}_i - \mu)(\bar{y}_i - \bar{y})}{1 + n_i P} = \sigma_a^2 Z'V^{-1}MZ \tag{2.8}$$

The distribution of the quadratic form, $Z'V^{-1}MZ$, (see Johnson Kotz, [9], can be shown like $\sum_{t=1}^{k=1} V_t$, where $V_t/\lambda_t$ are independently distributed as chi-squares with single degrees of freedom and $\lambda_t$ are the non-zero characteristic roots of the matrix $M$ (2.4).

The exact distribution of $\sum_{t=1}^{k=1} V_t$ is very tedious for practical applications (see Johnson and Kotz, [9], Kanji [10]. It is, therefore, approximated by some constant times central chi-square, say $gX_h^2$. The values of $g$ and $h$ are obtained by equating the first two moments of $\sum_{t=1}^{k=1} V_t$ and $gX_h^2$ (See Box, 1954).

Then, the equations for $g$ and $h$, obtained by equating the mean and variance of $\sum_{t=1}^{k=1} V_t$ and $gX_h^2$, are given by

$$\text{mean}\left(\sum_{t=1}^{k=1} V_t\right) = g \cdot h \tag{2.9}$$

and

$$\text{variance}\left(\sum_{t=1}^{k=1} V_t\right) = 2g^2h$$

The mean and variance of $\sum_{t=1}^{k=1} V_t$ can be obtained by using the results of theorem 2, (Box, [3]), as

$$\text{mean} = t_r M$$

$$= N - \sum_{i=1}^{k} \frac{n_i^2}{N}$$

and,

$$\text{Variance} = 2t_r M^2$$

$$= 2\left[\sum_{i=1}^{k} n_1^2 - 2\sum_{i=1}^{k} \frac{n_i^3}{N} + \left(\sum_{i=1}^{k} \frac{n_i^2}{N}\right)^2\right].$$

By substituting the values of the mean and variance of $\sum_{t=1}^{k=1} V_t$ in (2.9) and then solving the same, we obtain the values of $g$ and $h$ as

$$g = \frac{\sum_{i=1}^{k} n_i^2 - 2\sum_{i=1}^{k} \frac{n_i^3}{N} - \left(\sum_{i=1}^{k} \frac{n_i^2}{N}\right)^2}{N - \sum_{i=1}^{k} \frac{n_i^2}{N}}, \qquad (2.10)$$

and

$$h = \frac{\left(N - \sum_{i=1}^{k} \frac{n_i^2}{N}\right)^2}{\sum_{i=1}^{k} n_i^2 - 2\sum_{i=1}^{k} \frac{n_i^3}{N} + \left(\sum_{i=1}^{k} \frac{n_i^2}{N}\right)^2}.$$

In this way, we have that

$\sum_{i=1}^{k} (n_i^2 P/1 + n_i P)(\bar{y}_i - \mu)(\bar{y}_i - \bar{y}) = X$ (say) is distributed, approximately, as $\sigma_a^2 g \chi_h^2$, where, $g$ and $h$ are given in (2.9) above.

Now the confidence interval of $\sigma_a^2$ can be obtained as follows :
We know that

$$P\left(X^2_{h,1-\alpha/2} \leqslant \frac{X}{g\sigma_a^2} \leqslant X^2_{h,\alpha/2}\right) = 1 - \alpha;$$

which implies that

$$P\left(\frac{X}{gX^2_{h,\alpha/2}} \leqslant \sigma_a^2 \leqslant \frac{X}{gX^2_{h,1-\alpha/2}}\right) = 1 - \alpha; \qquad (2.10)$$

where,

$X^2_{h,\alpha/2}$ and $X^2_{h,1-\alpha/2}$ are the points on the abscisa of chi-square probability curve such that the probability area on the right hand side of these points is $\alpha/2$ and $1 - \alpha/2$, respectively.

