Jour. Ind. Soc. Ag. Statistics
Vol. XLI, No. 1 (1989) pp. 77-84

MOMENTS OF GROUP VARIANCE COMPONENT ESTIMATOR IN ONE-WAY UNBALANCED CLASSIFICATION

B. SINGH*
Central Institute for Research on Goats Makhdoom, P.O. Farah-281122, Mathura (U.P.)

(Received : March, 1984)

SUMMARY

The expression for the distribution of the between groups sum of squares is obtained as a linear combination of central chi-squares and hence that of the estimator of group variance component, in one-way unbalanced random classification. The rth cumulant of the group variance component estimator is derived in terms of trace of a power matrix, whose elements depend only on group sizes.

Keywords : Chi-square, Characteristic roots, Variance Covariance matrix, Efficiency.

Introduction

In balanced situations, the components sums of squares are distributed independently as some constant times chi-squares for normal populations. When group sizes are unequal, the non-null distribution of the between groups sum of squares in one-way random classification, is not a constant times chi-square. Hence, under unbalanced situations, the distribution of analysis of variance estimator of the group variance component is not a linear combination of two central chi-squares as the case with balanced situations. However, the estimator can be expressed as a linear combination of many central chi squares some of them with negative coefficients (Harville, [4]).

*Present Address : L.E.S. Division, I.V.R.I., Izatnagar (U.P.) 243122.

For the one-way classification Crump [2] and Searle [10] give separate expressions for the variance of the between groups variance component estimator with some typographical errors. The expressions for higher order moments are not yet available in the literature. In this paper we obtain an expression for the analysis of variance estimator of the group variance component as a linear combination of independent central chi-square variables in one-way unbalanced random classification. Using these expressions we derive the rth cumulant in general and first four moments as special cases, for the group variance component estimator.

2. Distribution of Components Sums of Squares

The jth observation in the ith group Y_{ij}, in one-way unbalanced classification, is represented by an equation.

$$Y_{ij} = m + a_i + e_{ij}, \tag{2.1}$$

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

where m is the grand mean (fixed), a_i, effects due to groups, are i i d normal with mean zero and variance σ_a^2; e_{ij}, error variables independent of a_i, are i i d normal with mean zero and variance σ_e^2; n_i is the ith group size; k is the number of groups and N is the total number of observations. Here σ_a^2, the group variance, and σ_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

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

where

$$\overline{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{and} \quad \overline{Y} = \frac{1}{N} \sum_{i=1}^{k} n_i \overline{Y}_i$$

are the means.

Using the model equation (2.1), SSB in (2.2) can be written as

$$SSB = \sum_{i=1}^{k} n_i (a_i + \bar{e}_i - \bar{a} - \bar{e})^2,$$

where

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

and $\bar{a} = \frac{1}{N} \sum_{i=1}^{k} n_i a_i$.

Let $Z_i = a_i + \bar{e}_i$, then SSB becomes

$$SSB = \sum_{i=1}^{k} n_i (Z_i - \bar{Z})^2$$
$$= \underset{\sim}{Z}' M \underset{\sim}{Z} \quad \text{(say)}, \tag{2.3}$$

where

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^{k} n_i Z_i, \quad \underset{\sim}{Z}' = (Z_1, \ldots, Z_k)$$

and

$$M = (m_{ij}) \text{ with } m_{ij} = \begin{bmatrix} n_i \left(1 - \frac{n_i}{N}\right), & i = j \\ -\frac{n_i n_j}{N}, & i \neq j \end{bmatrix} \tag{2.4}$$

It can be seen that, under the assumptions laid down in model (2.1), $\underset{\sim}{Z}$ is a multivariate normal vector with mean as null vector and variance covariance matrix, as a diagonal matrix V, given by

$$V = d_{iag}\left(\frac{\sigma_1^2}{n_1}, \ldots, \frac{\sigma_k^2}{n_k}\right)$$

with $\sigma_i^2 = \sigma_e^2 + n_i \sigma_a^2$.