Thus, equation (2.10) gives the $(1 - \alpha)\%$ confidence interval for $\sigma_a^2$ in terms of $X$, where $\alpha$ is an arbitrary. That is, if $\mu_0$ and $P_0$ are some specified values of population mean and the ratio of variance components, then the confidence interval of $\sigma_a^2$, with approximately $(1\ \alpha)\%$ coverage, can be obtained as

$$\left(\frac{X}{gX^2_{h,\alpha/2}}, \frac{X}{gX^2_{h,1-\alpha/2}}\right), \qquad (2.11)$$

with

$$X = \sum_{i=1}^{k} \frac{n_i^2 P_0}{1 + n_i P_0} (\bar{y}_{i.} - \mu_0)(\bar{y}_i - \bar{y}).$$

Such type of studies have also been reported by William [13]. Townsend & Searle [11], and others, where the confidence or the point estimators for variance components have been obtained in terms of the unknown parameters like mean ($\mu$) or the error variance ($\sigma_e^2$).

## 3. Approximation for $X$

Supposed $n_i^2$ are large enough so that

$$\frac{n_i P}{1 + n_i P} \simeq 1, \qquad V_i = 1, 2, \ldots, k. \tag{3.1}$$

Then $X$ can be approximated as

$$X = \sum_{i=1}^{k} n_i (\bar{y}_i - \mu)(\bar{y}_i - \bar{y})$$

$$= \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2. \tag{3.2}$$

The above expression (3.2) for $X$ is nothing but the between groups sum of squares, SSB. Thus, the confidence interval, using $X =$ SSB, can be worked out without knowing the values of $\mu$ and $P$. However, this confidence interval will over estimate moderately, depending upon the ratio of $n_i P/1 + n_i P$, the bounds of the $(1 - \alpha)\%$ confidence interval of the group variance components $\sigma_a^2$.

ACKNOWLEDGEMENT

I am grateful to Dr. D. D. Joshi, Professor of Statistics and Director, Institute of Social Sciences, Agra, for their valuable suggestions and to the referee for his critical comments for improving the manuscript.

REFERENCES

[1] Bartlett, R. C. (1953) : Approximate confidence intervals II. More than one unknown parameter, *Biometrika*, 40 : 306-17.

[2] Boardman, T. J. (1974) : Confidence intervals for variance components. A comparative Monte Carlo study, *Biometrics*, 30 : 251-62.

[3] Box, G. E. P. (1954) : Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification, *Ann. Math. Stat.*, 25 : 290-302.

[4] Bross; I. (1950) : Fiducial intervals for variance components, *Biometrics*, **6** : 136-44.

[5] Bulmer, M. G. (1957) : Approximate confidence limits for components of variance, *Biometrika*, **44**; 159-67.

[6] Green, J. R. (1954) : A confidence interval for variance components, *Ann. Math. Stat.*, **25** : 671-86.

[7] Healy, W. C. (1961] : Limits for a variance components with an exact confidence coefficient, *Ann. Math. Stat.*, **32** ; 466-76.

[8] Huitson, A. (1955) : A method of assigning confidence limits to linear combinations of variance, *Biometrika*, **42** : 471-79.

[9] Johnson, N. L. and Kotz, S. (1970): *Continuous Unvariate Distributions*—2, Chapter 29, John Wiley & So s, Inc., New York.

[10] Kanji, G. K. (1979) : Power aspects of Anova in fixed effect model one-way classification, *The Australian Jour. of Stat.*, **22**(1) : 36-44.

[11] Towensend, E. C. and Searle, S. R. (1971) : Bert-quadratic unbiased estimators of variance components from unbalanced data in the one-way classification, *Biometrika*, **27** : 643-657.

[12] Welch, B. L. (1956) : On linear combinations of several variances, *JASA*, **51** : 132-48.

[13] William, J. S. (1962) : A confidence interval for variance components, *Biometrika*, **49** : 278-81.