Let $\lambda_1, \lambda_2, \ldots \lambda_{k-1}$ are the non-zero characteristic roots of matrix $U = VM$, then (2.3) can be expressed as (see Box, [1])

$$SSB = \sum_{t=1}^{k-1} \lambda_t U_t, \tag{2.5}$$

where $U_1, \ldots, U_{k-1}$ are the independent chi-squares each with single degrees of freedom.

By elementary matrix operations for the determinantal equation $| U - \lambda I | = 0$ and following Lamotte [7], the characteristic roots λ_t of the matrix U can be expressed as

$$\lambda_t = \lambda_t^* \sigma_a^2 + \sigma_e^2, \tag{2.6}$$

where

λ_t^*, $(t = 1, 2, \ldots k - 1)$, are the non-zero roots of the matrix M (2.4), whose elements depend on group sizes only. Now using (2.6), the expression for SSB in (2.5) becomes

$$SSB = \sum_{t=1}^{k-1} U_t \, (\lambda_t^* \sigma_a^2 + \sigma_2), \tag{2.7}$$

where the non zero characteristic roots λ_t^* of M are exclusively functions of group sizes.

It is known that, under the assumptions laid down in the model (2.1), the within groups sum of squares

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2$$

is distributed as

$$\sigma_e^2 \chi^2 N - k. \tag{2.8}$$

Singh [12] has used the expressions (2.7) and (2.8) to derive the non-null distribution of ANOVA F-ratio and whence the probability of getting negative estimates of the group variance component estimator in a two variance components model. The next section uses these expressions for deriving the moments for the estimator of the group variance component, σ_a^2, in the model (2.1).

3. Moments of the Group Variance Components Estimator

The analysis of variance estimator of the group variance component σ_a^2 in the model (2.1) (Henderson, [5]) is defined by

$$\hat{\sigma}_a^2 = \frac{1}{k^1}\left[\frac{SSB}{k-1} - \frac{SSE}{N-k}\right], \tag{3.1}$$

where

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

Now using the results (2.7) and (2.8) the estimator $\hat{\sigma}_a^2$ can be expressed as a linear combination of independent chi-squares variables as given by

$$\hat{\sigma}_a^2 = \frac{1}{k^1}\left[\sum_{t=1}^{k-1} \frac{\lambda_t^* \sigma_a^2 + \sigma_e^2}{k-1} \, U_t - \frac{\sigma_e^2}{N-k} \chi^2_{N-k}\right]$$

$$= \sum_{t=1}^{N-1} \lambda_t' \, U_t, \quad \text{say} \tag{3.2}$$

where

$$\lambda_t = \begin{cases} \dfrac{\lambda_t^* \sigma_a^2 + \sigma_e^2}{k^1 (k-1)}, & t = 1, 2, \ldots, k-1 \\ -\dfrac{\sigma_e^2}{k^1 (N-k)}, & t = k, k+1, \ldots, N-1 \end{cases}$$

and $U_1, U_2, \ldots, U_{N-1}$, are independent chi-squares each with single degree of freedom.

The expression for the rth cumulant of $\hat{\sigma}_a^2$ (3.1) can be, obtained as (see Box [1], theorem 2.2)

$$k_r = 2^{r-1} \underline{|r-1} \sum_{t=1}^{N-1} (\lambda_t)^r$$

$$= 2^{r-1} \underline{|r = 1} \left[\sum_{t=1}^{k-1} \left(\frac{\lambda_t^* \sigma_a^2 + \sigma_e^2}{k^1 (k-1)} \right)^r + (-1)^r \left(\frac{\sigma_e^2}{(N-k)} \right)^r (N-k) \right].$$

Expanding the term $(\lambda_t^* \sigma_a^2 + \sigma_e^2)^r$ by the binomial expansion and rearranging it, we get

$$K_r = \frac{2^{r-1} \underline{|r-1}}{(k^1)^r} \left[\frac{1}{(k-1)^r} \sum_{j=0}^{r} \binom{r}{j} (\sigma_a^2)^j \sum_{t=1}^{k-1} (\lambda_t^*)^j (\sigma_e^2)^{r-j} + (-1)^r \frac{(\sigma_e^2)^r}{(N-k)^{r-1}} \right]. \tag{3.3}$$

By using the relation of Box [1]

$$\sum_{t=1}^{k-1} (\lambda_t^*)^r = t_r M^r, \quad r \geqslant 1 \tag{3.4}$$

where t_r denotes trace of a matrix, the expression for k_r, in (3.3), becomes

$$k_r = \frac{2^{r-1} \underline{|r-1}}{(k^1)^r} \left[\frac{1}{(k-1)^r} \sum_{j=1}^{r} \binom{r}{j} (\sigma_e^2)^{r,j} (\sigma_a^2)^j t_r M^j + (\sigma_e^2)^r \left(\frac{1}{(k-1)^{r-1}} + \frac{(-1)^r}{(N-k)^{r-1}} \right) \right]. \tag{3.5}$$

The simplified form of k_r, the rth cumulant of $\hat{\sigma}_a^2$, given in (3.5) above, has been derived with the help of the result (2.7) which has been obtained using the relation (2.6) of characteristic roots.

The result (2.7), for distribution of *SSB*, has made the binomial expansion applicable in separating out the group variance component σ_a^2 and the error variance component σ_e^2 from the variance covariance matrix. Now the value of moments/cumulants for σ_a^2 for any order can be obtained from (3.5) for any value of σ_a^2 or σ_e^2 just by evaluating the trace of the power of the matrix M. As the elements of M are functions of group sizes only, so there is no need of evaluating the trace of the power matrix for each value of the variance components, σ_a^2 or σ_e^2, respectively for computing the moments of σ_a^2.

The first four cumulants of σ_a^2, obtained from (3.5) are given by

$$k_1 = \sigma_a^2,$$

$$k_2 = 2\left[\left(\frac{\sigma_e^2}{k^1}\right)^2 \frac{N-1}{(k-1)(N-k)} + \frac{2\,\sigma_e^2\,\sigma_e^2}{k^1(k-1)} + \left(\frac{\sigma_a^2}{k^1}\right)^2 \frac{t_r\,M^2}{(k-1)^2}\right],$$

$$k_3 = 8\left[\left(\frac{\sigma_e^2}{k^1}\right)^3 \frac{(n-1)(N-2k+1)}{(k-1)^2(N-k)^2} + 3\left(\frac{\sigma_e^2}{k^1}\right)^2 \frac{\sigma_a^2}{k^1}\,\frac{t_r\,M}{(k-1)^2}\right.$$
$$\left. + 3\,\frac{\sigma_e^2}{k^1}\left(\frac{\sigma_a^2}{k^1}\right)^2 \frac{t_r\,M^2}{(k-1)^3} + \left(\frac{\sigma_a^2}{k^1}\right)^3 \frac{t_r\,M^3}{(k-1)^3}\right]$$

and

$$k_4 = 48\left[\left(\frac{\sigma_e^2}{k^1}\right)^4\left(\frac{1}{(k-1)^3} + \frac{1}{(N-k)^3}\right) + 4\left(\frac{\sigma_e^2}{k^1}\right)^3 \frac{\sigma_a^2}{k^1}\,\frac{t_r\,M}{(k-1)^4}\right.$$
$$+ 6\left(\frac{\sigma_a^2}{k^1}\right)^2\left(\frac{\sigma_e^2}{k'}\right)^2 \frac{t_r\,M^2}{(k-1)^4} + 4\,\frac{\sigma_e^2}{k^1}\left(\frac{\sigma_a^2}{k^1}\right)^3 \frac{t_r\,M^3}{(k-1)^4}$$
$$\left. + \left(\frac{\sigma_a^2}{k^1}\right)^4 \frac{t_r\,M^4}{(k-1)^4}\right], \tag{3.6}$$

where $t_r\,(M)^r$ for $r = 1, 2, 3, 4 \ldots$ can be obtained from (2.4)

$$t_r\,M = N - \sum_{i=1}^{k} n_i^2/N$$
$$= k^1(k-1),$$

$$t_r\,M^2 = \left[\left(\sum_{i=1}^{k} \frac{n_i^2}{N}\right)^2 - 2\sum_{i=1}^{k} n_i^3/N + \sum_{i=1}^{k} n_i^2\right],$$

$$t_r\,M^3 = \sum_{i=1}^{k}\sum_{i'=1}^{k} \frac{n_i^2\,n_{i'}^2}{N^2}\left(n_i + n_i' - \sum_{i=1}^{k} \frac{n_i^2}{I}\right)$$
$$+ \sum_{i=1}^{k} n_i\left(1 + \sum_{i=1}^{k} \frac{n_i^2}{N^2} - 3\,\frac{n_i}{N}\right)$$

and

$$t_r M^4 = \sum_{i=1}^{k}\sum_{i'=1}^{k} \frac{n_i^2 n_{i'}^{2}}{N^2}\left(n_i + n'_i - \sum_{i=1}^{k} \frac{n_i^2}{N}\right)^2$$

$$+ \sum_{i=1}^{k} n_i^4 \left(1 + 2\sum_{i=1}^{k} \frac{n_i^2}{N^2} - 4\frac{n_i}{N}\right). \quad (3.7)$$

The central moments of $\hat{\sigma}_a^2$ can now be obtained from (3.7) above, with the following relations

$$\mu_2 = k_2, \quad \mu_3 = k_3 \quad \text{and} \quad \mu_4 = k_4 + 3k_2^2. \quad (3.8)$$

Leone and Nelson [8] and Leone *et al.* [9], while investigating empirically the sampling distribution of the variance components estimators in nested designs, found that for normal populations the Pearson type III curves are suitable for approximating the distribution of variance components estimators. Using the moments of the components sums of squares in one way ANOVA model David and Johnson [3] have approximated the distribution of a linear combination of sums of squares correspond, in general to curves of Pearson type IV to study the effect of non-normality and heterogeneity of variance on tests of general linear hypothesis. Similarly, the moments expressions (3.6) can be used to approximate the sampling distribution of the group variance correspondent estimator $\hat{\sigma}_a^2$ (3.1) to a suitable curve. The other use of these moments can be for estimating the sampling error or the efficiency of the estimator $\hat{\sigma}_a^2$.

ACKNOWLEDGEMENT

The author is grateful to Dr. D. D. Joshi, Professor of Statistics and Director, Institute of Social Sciences, Agra for his valuable suggestions and to the referee for his critical comments in improving the manuscript.

REFERENCES

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

[2] Crump, S. L. (1951) : Present status of variance component analysis, *Biometrics*, **7** : 1-16.

[3] David, F. N. and Johnson, N. L. (1951) : A method of investigating the effect of non-normality and heterogeneity of variance on tests of general linear hypothesis, *Ann. Math. Stat.*, **32** : 466-476.

[4] Harville, D. A. (1969) : Expression of variance component estimator as linear

combination of independent non-central chi-squares variates, *Ann. Math. Stat.*, 40 : 2189-2194.

[5] Henderson, C. R. (1953) : Estimation of variance and covariance components, *Biometrics*, 9 : 226-252.

[6] Johnson, N. L. and Kotz, S. (1970) : *Continuous Univariate Distribution*, Vol. II. John Wiley and Sons.

[7] Lamotte, L. R. (1976) : Invariant quadratic estimators in the random one-way ANOVA model, *Biometrics*, 32 : 793-804.

[8] Leone, F. C. and Nelson, L. S. (1966) : Sampling distributions of variance components. I. Empirical studies of balanced nested designs, *Technometrics*, 8 : 457-468.

[9] Leone, F. C., Nelson, L. S., Johnson, N. L. and Eisenhart, S. (1968) : Sampling distributions of variance components. II. Empirical studies of unbalanced nested designs, *Technometrics*, 10 : 719-737.

[10] Searle, S. R. (1956) : Matrix methods in variance and covariance components analysis, *Ann. Math. Stat.*, 2 : 737-748.

[11] Searle, S. R. (1971) : Topics in variance component estimation, *Biometrics*, 27 : 1-76.

[12] Singh, B. (1986) : Distribution of variance components estimators in a two-variance components model, *Sankhya Ser. B*, 48 : (in press).